Appendix A: State and Site Award Data

Appendix A presents additional information on when Reading First Impact Study sample sites first received Reading First awards (Exhibit A.1).

Exhibit A.1: Award Date by Site in Order of Date when Reading First Funds Were First Made Available for Implementation

	Date Initial Reading First Award Was Announced	Date when Reading First Funds Were First Made Available for Implementation
Site 9	03/2003	04/2003
Site 12	04/2003	05/2003
Site 2	06/2003	06/2003
Site 6	05/2003	06/2003
Site 5	02/2003	07/2003
Site 4	05/2003	07/2003
Site 18	06/2003	08/2003
Site 10*	10/2003	08/2003
Site 11*	10/2003	10/2003
Site 17*	08/2003	12/2003
Site 14	01/2004	02/2004
Site 8	01/2004	03/2004
Site 3	03/2004	04/2004
Site 13	01/2004	04/2004
Site 15	10/2003	05/2004
Site 1	05/2004	06/2004
Site 7	05/2004	06/2004
Site 16	03/2004	08/2004

NOTE:

Sites 10, 11 and 17 "backdated" the point at which schools could begin spending their grant money. It is not an error that the schools appear to have been given their money before their grants were announced.

SOURCE: Reading First District Coordinators

Appendix B: Methods

This appendix describes the general regression discontinuity approach used to estimate the impacts of Reading First and presents the specific models used to estimate impacts. In addition, it describes how the issue of multiple hypothesis testing was addressed and provides information about statistical precision.

Part 1: Regression Discontinuity Design

Approach

The Reading First Impact Study is based on a regression discontinuity design that capitalizes on the systematic process used by a number of school districts to allocate their Reading First funds.[1] A regression discontinuity design is the strongest quasi-experimental method that exists for estimating program impacts. Under certain conditions (which are met by the present study) this method can approach the rigor of a randomized experiment.[2] The conditions include:

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.
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.
3) Funding decisions were based only on whether a school's rating was above or below its local cut-point; nothing superseded these decisions; and further,
4) The shape of the relationship between schools' ratings and outcomes is correctly modeled.

To see how the method works, consider a hypothetical school district that allocates its $2 million annual Reading First grant to 10 schools in equivalent allotments of $200,000, per year, per school. The district also has prioritized the schools with the highest rates of poverty, as measured by the percentage of students eligible for free or reduced priced meals. The district therefore awards grants first to the school with the highest poverty rate, then to the school with the next-highest poverty rate, and so on, until ten schools receive grants and all of the Reading First funding has been allocated.

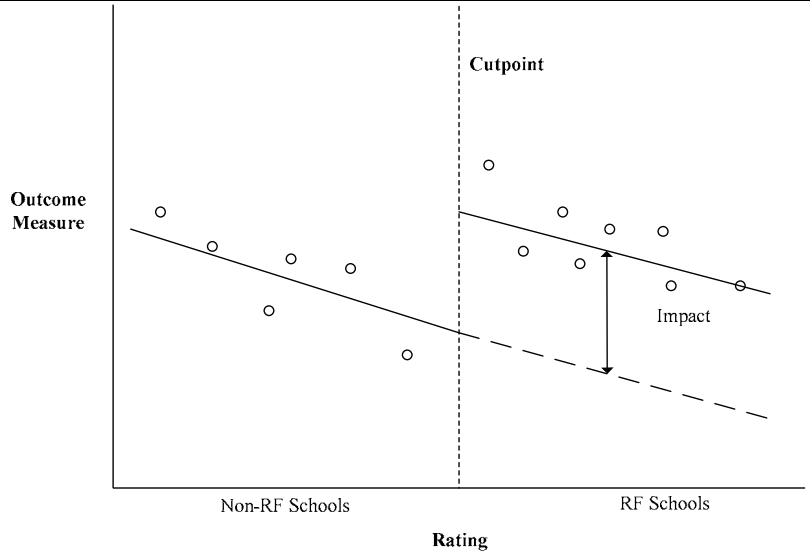
Exhibit B.1 illustrates how the dividing line, or "cut-point," between the last funded school and the first school *not funded* on the district's priority list (or between the 10th and 11th schools on this hypothetical district's list) creates a "discontinuity" that makes it possible to estimate program impacts on future outcomes. The vertical axis of the exhibit represents a future outcome measure for each school, such as its

[1] The Reading First Impact Study was originally planned as a randomized control study, in which eligible schools from a sample of districts were to receive Reading First funds or become members of a non-Reading First control group. The approach was not feasible, however, in the 38 states that had already begun to allocate their Reading First grants before the study began. Furthermore, in the remaining states, randomization was counter to the spirit of the Reading First Program, which strongly emphasizes serving the schools most in need. It was possible, however, to randomize schools in one site.

[2] Regression discontinuity analysis was introduced by Thistlethwaite and Campbell (1960) and has more recently experienced a resurgence of interest (e.g., Cappelleri et al., 1991; Cook, 2008; Goldberger, 1972; Hahn, Todd and Van Der Klaauw, 2001; Mohr, 1995; Reichardt, Trochim, and Cappelleri, 1995; and Trochim, 1990).

average student reading score in a subsequent year. The horizontal axis represents the rating used to determine each school's priority for Reading First (in this example, the percentage of past students eligible for free or reduced price meals). Schools to the left of the cut-point do not receive Reading First funding and serve as a "comparison group" for the impact analysis; these schools are referred to as non-Reading First schools. Schools to the right of the cut-point receive Reading First funding; these schools represent the "treatment group" for the impact analysis, and are referred to as Reading First schools.

Exhibit B.1: Regression Discontinuity Analysis for a Hypothetical School District



The exhibit illustrates a downward-sloping relationship between schools' ratings and their future outcomes. This implies that schools with a higher proportion of past (and thus future) students who live in poverty will tend to have lower levels of future student achievement. In the absence of Reading First, average student achievement at non-Reading First schools would therefore tend to be higher than at Reading First schools. Consequently, the average outcome for non-Reading First schools most likely over-states what this average would have been for Reading First schools without the program (their "counterfactual"). Because of this, a simple comparison of average outcomes for Reading First schools and non-Reading First schools would understate the impact of Reading First.

Given the way that schools were selected for Reading First, however, it is possible to obtain unbiased estimates of the program's impacts on future outcomes by controlling statistically for the relationships that exist between school outcomes and ratings. (These relationships comprise the "regression" part of regression discontinuity analysis.) Intuitively, this analysis would proceed as follows. The first step is to fit a regression line through the data points for non-Reading First schools, as indicated by the solid line to the left of the cut-point in Exhibit B.1. The second step is to extrapolate the fitted line across the cut-point to predict what student achievement would have been for Reading First schools—in the absence of the program. This is indicated by the dashed line in the exhibit. The third step is to fit a regression line through the data points for Reading First schools, as indicated by the solid line to the right of the cut-point. (For the purpose of this hypothetical example, the two fitted lines are assumed to have the same slope and are thus parallel, which simplifies the analysis but is not necessary.) The impact of Reading First thus can be measured by the vertical distance between the solid fitted line for Reading First schools

(what actually happened in Reading First schools after the program was launched) and the dashed extrapolated line for Reading First schools (the counterfactual prediction of what would have happened in Reading First schools without the program). This distance is indicated by a two-sided arrow.

In short, the analysis uses the observable discontinuity in the regression relationship to identify the impact of Reading First. The magnitude of the discontinuity indicates the magnitude of the impact. If the regression model has the correct shape for the data being modeled (for example, two parallel straight lines for Reading First and non-Reading First schools), the discontinuity provides an unbiased impact estimate.

The approach works properly, if schools' ratings are the only thing that determines their selection for Reading First. Consequently, only background characteristics that are correlated with ratings can be correlated with selection for the program. In other words, the only characteristics that can differ systematically between Reading First schools and non-Reading First schools are those correlated with their ratings. Controlling statistically for the ratings thereby controls for any systematic pre-existing differences between the two groups of schools.[3] It is this control that makes unbiased impact estimates possible.

Seventeen of the 18 sites in the Reading First Impact Study (16 school districts and one state program) allocated their Reading First grants in ways that meet the requirements of a regression discontinuity design. Each site prioritized its eligible schools according to a specified quantitative indicator, in most cases, an indicator based on a measure of student poverty, student performance, or both.[4] (See Exhibit B.2 for the criteria used by each site to rate its schools for Reading First.) Each site then allocated its Reading First funds according to the prioritized list, funding the top priority school first, the second priority school next, and so on through the list, until all available resources were allocated. In the context of this study, these sites are referred to as regression discontinuity design (RDD) sites.

The study sample was drawn from Reading First schools and non-Reading First schools whose ratings were as close as possible to their sites' local cut-point.[5] Half of the schools in the study sample are Reading First schools and half are non-Reading First schools.[6] Only 9 of the 248 sample schools from study sites had their rating-based Reading First funding status changed. Consequently, the study's sites support what is called a "sharp" regression discontinuity analysis, which is the strongest form of the design.[7]

---

[3] It is because regression discontinuity analysis utilizes "selection on observables" (i.e., values of the rating) that it can produce unbiased impact estimates (Cain, 1975). This feature is what distinguishes the approach from other quasi-experimental designs.

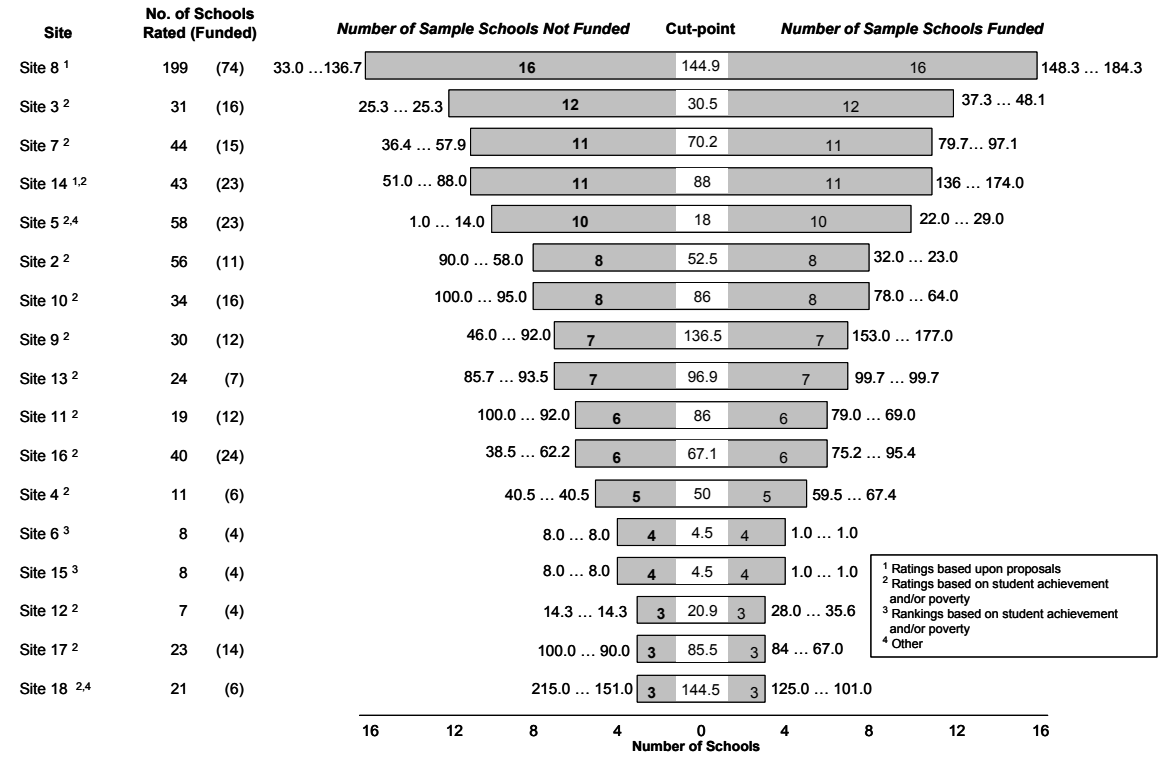
[4] A separate rating coefficient (in the impact estimation model) was specified for each site to account for differences in rating variables and cut-points. These differences enhance the generalizability of the present study because it comprises 17 regression discontinuity analyses from different parts of the United States.

[5] Note that the RDD can be compromised if there is little or no variation on the rating variable within treatment and comparison groups in a given site. As illustrated in Exhibit B.2, however, the schools selected for the study sample were both close to their local cutpoints and varied with respect to the rating variable. Therefore, this potential problem was not present in the study sample.

[6] These proportions were exact for the original study sample of 258 schools. With the subsequent loss of 10 schools, they remain almost exact.

[7] A sharp regression discontinuity analysis has very few cases where assignment to treatment or comparison status based on ratings is changed due to other considerations. A "fuzzy" regression discontinuity design has more such aberrant cases. A fuzzy regression discontinuity analysis is more complex and requires further assumptions (Shadish, Cook and Campbell, 2002).

Exhibit B.2: Numbers, Ratings, and Cut-points for Selection of Reading First and Reading First Impact Study Schools, by Site (Initial Sample for 17 Sites, Excluding Random Assignment Site)



NOTES:

Ratings varied in directionality and metrics; in some sites, higher scores indicated greater needs; in other sites, lower scores indicated greater needs.

EXHIBIT READS: Site 8 rated 199 schools, and funded 74 schools. The RFIS sample in Site 8 included 32 schools—16 non-Reading First schools and 16 Reading First schools—that were rated from 136.7 to 148.3, shown at the left and right sides of the shaded bar, respectively. The cut-point was at 144.9. The lowest school rating was 33, and the highest school rating was 184.3.

SOURCES: Interviews with sites' Reading First coordinators in 2004.

In the 18th study site (a school district), it was possible to randomly assign a subset of its Reading First-eligible schools to receive or not receive Reading First funds. In this site, five candidate schools were assigned to Reading First and five were assigned to a control group. Hence, this site provides a group-randomized experiment. This site is referred to as the experimental site.

Sample Size

Although regression discontinuity analysis can provide unbiased impact estimates under the conditions met by this study—and thus is comparable to a true experiment in this regard—the quasi-experimental approach requires a much larger sample of schools to provide the same precision as an experiment because one must include the rating variable in any models to account for the design effect (Bloom, Kemple and Gamse, 2004). The study team conducted analyses of the effect of including the rating of schools as a covariate for a regression discontinuity analysis of program impacts. The team determined that if ratings are used as a covariate, the variance of the impact estimator for a regression discontinuity

analysis will be four times that for a corresponding experiment.[8] Hence, to achieve the same minimum detectable effect the regression discontinuity analysis would need four times as many schools as the experiment.

Based on these analyses and extensive discussions among members of the research team, IES staff, and the project's technical work group , it was decided that a sample of roughly 240 schools was needed, which is four times the sample size planned for the original experimental design. This larger sample size was necessary for the study to achieve a minimum detectable effect size of 0.20 standard deviations. As noted above, initial recruitment efforts produced a sample of 258 schools from one state site and 17 district sites. These 18 sites represent a total of 13 states. Due to refusals, school closings, reconfiguring, or redistricting, 10 schools (4 RF schools and 6 non-RF schools) subsequently dropped out of the study. For results presented in this report, a final analytic sample of 248 schools was used. (See Exhibit B.3 for a flowchart of sample selection from regression discontinuity design target sample to the final analytic sample.)

Specification Tests

In developing the study sample, Reading First schools and non-Reading First schools were selected to be as close as possible to their local cut-points for receipt of Reading First funding. This was done to yield two groups of schools that were as similar as possible. (See Exhibit B.4 for unadjusted baseline characteristics of schools in the study sample.) In addition, program impacts were estimated using a linear regression discontinuity model that controls for values of the ratings used to choose schools for program funding. Furthermore, estimates of impacts on measures of student reading achievement control explicitly for school-level baseline measures of reading achievement. This *combination* of sample design and statistical analysis was expected to provide internally valid estimates of program impacts.

Three sets of specification tests were conducted to assess whether this expectation was met.[9] Although none of these tests by itself can *prove* that internal validity was achieved, in combination they provide evidence that this is most likely the case. The most important such test used a linear regression discontinuity model to compare baseline characteristics of Reading First schools and non-Reading First schools. If a linear regression discontinuity model is an appropriate way to control for all pre-existing differences between the two groups, observable or not, then it should eliminate their observed baseline differences.

Baseline specification tests were conducted using aggregate school-level baseline characteristics.[10] The results of these tests in Exhibit B.5 show that none of the adjusted residual differences between Reading First schools and non-Reading First schools for the selected baseline characteristics were statistically significant. Hence, there is little evidence of residual differences in these school-level baseline characteristics. Results of these tests do not provide statistical evidence of substantial bias in impact estimates for the present report. Also, because impact estimates for student reading comprehension control explicitly for observed differences in school-level mean baseline test scores (typically the strongest predictor of future test scores), they provide further protection against bias.

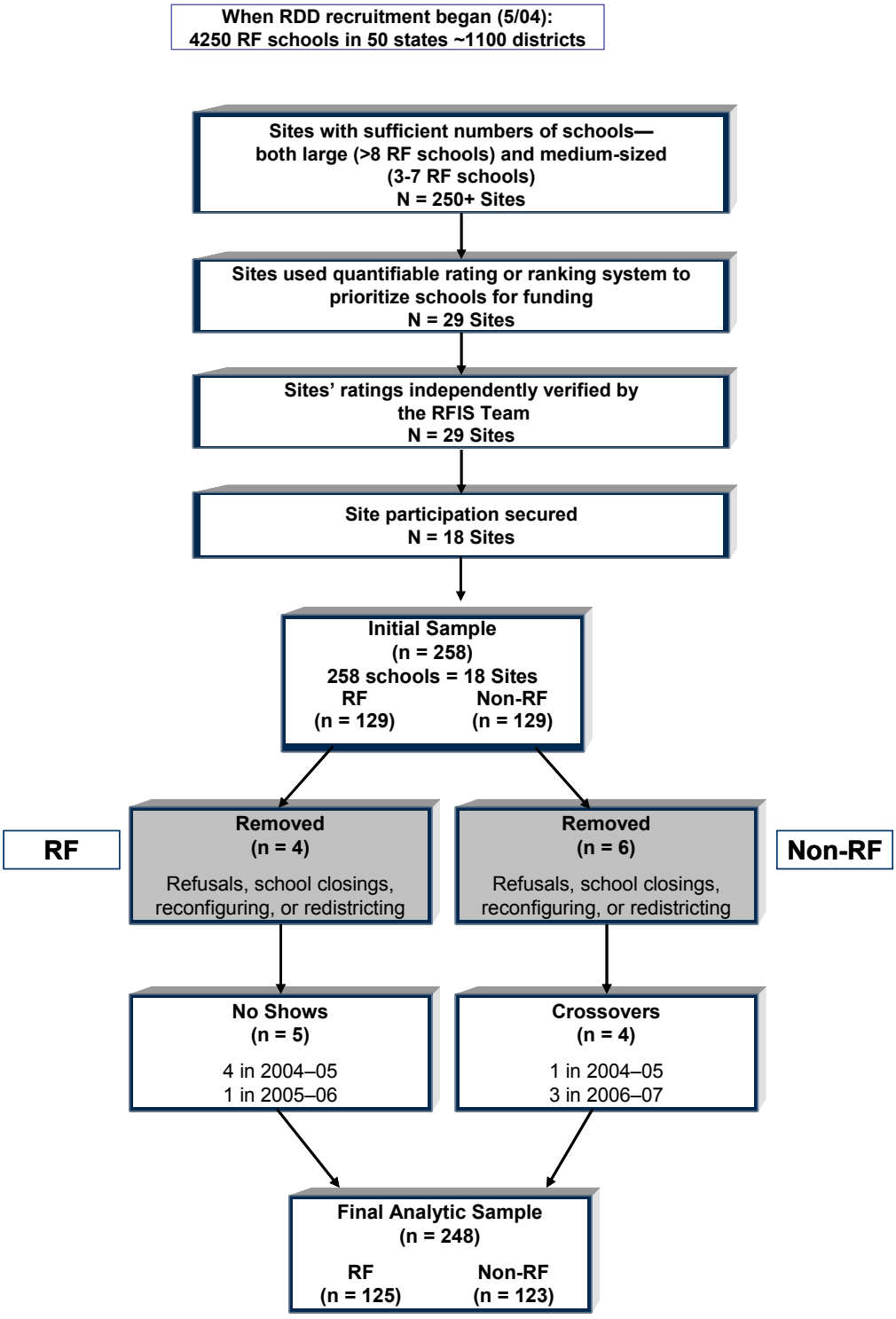
---

[8] See the study's Interim Report Appendix B, Part 5 (Gamse, Bloom, Kemple & Jacob, 2008) for details of these analyses.

[9] See the study's Interim Report Appendix B (Gamse, Bloom, Kemple & Jacob, 2008) for a detailed presentation of the specification tests conducted to assess the study's internal validity.

[10] Baseline data were available at the school level only.

Exhibit B.3: RFIS Sample Selection: From Regression Discontinuity Design Target Sample to Analytic Sample



*The final analytic sample includes 146 schools from 7 sites that have 8 or more RF schools (74 RF, 72 non-RF schools) and 102 schools from 6 sites that have between 3 and 7 RF schools (51 RF, 51 non-RF schools).

Exhibit B.4: Observed Differences in Baseline Characteristics of Schools in the Study Sample: 2002-2003

Characteristic	Actual Mean for Reading First Schools	Actual Mean for Non-Reading First Schools	Difference	Statistical Significance of Difference (p-value)
Students				
Male (%)	52.3	51.6	0.7*	(0.049)
Race (%)				
Asian	3.1	3.3	-0.2	(0.670)
Black	35.6	33.9	1.7	(0.532)
Hispanic	26.7	22.5	4.1*	(0.021)
White	34.2	39.8	-5.6*	(0.006)
American Indian/Alaskan	0.5	0.5	0.0	(0.847)
Free Lunch and Reduced Lunch (%)	74.4	68.9	5.5*	(0.002)
Schools				
Eligible for Title I (%)	97.6	90.7	6.9*	(0.013)
Locale (%)				
Large City	39.2	37.4	1.8	(0.476)
Mid-size City	36.8	34.6	2.2	(0.434)
Other[a]	24.0	28.0	-4.0	(0.286)
Size				
Total Number of Students	474.8	488.7	-13.9	(0.462)
Number of Students in Grade 3	71.6	76.0	-4.4	(0.162)
Student/Teacher Ratio	15.1	15.2	-0.1	(0.613)
Third Grade Reading Performance				
Deviation from State RF Mean Proficiency Rate (%)[b]	-1.3	1.8	-3.0*	(0.019)

NOTES:

The complete RF 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.

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

[a] Other Locale includes urban fringe of a large city, urban fringe of a mid-sized city, large town, small town, and rural.

[b] A school-'s proficiency score is defined as the percentage of third grade students (or fourth or fifth grade when third grade is unavailable) in the school that score at or above the state-defined proficiency threshold on the state's reading assessment. The values in this row represent the average percentage point deviation from the mean proficiency score for the Reading First schools in the state.

EXHIBIT READS: On average, , 52.3 percent of students in Reading First schools and 51.6 percent of students in non-Reading First schools were male. The difference on the percent of male students between Reading First and non-Reading First schools was 0.7 percentage points. The difference was statistically significant at the $p \leq .05$ level ($p=.049$).

SOURCES: Data on baseline characteristics are from the Common Core of Data.

Exhibit B.5: Estimated Residual Differences in Baseline Characteristics of Schools in the Study Sample: 2002-2003

Characteristic	Estimated Residual Difference	Statistical Significance of Difference (p-value)
Students		
Male (%)	0.9	(0.246)
Race (%)		
Asian	0.9	(0.363)
Black	-7.2	(0.199)
Hispanic	3.3	(0.345)
White	2.8	(0.503)
American Indian/Alaskan	0.2	(0.182)
Free Lunch and Reduced Lunch (%)	-6.0	(0.073)
Schools		
Eligible for Title I (%)	-1.4	(0.802)
Locale (%)		
Large City	4.3	(0.419)
Mid-size City	9.1	(0.108)
Other[a]	-13.4	(0.083)
Size		
Total Number of Students	-0.9	(0.982)
Number of Students in Grade 3	-3.8	(0.558)
Student/Teacher Ratio	0.1	(0.861)
Third Grade Reading Performance		
Deviation from State RF Mean Proficiency Rate (%)[b]	4.3	(0.085)

NOTES:

The complete RF 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 "Estimated Residual Difference" is the adjusted residual differences between Reading First schools and non-Reading First schools estimated using the regression discontinuity model, which controls for each school's rating.

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

[a] Other Locale includes urban fringe of a large city, urban fringe of a mid-sized city, large town, small town, and rural.

[b] A school's proficiency score is defined as the percentage of third grade students (or fourth or fifth grade when third grade is unavailable) in the school that score at or above the state-defined proficiency threshold on the state's reading assessment. The values in this row represent the average percentage point deviation from the mean proficiency score for the Reading First schools in the state.

EXHIBIT READS: The estimated residual difference on the percent of male students between Reading First and non-Reading First schools was 0.9 percentage points. The difference was not statistically significant at the $p \leq .05$ level (p=.246).

SOURCES: Data on baseline characteristics are from the Common Core of Data.

Part 2: Estimation Methods

The slightly different statistical models used to estimate the impact of Reading First on the three major outcome domains (student reading achievement, classroom instruction, and student engagement with print), as well as surveys, shared most elements. However, because there were some differences in the models for reading achievement, classroom instruction and student engagement with print and surveys, the approach for each is described separately below.

Impact tables throughout the report and appendices contain: 1) the actual, unadjusted mean outcomes for Reading First schools in the study sample ("Actual Mean with Reading First"), 2) the best estimate of what would have happened in RF schools absent RF funding ("Estimated Mean without Reading First"), 3) the impact estimate, [11] and 4) the effect size of the impact estimate.[12, 13]

Impact Estimation Method for Reading Achievement

The statistical model used to estimate RF impacts on student reading comprehension and decoding is described by (1) below:

$$Y_{ijkm} = \sum_{mt} \beta_{0m} S_{mk} YR_t + \sum_{m} \beta_{1m} S_{mk} T_k + \sum_{mt} \beta_{2m} S_{mk} R_k YR_t + \sum_{mt} \beta_{3m} S_{mk} \overline{Y_{-1km}}\, YR_t + \sum_{t} \gamma_t Z_{jk} YR_t + \sum_{nt} \theta_n X_{nijkm} YR_t + \mu_k + \upsilon_{jk} + \varepsilon_{ijk} \quad (1)$$

where:

Y_{ijkm} = the post-test for student i from classroom j in school k in site m,

S_{mk} = one if school k is in site m and zero otherwise, m = 1 to 18,

T_k = one if school k is a treatment school and zero otherwise,

R_k = the rating for school k (standardized and centered by site),

$\overline{Y_{-1km}}$ = the mean baseline pretest for school k (standardized and centered by site),

YR_t = indicator for follow-up years, 2005, 2006, or 2007[14]

Z_{jk} = a variable indicating when the post-test was given for classroom j in school k (site-centered),

X_{nijkm} = demographic characteristic n of student i from classroom j in school k, in site m

μ_k , υ_{jk} and ε_{ijk} = school-level, classroom- level, and student-level random error terms, respectively, assumed to be independently and identically distributed.

---

[11] The estimates of what would have happened in RF schools absent RF funding are calculated by subtracting the impact estimates from the RF schools' actual mean values.

[12] 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 all years for which the outcome was available.

[13] When calculating the effect sizes, standard deviation from the non-Reading First schools were used instead of the pooled standard deviation from Reading First and non-Reading First schools because the treatment could have effected the distribution—hence the standard deviation—of the outcomes in Reading First schools but not in non-Reading First schools. The study team wanted to use a stable standard deviation and non-Reading First schools provided that. It is also important to note that the standard deviations for the student outcomes obtained from non-Reading First schools are very close to those observed in the national norming sample.

[14] For decoding, this indicator is not used because only one year of data is available for this outcome.

The average estimated value of β_{1m} (m = 1, 2, …, 18), weighted by the number of RF schools in each site, is the program impact for the average RF school in the study sample.

The student achievement impact model (Equation 1) used to estimate impacts on reading comprehension and decoding has the following characteristics:

- It is a multi-level model that reflects the nested structure of the data by accounting for three levels of clustering in the estimation of standard errors: clustering of students within classrooms, classrooms within schools, and schools within study sites.
- Baseline covariates are added to the model to improve precision. These covariates include student gender, student age at start of school year,[15] date of the post-test at the classroom level, and a school-level pre-program reading performance measure.[16, 17]
- The rating variable was not included in the model for the one site that assigned schools to Reading First and non-Reading First groups randomly.
- In estimating reading comprehension pooled impacts for the combined sample from 2005, 2006 and 2007, the covariates for site, rating, pretest, test date, and demographic characteristics were interacted with an indicator for follow-up year (2005, 2006 or 2007).
- In estimating decoding impacts, the covariates for site, rating, pretest, test date, and demographic characteristics were not interacted with an indicator for follow-up year because there is only one year of data for this outcome.

Impacts on Classroom Instruction and Student Engagement with Print

The impacts of Reading First on classroom instruction and student engagement with print were estimated using the following three-level model (with observations at level one, classrooms at level two, and schools at level three):

$$Y_{ijkm} = \sum_{mt} \beta_{0m} S_{mk} YR_t + \sum_{m} \beta_{1m} S_{mk} T_k + \sum_{mt} \beta_{2m} S_{mk} R_k YR_t + \mu_k + \nu_{jk} + \varepsilon_{ijk} \qquad (2)$$

Where:

Y_{ijkm} = the outcome measure for observation i from classroom j in school k in site m,
S_{mk} = one if school k is in site m and zero otherwise, (m= 1,2, …, 18),
T_k = one if school k is a treatment school and zero otherwise,
R_k = the rating for school k (standardized and centered by site),
YR_t = indicator for follow-up years, 2005, 2006, or 2007,[18]

---

[15] Age at start of the school year is each student's age as of September 1 of the given year. For example, age as of September 1, 2005 for the 2005-2006 school year.

[16] Different pre-program performance measures were constructed for early and late award sites. For the 10 early award sites and one late award site (which had no fall 2004 test data due to a hurricane), performance on a state reading test (when available, we used an average of test scores from up to three pre-RF years) was used as a school level pretest measure. For late award sites except for the one without available fall 2004 data, the mean fall 2004 SAT 10 test scores for each school/grade were used as the pretest measure.

[17] As a robustness test, the analysis was conducted without some or all of these additional covariates and the impact estimates stayed virtually unchanged. *Results for these additional tests are available upon request*.

[18] For the STEP, only two year indicators are included in the model, since STEP data was not collected in the first year.

μ_k, ν_{jk} and ε_{ijk} = school-level, classroom-level, and observation-level random error terms, respectively, assumed to be independently and identically distributed.

The impact estimate is the average estimated value of β_{1m} (m = 1, 2, …, 18) weighted by number of treatment schools in each site.

The impact estimation model for classroom instruction and student engagement with print described by (Equation 2) has the following characteristics:

- It is a multi-level model that reflects the nested structure of the data by accounting for three levels of clustering in the estimation of standard errors: clustering of observation days within classrooms, classrooms within schools, and schools within sites.
- A rating variable was not included in the model for the one site that assigned schools to Reading First and non-Reading First groups randomly.
- In estimating pooled impacts for the combined sample from 2005, 2006 and 2007, the covariates for site and rating were interacted with an indicator for follow-up year (2005, 2006, or 2007).

Estimation Method for Surveys

Data from self-report surveys of teachers, reading coaches, and principals were used to estimate the impact of Reading First on the key components of scientifically based reading instruction.[19] Two models were needed to estimate differences for survey data—one for classroom level data (i.e., from teacher survey) and a second for school level data (i.e., from reading coach survey or principal survey). Differences for classroom level survey data were estimated using the following two-level model (with classrooms at level one and schools at level two):

$$Y_{jkm} = \sum_{mt} \beta_{0m} S_{mk} + \sum_{m} \beta_{1m} S_{mk} T_k + \sum_{mt} \beta_{2m} S_{mk} R_k + \nu_k + \varepsilon_{jk} \qquad (3)$$

Where:

Y_{ijkm} = the outcome measure for classroom j in school k in site m,[20]

S_{mk} = one if school k is in site m and zero otherwise, (m= 1,2, …, 18),

T_k = one if school k is a treatment school and zero otherwise,

R_k = the rating for school k (standardized and centered by site),

ν_k and ε_{jk} = school-level and classroom-level random error terms, respectively, assumed to be independently and identically distributed.

The difference estimate is the average estimated value of β_{1m} (m = 1, 2, …, 18) weighted by number of treatment schools in each site.

The impact estimation model for classroom level survey data described by (Equation 3) has the following characteristics:

---

[19] Only 2007 survey data is included in these analyses due to low survey response rates in 2005.

[20] To maintain parallel structure with other estimation models presented in this appendix, the nomenclature for classroom (j), school (k), and site (m) remains the same even in the absence of student or observation level survey data (i).

- It is a multi-level model that reflects the nested structure of the data by accounting for two levels of clustering in the estimation of standard errors: clustering of classrooms within schools and schools within sites.
- A rating variable was not included in the model for the one site that assigned schools to Reading First and non-Reading First groups randomly.
- Only one year of data were used for survey data, so no interactions with the follow-up year were included in the estimation model.

Differences for school level survey data were estimated using the following ordinary least squares regression model:

$$Y_{km} = \sum_{mt} \beta_{0m} S_{mk} + \sum_{m} \beta_{1m} S_{mk} T_k + \sum_{mt} \beta_{2m} S_{mk} R_k + \varepsilon_k \quad (4)$$

Where:

Y_{ijkm} = the outcome measure for school k in site m,[21]

S_{mk} = one if school k is in site m and zero otherwise, (m= 1,2, …, 18),

T_k = one if school k is a treatment school and zero otherwise,

R_k = the rating for school k (standardized and centered by site),

ε_k = school-level random error term assumed to be independently and identically distributed.

The difference estimate is the average estimated value of β_{1m} (m = 1, 2, …, 18) weighted by number of treatment schools in each site.

The impact estimation model for school level survey data described by (Equation 4) has the following characteristics:

- It is a single-level ordinary least squares regression model that accounts for one level of clustering in the estimation of standard errors: clustering of schools within sites.
- A rating variable was not included in the model for the one site that assigned schools to Reading First and non-Reading First groups randomly.
- Only one year of data were used for survey data, so no interactions with the follow-up year were included in the estimation model.

Part 3: Approach to Multiple Hypothesis Testing

This section addresses the issue of multiple hypothesis testing. It first summarizes the five core principles that were used as a guide for addressing the issue in the current study, and then describes a two-stage approach for operationalizing these principles.

[21] To maintain parallel structure with other estimation models presented in this appendix, the nomenclature for school (k) and site (m) remains the same even in the absence of student or observation level survey data (i) or classroom level survey data (j).

Principle #1: Qualify tests instead of adjusting them: The present analysis qualifies specific hypothesis tests using composite tests of pooled hypotheses rather than (1) adjusting significance levels (through Bonferroni methods) or (2) adjusting significance thresholds (through Benjamini and Hochberg methods) of specific tests.

Principle #2: Address multiple testing differently for the central research questions of the study and for supplemental analyses. The analysis specifies two tiers of hypotheses: Tier I comprises a small number of hypotheses about the central research questions of the study, and Tier 2 represents supplemental research questions. Multiple testing is treated separately and differently within the two tiers. Statistical tests of Tier I hypotheses are considered confirmatory. To address the issue of multiplicity within Tier I, the present study tested a reduced set of outcomes by conducting pooled tests of composite hypothesis that represent a set of hypotheses that have been tested separately. The Tier 2 hypothesis tests are allowed to be much larger and less confirmatory. It may or may not be necessary to qualify these findings for multiple testing since they are not confirmatory.

Principle #3: Delineate separate domains that reflect key clusters of constructs represented by the central research questions of a study. Domains comprise broad clusters of outcome constructs that can contain multiple measures, subgroups, or follow-up observations. Domains are defined conceptually, and do not provide narrow "silos" for collecting findings. The central domains for the present study are student reading achievement, classroom reading instruction, and student engagement with print. In addition, survey data is a domain for exploratory analyses of support for scientifically based reading instruction across study schools.

Principle #4: Report analyses to address multiple comparisons in the background of research reports, not in the foreground. For the present study references to the qualifying tests occur in the main text but not in tables.

Principle #5: Use tests for interactions as a composite test (and thus a guide) for focusing on subgroup findings.[22]

Based on the above five principles, the present study uses the following two-stage approach to address multiple hypothesis testing. The first stage involves prioritizing outcomes and subgroups for the impact analysis. The second stage encompasses strategies for conducting composite tests on pooled key outcomes. The core features of each stage are described below.

Stage 1: Creating a Parsimonious List of Outcomes and Subgroups and Prioritizing Key Outcomes

The first stage of the framework involves a process of carefully categorizing and prioritizing the outcomes and subgroups for the impact analysis. The goal of this exercise is to create the shortest possible list of outcomes and subgroups that reflect the most proximal and policy relevant indicators of Reading First's effectiveness. Analytically, the shorter the list, the less likely it is that one would attribute

[22] If differences between impacts for subgroups are not statistically significant, then individual subgroup results should be interpreted with caution.

statistical significance to an impact that did not truly occur. These outcomes and subgroups were selected within distinct measurement domains to correspond to key components of the program's theory of action and the key research questions posed by the program's evaluation.

The impact analysis focuses on two components of the Reading First theory of action: 1) aligning teachers' instructional practices and behaviors with the five dimensions of reading instruction,[23] and 2) improving students' reading achievement.[24] The highest priority outcomes within each of these measurement domains would constitute "Tier 1" outcomes for the impact analysis.

Recognizing that a short list of outcomes will almost certainly exclude important policy-relevant indicators of Reading First's effectiveness (a form of Type II error), this first stage of the framework also includes the development of a secondary, or "Tier 2," list of outcomes and subgroups. As discussed below, the present study treats Tier 1 and Tier 2 outcomes and their accompanying subgroups separately, and potentially differently, if or when making adjustments to the standards used for judging statistical significance.

Exhibit B.6 provides a list of the Tier 1 and Tier 2 outcomes defined for each measurement domain for this report. Also displayed are the grade levels and follow-up periods on which the impact analyses focus.

Stage 2: Conducting Composite Tests to Qualify Specific Hypothesis Tests

One approach to qualifying multiple hypothesis tests is to test whether the overall effect of treatment on a family of outcomes is significantly different from zero. For example, a policy maker may be interested in the effect of an intervention on test scores in general, rather than on each subject separately. Measurement of such *overall* effects has its roots in the literature on clinical trials and on meta-analysis (O'Brien, 1984; Logan and Tamhane, 2002; and Hedges and Olkin, 1985). The present analysis constructs summary indices that aggregate information over multiple treatment effect estimates within each domain for Tier 1 outcomes, as well as for survey constructs in Tier 2. See Exhibit B.7.

Reading Comprehension

To qualify the impact estimates for each outcome measure for each grade in the reading comprehension domain, the present analysis ran a composite regression that pooled the sample across grades 1, 2, and 3 and two measures: scaled scores and an indicator of whether or not a student scored at or above grade

[23] The RFIS observational instrument, the IPRI, focused primarily on teacher behaviors. In order to ensure that the study also collected some data on student behaviors during observed reading instruction, the RFIS team developed the Student Time-on-Task Engagement with Print (STEP) instrument. Because student engagement with print is an outcome that is distinct from the student reading comprehension or classroom reading instruction domains, it is treated separately.

[24] The Reading First theory of action also includes allocating additional resources for districts and schools to purchase reading curricula, materials, and assessments; exposing teachers to professional development and coaching focused on the five dimensions of effective reading programs; and holding districts and schools accountable for improved reading achievement. The present study was able to measure the impact of Reading First on some of these other elements using survey data.

Exhibit B.6: Outcome Tiers for the Reading First Impact Analysis

Tier	Domain	Outcome	Impacts Estimate			Variation
			Year	Grade	Sample	
Tier 1	**Student Achievement**	***Reading Comprehension Scaled Score (SAT 10)***	2005, 2006, 2007 Pooled	Separate for Grade 1, 2, 3	Full	**N/A**
		% At or Above Grade Level (SAT 10)	2005, 2006, 2007 Pooled	Separate for Grade 1, 2, 3	Full	
		TOSWRF Standard Score	2007	Separate for Grade 1	Full	
	Instruction	***Time on Five Dimensions***	2005, 2006, 2007 Pooled	Separate for Grade 1, 2	Full	**N/A**
		Highly Explicit Instruction	2005, 2006, 2007 Pooled	Separate for Grade 1, 2	Full	
		High Quality Practice	2005, 2006, 2007 Pooled	Separate for Grade 1, 2	Full	
	Student Engagement with Print	***% Students Engaged with Print***	2006, 2007 Pooled	Separate for Grade 1, 2	Full	**N/A**
Tier 2	**Student Achievement**	***Reading Comprehension Scaled Score (SAT 10)***	2005	Separate for Grade 1, 2, 3	Full, Award Subgroup	• **Variation over time** • **Variation across sites** • **Variation by Award Subgroup**
			2006	Separate for Grade 1, 2, 3	Full, Award Subgroup	
			2007	Separate for Grade 1, 2, 3	Full, Award Subgroup	
			2005, 2006, 2007 Pooled	Separate for Grade 1, 2, 3	Full, Award Subgroup	
			2007	Separate for Grade 3	2005/2007 Stayer Subgroup	• **Variation by Student Exposure**
		% At or Above Grade Level (SAT 10)	2005	Separate for Grade 1, 2, 3	Full	**N/A**
			2006	Separate for Grade 1, 2, 3	Full	
			2007	Separate for Grade 1, 2, 3	Full	
			2005, 2006, 2007 Pooled	Separate for Grade 1, 2, 3	Full, Award Subgroup	
	Instruction	***Time on Five Dimensions (Combined and for Five Dimensions separately)***	2005	Separate for Grade 1, 2	Full, Award Subgroup	• **Variation over time** • **Variation across sites** • **Variation by Award Subgroup**
			2006	Separate for Grade 1, 2	Full, Award Subgroup	
			2007	Separate for Grade 1, 2	Full, Award Subgroup	
			2005, 2006, 2007 Pooled	Separate for Grade 1, 2	Full, Award Subgroup	

Exhibit B.6: Outcome Tiers for the Reading First Impact Analysis (continued)

Tier	Domain	Outcome	Impacts Estimate			Variation
			Year	Grade	Sample	
Tier 2	Instruction	*Highly Explicit Instruction*	2005	Separate for Grade 1, 2	Full	N/A
			2006	Separate for Grade 1, 2	Full	
			2007	Separate for Grade 1, 2	Full	
			2005, 2006, 2007 Pooled	Separate for Grade 1, 2	Full, Award Subgroup	
		High Quality Practice	2005	Separate for Grade 1, 2	Full	N/A
			2006	Separate for Grade 1, 2	Full	
			2007	Separate for Grade 1, 2	Full	
			2005, 2006, 2007 Pooled	Separate for Grade 1, 2	Full, Award Subgroup	
	Student Engagement with Print	*% Students Engaged with Print*	2006	Separate for Grade 1, 2	Full	N/A
			2007	Separate for Grade 1, 2	Full	
	Professional Development in SBRI	*Amount of PD in reading received by teachers*	2007	Grades 1, 2, 3 Pooled	Full	N/A
		Teacher receipt of PD in the five essential components of reading instruction	2007	Grades 1, 2, 3 Pooled	Full	
		Teacher receipt of coaching	2007	Grades 1, 2, 3 Pooled	Full	
		Amount of time dedicated to serving as K-3 reading coach	2007	Grades 1, 2, 3 Pooled	Full	

Exhibit B.6: Outcome Tiers for the Reading First Impact Analysis (continued)

Tier	Domain	Outcome	Impacts Estimate			Variation
			Year	Grade	Sample	
Tier 2	Amount of Reading Instruction	*Minutes spent on reading instruction per day*	2007	Grades 1, 2, 3 Pooled	Full	N/A
	Supports for Struggling Readers	*Availability of differentiated instructional materials for struggling readers*	2007	Grades 1, 2, 3 Pooled	Full	N/A
		Provision of extra classroom practice for struggling readers	2007	Grades 1, 2, 3 Pooled	Full	
	Use of Assessments	*Use of assessments to inform classroom practice*	2007	Grades 1, 2, 3 Pooled	Full	N/A

Exhibit B.7: Summary of Impacts and Results of Composite Tests

Outcome Measure	Impact (p-value)			Result of Composite Test
	Grade 1	Grade 2	Grade 3	
Reading Comprehension				
• Standard scaled score	4.74 (p=0.083)	1.69 (p=0.462)	0.30 (p=0.887)	p=0.957 for composite test across 3 grades and 2 outcomes
• Percent reading at or above grade level	4.22 (p=0.104)	1.60 (p=0.504)	-0.08 (p=0.973)	
Instruction				
• Minutes of instruction in 5 reading dimensions	6.92* (p=0.005)	9.79* (p<0.001)	--	p<0.001 for composite test across 2 grades and 3 outcomes
• Highly explicit instruction	3.29* (p=0.018)	3.00* (p=0.040)	--	
• High quality student practice	0.82 (p=0.513)	2.94* (p=0.019)	--	
Student Engagement with Print				
• Percent of students engaged with print	5.33 (p=0.070)	-4.75 (p=0.104)	--	p=0.845 for composite test across 2 grades and 1 outcome
Key Components of Scientifically Based Reading Instruction at the School-Level				
• Amount of time dedicated to serving as K-3 reading coach	33.49* (p<0.001)			p=0.009 for composite test across 2 outcomes
• Availability of differentiated instructional materials for struggling readers	0.01 (p=0.661)			
Key Components of Scientifically Based Reading Instruction at the Classroom-Level (aggregated across all three grade levels)				
• Amount of PD in reading received by teachers	12.13* (p<0.001)			p<0.001 for composite test across 3 grades and 6 outcomes
• Teacher receipt of PD in the five essential components of reading instruction	0.55* (p=0.010)			
• Teacher receipt of coaching	0.20* (p<0.001)			
• Minutes of reading instruction per day	18.47 (p<0.001)			
• Provision of extra classroom practice for struggling readers	0.19 (p=0.018)			
• Use of assessments to inform classroom practice	0.18 (p=0.090)			

NOTES:

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.

EXHIBIT READS: The result of the composite test for reading comprehension test scores, across three grades and two outcomes, is not statistically significant (p=.957).

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; RFIS Student Time-on-Task and Engagement with Print, fall 2005, spring 2006, fall 2006, and spring 2007, and RFIS Survey administration, spring 2007.

level.[25] To qualify the six multiple hypotheses tests for these outcomes, the RFIS Team created one parsimonious index. The aggregation improves statistical power to detect effects that go in the same direction within a domain. The summary index is defined to be the equally weighted average of z-score outcome components, with the sign of each measure oriented so that more beneficial outcomes have higher scores.[26]

Specifically, the present analysis took the following steps in creating a composite index and conducting the analysis:[27]

1. First, z-scores were created for each outcome component in the reading comprehension domain by subtracting the unadjusted non-RF mean (pooled across years and grade levels) and dividing by its standard deviation (pooled across years and grade levels). Thus, each component of the index has a mean of zero and a standard deviation of one for the non-RF group.
2. If an observation unit has a valid response to at least one component measure of the index, then any missing values of other component measures are imputed as the random assignment group mean. This results in differences between RF and non-RF means of an index being the same as the average of those two groups' means of the components of that index (when the components are divided by their comparison group standard deviation and have no missing value imputation), so that the index can be interpreted as the average of results for separate measures scaled in standard deviation units.[28]
3. The z-scores from each component were averaged to obtain the index and an impact analysis was run on this index using a sample that pooled all years and all grade levels together.

This regression addresses the question whether overall the program "worked" in terms of improving student achievement. This result serves as a "qualifier" to the small number of specific hypothesis tests shown in impact tables.[29]

---

[25] Although decoding is considered to be in the same domain as comprehension, it was not possible to include the TOSWRF scores in the composite because scores are available for only one grade in one year.

[26] An alternative is to use seemingly unrelated regression effects for specific outcomes to estimate the covariance of the effects and then to calculate the mean effect size for groups of estimates in a second step. The average z-score index approach is much simpler to work with. The two approaches yield identical treatment effects when there is no item nonresponse and no regression adjustment (Kling, Liebman, and Katz, 2007).

[27] The discussion and method presented here draw from Kling, Liebman, and Katz (2007).

[28] No data imputation was done in constructing the reading achievement composite index.

[29] Though decoding is included in the student achievement domain, it is not possible to include this outcome measure in the summary index with the reading comprehension outcomes. As noted earlier, the decoding test was administered in only one grade for one year. In addition, 2,158 students who were administered the reading comprehension test in that year and grade were not given the decoding test. Therefore, attempts to calculate a common index for the decoding and reading comprehension measures would have required collapsing the reading comprehension data to a single year and grade and imputing missing decoding scores. This approach was not taken because this process would have resulted in a significant loss of statistical power and would have weakened the usefulness of the index as a qualifier for reading comprehension impacts.

Classroom Instruction

A similar composite analysis was conducted for the instructional domain. To qualify the impact estimates for each outcome measure for each grade in the instructional domain, the analysis ran a composite regression which pooled the sample across grades and used an index constructed from z-scores for all three instructional outcome measures as the dependent variable. The index of instruction averaged together minutes in the five dimensions of reading instruction, percentage of highly explicit instruction, and percentage of high quality student practice.[30]

The results from this analysis help to answer the research question whether *overall* the Reading First program has an impact on instructional practice.

In addition, program impacts for time spent on each of the five dimensions will be reported separately. Since the impact on total time spent on the five dimensions will already have been reported, any additional qualifying test is not necessary for these analyses.

Student Engagement with Print

A similar composite analysis was conducted for the student engagement with print outcome domain. For this domain impacts are reported for the full sample in grades 1 and 2 as the percentage of students engaged with print. To qualify the two multiple hypotheses tests for these outcomes, the RFIS Team reports the result from a composite regression which pools two grades together and represents the outcome measure in one parsimonious index, created in the same way that the composite index for reading comprehension and instruction was created.[31] This regression addresses the question whether overall the program "worked" in terms of having an impact on the percentage of students engaged with print. This result serves as a "qualifier" to the small number of specific hypothesis tests shown in impact tables.

Implementation of Key Components of Scientifically Based Reading Instruction (Surveys)

Because survey data was collected at the school level (i.e., reading coach and principal surveys) and at the classroom level (i.e., teacher surveys), two composite tests *across* domains were conducted, one at the school level and one at the classroom level. To qualify the impact estimates for each *school level* survey outcome measure, the analysis team ran a composite regression where the dependent variable was a single index created from z-scores for each of the two school level outcome variables. The index of key components of scientifically based research instruction at the school level averaged together the following

[30] No data imputation was done in constructing the reading instruction composite index.

[31] No data imputation was done in constructing the student engagement with print composite index.

two survey outcomes: amount of time dedicated to serving as K-3 reading coach and availability of differentiated instructional materials for struggling readers.[32]

To qualify the impact estimates for each *classroom* level survey outcome measure, the analysis team ran a composite regression where the dependent variable was a single index created from z-scores for each of the six classroom level outcome variables. The index of key components of scientifically based research instruction at the classroom level averaged together the following six survey outcomes: amount of PD in reading received by teachers, teacher receipt of PD in the five essential components of reading instruction, teacher receipt of coaching, minutes of reading instruction per day, provision of extra classroom practice for struggling readers, and use of assessments to inform classroom practice.[33]

These regression analyses address the question whether overall the program "worked" in terms of having an effect on the implementation of key components of scientifically based reading instruction, in the school and in the classroom. These results serve as a "qualifier" to the small number of specific hypothesis tests shown in impact tables.

Part 4: Statistical Precision

The statistical precision of an impact estimator is its ability to detect true intervention effects when they exist. A common way to represent statistical precision is a minimum detectable effect. This measure indicates the smallest true effect that an estimator has a "good chance" of detecting. The current analysis uses the common convention of defining a minimum detectable effect as the smallest true program effect (impact) that has an 80 percent chance of being found to be statistically significant (i.e., it has 80 percent statistical power) at the 0.05 level of statistical significance for a two-tailed test of the null hypothesis of no effect. When a minimum detectable effect is expressed as a standardized effect size (in standard deviation units), it is usually referred to as a minimum detectable effect size (MDE).

Exhibit B.8 lists the minimum detectable effect (or effect size) for full-sample estimates of program impacts on key study outcomes when the data are pooled across the school years for which data are available. These minimum detectable effects are based on the experience of students and schools in the study sample during the follow-up period, and not on the initial assumptions that guided the study design. Hence, the findings in Exhibit B.8 represent the actual precision of the present design as it materialized in the field.[34]

---

[32] Six schools were dropped from these analyses because both survey outcomes were missing. All other schools had a valid response on at least one of the survey outcomes. Missing values on individual survey outcomes ranged from 0% to 7% overall (RF: 0% to 1% and non-RF: 0% to 14%). Missing values were imputed as the random assignment group mean.

[33] All classrooms had a valid response on at least one of the six survey outcomes. Missing values on individual survey outcomes ranged from 0% to 3% overall (RF: 0% to 2% and non-RF: 0% to 5%). Missing values were imputed as the random assignment group mean.

[34] Because for the present full sample the number of degrees of freedom for estimating the standard error of an impact estimator is well beyond 30, the minimum detectable effect of an estimator equals 2.8 times its standard error. For further discussion see 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.

Exhibit B.8: Minimal Detectable Effects for Full Sample Impact Estimates

	Grade Level		
	Grade 1	Grade 2	Grade 3
Panel 1			
Student Achievement			
Reading Comprehension			
Mean Scaled Score	7.62	6.41	5.93
Effect Size	0.16	0.15	0.15
Percent at or above Grade Level	7.22	6.67	6.70
Decoding			
Mean Standard Score	3.14	--	--
Effect Size	0.21	--	--
Panel 2			
Instructional Outcomes			
Instruction in the Five Dimensions Combined			
Minutes	6.84	6.88	--
Effect Size	0.33	0.32	--
Percentage of Intervals in Five Dimensions with Highly Explicit Instruction			
Percentage	3.88	4.06	--
Effect Size	0.22	0.21	--
Percentage of Intervals in Five Dimensions with High Quality Student Practice			
Percentage	3.50	3.49	--
Effect Size	0.21	0.19	--
Panel 3			
Student Engagement with Print			
Percentage of Students Engaged with Print	8.18	8.15	--
Effect Size	0.28	0.28	--

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.

Minimal detectable effects are based on the standard errors of the impact estimates for the full sample pooled across three school years (except for Student Engagement with Print, which is based on two years of data) 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 data.

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.

EXHIBIT READS: The minimal detectable effect of the Reading First program on reading comprehension for a mean scaled score in grade 1 is 7.62 scaled score points. The minimal detectable effect of the Reading First program on reading comprehension for a mean scaled score in grade 2 is 6.41 scaled score points. The minimal detectable effect of the Reading First program on reading comprehension for a mean scaled score in grade 3 is 5.93 scaled score points.

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 TOSWRF administration in spring 2007; RFIS Instructional Practice in Reading Inventory, spring 2005, fall 2005, spring 2006, fall 2006, and spring 2007; and RFIS Student Time-on-Task and Engagement with Print, fall 2005, spring 2006, fall 2006 and spring 2007.

The three panels in the exhibit present minimum detectable effects for the Tier 1 outcomes of the present study. The three columns in the exhibit present minimum detectable effects for grades one, two, and three separately.

The top panel focuses on measures of student reading comprehension and decoding. Findings in this panel indicate that the present study design and impact estimation model have minimum detectable effects for reading comprehension that range from approximately 6 to 8 scaled score points, which corresponds to 0.15 to 0.16 standard deviations or about 7 percentage points. The minimal detectable effect for decoding was approximately 3 scaled score points, which corresponds to 0.21 standard deviations. These findings indicate that the present study achieved its goal of providing minimum detectable effect sizes that are no larger than 0.20 standard deviations for estimates of the impacts of Reading First on student reading comprehension.[35]

Findings in the second panel of the exhibit indicate that the minimum detectable effect for instructional time spent in the five dimensions of reading instruction in grades 1 and 2 is approximately 7 minutes, which corresponds to 0.32 to 0.33 standard deviations.

Minimum detectable effects for the percentage of instructional intervals in the five dimensions that exhibited highly explicit instruction or that exhibited high quality student practice ranged from approximately 3 to 4 percentage points. The minimum detectable effect on the percentage of students engaged with print was between 10 and 11 percentage points, roughly twice as large as that for the preceding two measures.

On balance, the statistical precision of the present study design and its analytic framework achieve the initial goals of the study's design. The precision is adequate for full-sample impact estimates, which are the primary focus of the present study.

Part 5: Handling Missing Data

This section describes how the study handled missing data for each outcome measure and for covariates used in analyses.

Surveys

The study imputed values for several survey variables. When Reading Coach survey data were missing at the item-level, and the identical questions had been asked of school principals, principal data were used when available. The study used principals' responses for 13 schools (all non-Reading First) without Reading Coach surveys on the "availability of differentiated instructional materials for struggling readers" outcome. The study also imputed values of 0 for the "amount of time dedicated to serving as K-3 reading coach" for six (non-Reading First) schools that the study had confirmed did not have reading coaches. The study imputed data from other RF schools about the presence of a scheduled reading block in grades K-3 and about the length of the K-3 reading block when all other RF schools in the site had reported having a reading block and that the block was the same length (n=2). In one instance, where all other RF

[35] See Gamse et al. (2004).

schools in a site reported using the same core reading program, the study imputed the core reading program for that school using data from other RF schools in that site.

Missing data rates ranged from 0.1 to 3.3 percent for teacher survey outcomes (RF: 0.1 to 1.0 percent; non-RF: 0.0 to 4.9 percent) and 1.3 to 2.8 percent for reading coach and/or principal survey outcomes (RF: 0.0 to 1.6 percent; non-RF: 2.7 to 4.1 percent). Survey constructs (i.e., those outcomes comprised of multiple survey items) were constructed only for observations with complete data, with one qualification: the construct "minutes spent on reading instruction per day" was calculated as the total number of minutes reported for the previous week (across a maximum of 5 days) divided by the number of days with non-missing values. For this construct, teacher surveys with missing data for every day of the previous week were eliminated (0.9 percent).

Classroom Observations: IPRI

No imputations were made for observations that were missing in their entirety. Imputations were made for a small proportion of intervals within observations during data cleaning. When an individual observation record (comprised of 35 successive intervals, on average) contained gaps in time or blank intervals, the record was filled in to make observation interval times internally consistent, using verbatim information provided by the observer on classroom activities. When no such information was available, gaps in time were filled in with intervals coded as non-instructional. When an observation contained more than one blank interval, and the same instructional activity (at the dimension or Part B levels) had been recorded in the intervals immediately before and after the blank interval, the blank intervals were post-coded to be the same as the preceding and successive intervals' activities. If the surrounding intervals were not identical to each other, the blank intervals remained blank. In the pooled sample (spring 2005, fall 2005, spring 2006, fall 2006, spring 2007), imputations were made in 402 classroom observations (2.8 percent). Of those 402 observations, 205 were in Reading First classrooms and 197 were in non-Reading First classrooms, which corresponds to 2.9 percent of all Reading First classrooms and 2.8 percent of non-Reading First classrooms.

Classroom Observations: STEP

No data imputation was done for the STEP. Missing data were handled in the following manner. If classrooms had missing values for all three sweeps, they were dropped from the analytic sample. A sweep was coded as missing if the class was in transition between activities or the entire class was listening to a story. Only 78 classrooms (1 percent) were given a missing code data for all three sweeps. Of the 78 classrooms, 23 were Reading First and 55 were non-Reading First, which corresponds to 1 percent of Reading First classrooms and 2 percent of non-Reading First classrooms in the pooled analytic sample. In cases where classrooms had one or more non-missing values for the three sweeps, the percentage of students engaged with print was calculated by averaging across the number of sweeps available for that classroom. For the pooled analytic dataset (fall 2005, spring 2006, fall 2006, and spring 2007), 70 percent of classrooms had three sweeps of data; 23 percent had two sweeps of data; 5 percent had one sweep of data; and 1 percent were missing all three sweeps.[36]

[36] Numbers do not add to 100% due to rounding.

Student Reading Achievement: SAT 10 Reading Comprehension Subtest

No data imputation was done for the SAT 10. Students test scores were coded as missing and excluded from the analytic sample if they were deemed invalid according to SAT 10 scoring guidelines. For the pooled sample (Spring 2005, Spring 2006, Spring 2007), this amounted to 222 student test scores (0.2 percent). Of the missing scores, 92 were Reading First and 130 were non-Reading First, which corresponds to 0.1 percent of Reading First test scores and 0.2 percent of non-Reading First scores.

Data were imputed for three covariates used in student achievement analyses—student age, gender, and date of testing. If missing, student age and gender were imputed by using school-by-grade means, except when data on an entire grade within a school was missing, in which case the district-by-grade mean was used. When an imputed covariate was used in analysis, a dummy variable indicating the imputed observations was also included. For student age, values were imputed for 0.35 percent (0.19 percent Reading First and 0.53 percent non-Reading First) of the pooled sample. For gender, values were imputed for 0.36 percent (0.30 percent Reading First and 0.43 percent non-Reading First) of the pooled sample. For the date of testing covariate, which was recorded at the classroom level, missing values were imputed using the school mean. Values were imputed for 0.79 percent (1.44 percent Reading First and 0.06 percent non-Reading First) of the pooled sample.

Student Reading Achievement: TOSWRF

No data imputation was done for the TOSWRF. Students were coded as missing and were excluded from the analytic sample if birth dates were missing, or out of range (meaning that a standard score could not be computed for those students), or if students did not follow test instructions. Of the 465 test scores (5 percent) excluded from the analytic sample, 233 were Reading First and 232 were non-Reading First, which corresponds to 4 percent of Reading First scores and 4 percent of non-Reading First scores in the analytic sample.

Covariates used in the TOSWRF analyses were imputed in the same manner as described above for the SAT 10 analyses.

Appendix C: Measures

Appendix C, Parts 1 through 5, describe the data collection instruments and assessments used to create measures for all outcome domains assessed in the RFIS. These instruments and assessments include Reading Coach and Teacher Surveys, the Instructional Practice in Reading Inventory (IPRI), the Global Appraisal of Teaching Strategy (GATS), Student Time-on-Task and Engagement with Print (STEP), and assessments of students' reading performance (SAT 10 and TOSWRF). Parts 1 through 5 also include relevant information on properties of instruments, data collection procedures, and response rates.

Appendix C, Part 6 contains copies of each of the RFIS data collection instruments.

Part 1: Reading Coach and Teacher Surveys

Description of the Instruments

The RFIS developed survey instruments for reading coaches and classroom teachers in grades 1 through 3 to learn about how schools were implementing scientifically based reading programs. The surveys for the RFIS and the Reading First Implementation Evaluation[1] contain identical items in order to facilitate comparisons between the purposive RDD sample of the current study and the nationally representative sample of Reading First schools and personnel surveyed by the Implementation Evaluation.

The Reading Coach Survey targeted school level individuals designated as reading coaches, reading specialists, literacy coordinators or Title I/resource teachers. It included items on the coach's background and experience, core and supplemental reading materials, professional development offered to grades 1–3 teachers, specific coaching activities, characteristics of reading instruction in the school, changes that have taken place in reading instruction, and areas needing improvement. See Appendix C, Part 7, Exhibit C.17 for a copy of the Reading Coach Survey.

The Teacher Survey addressed student characteristics, reading instruction (e.g., materials, content, time allocation), assessment, interventions for struggling readers, participation in reading-related professional development, and collaboration and support from other teachers and staff. The surveys provide self-reported information on the instructional emphases across grades. See Appendix C, Part 7, Exhibit C.18 for a copy of the Grade 2-3 Teacher Survey.

Administration Procedures and Response Rates

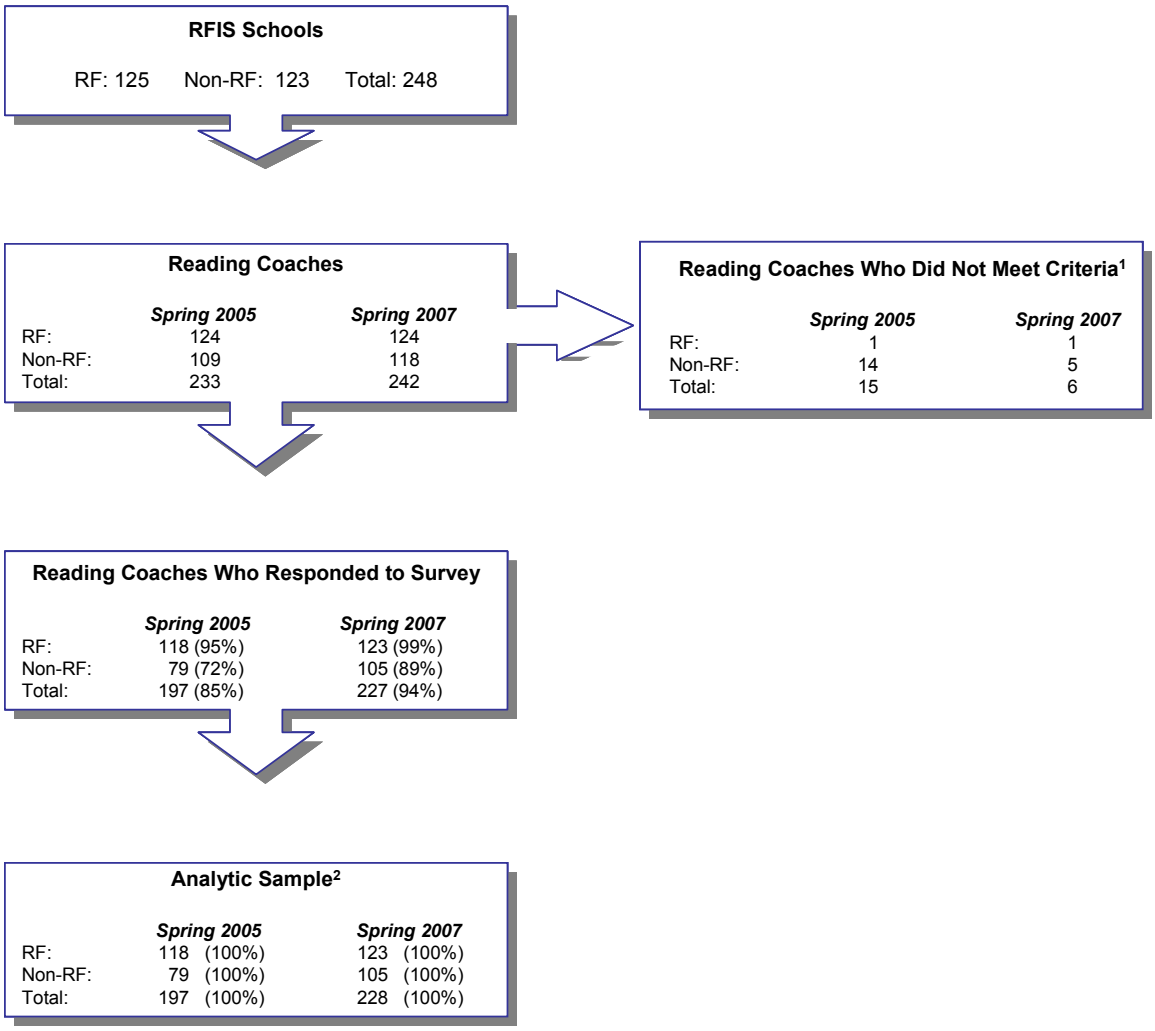
Surveys were mailed in March 2007 to building level study liaisons who then distributed sealed envelopes to an average of nine classroom teachers (three per grade level), the school's reading coach, and the building principal. Respondents were asked to return the completed surveys within two to three weeks. Follow-up was conducted to encourage potential respondents to complete and return the surveys.

[1] The Reading First Implementation Evaluation, commissioned by the Policy and Program Studies Service at the U.S. Department of Education, collected survey data from principals, teachers and reading coaches in nationally representative samples of RF schools and non-RF Title I schoolwide project (SWP) schools in the 2004–05 and 2006–07 school years.

Exhibit C.1 provides the response rate and sample size for each survey. In 2007, 227 reading coaches and 1,792 teachers completed a survey. The effective Reading Coach Survey response rate was 99 percent for RF schools and 89 percent for non-RF schools. The effective Teacher Survey response rate was 87 percent for RF schools and 83 percent for non-RF schools.

Exhibit C.1: Survey Data Collection: School, Reading Coach, and Teacher Sample Information

Reading Coach Survey



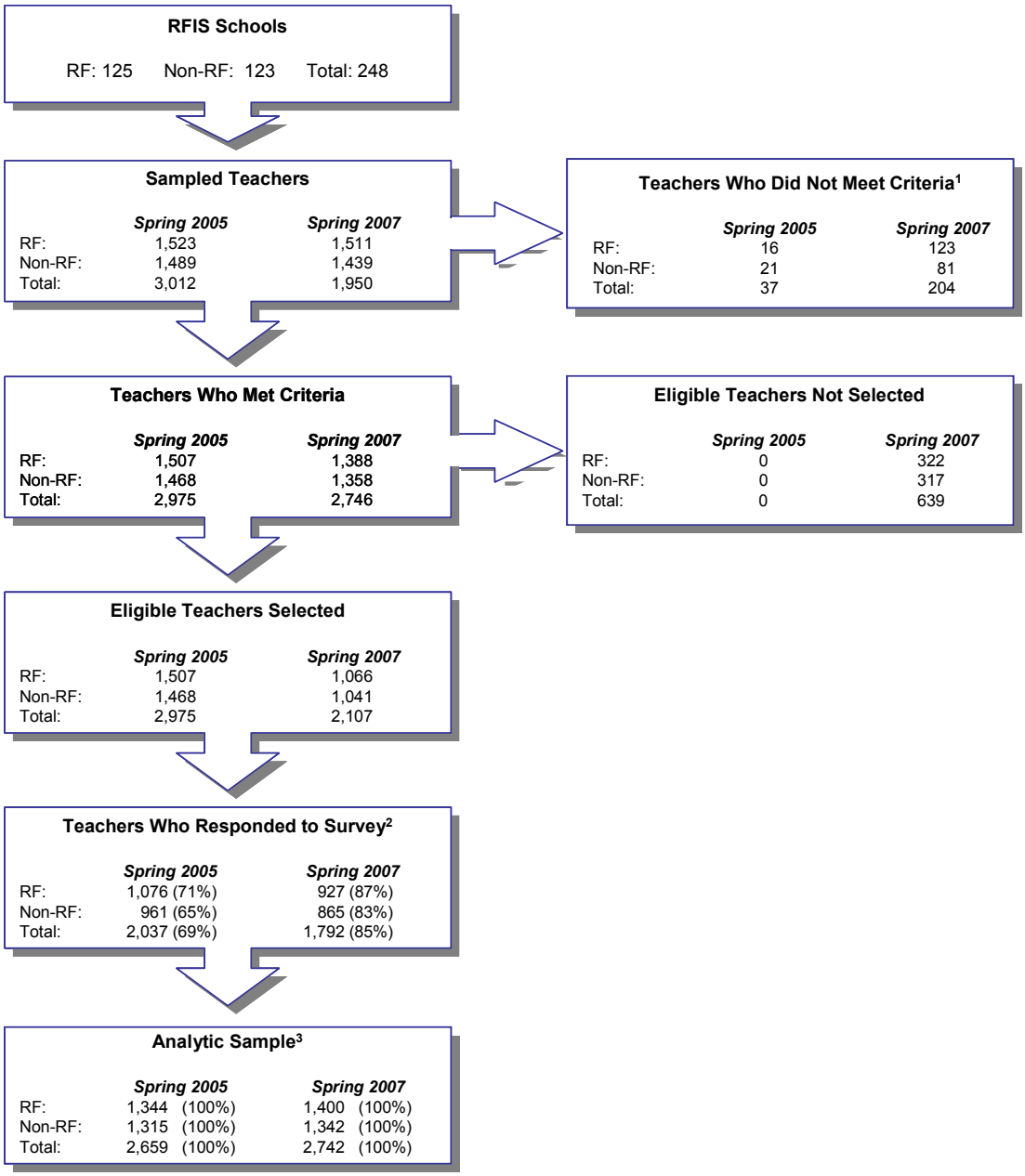
[1] Reading coach respondents who did not meet criteria included individuals who indicated that they did not serve any school(s) as a reading coach who provided ongoing training and support to school staff in delivering effective reading instruction.

[2] All completed surveys were used in the analytic sample. Information on item-level response rates is presented on tables where applicable, and Appendix B, Part 5 describes the overall approach to handling missing survey data.

SOURCE: Reading First Impact Study Reading Coach Surveys, spring 2005 and spring 2007.

Exhibit C.1: Survey Data Collection: School, Reading Coach, and Teacher Sample Information (continued)

Teacher Survey



NOTES:

[1] Respondents who did not meet criteria or were not selected included student teachers, substitute teachers, or teachers whose classrooms were not observed (for grades 1 and 2) or tested (grades 1, 2, and 3).

[2] A total of 23 teachers (15 in 2004-05, 8 in 2006-07) returned surveys but were dropped because they indicated that they did not teach reading or grades 1, 2, or 3.

[3] All completed surveys were used in the analytic sample. Information on item-level response rates is presented on tables where applicable, and Appendix B, Part 5 describes the overall approach to handling missing survey data.

SOURCE: Reading First Impact Study Teacher Surveys, spring 2005 and spring 2007.

Composition, Scale, Internal Consistency and Scientifically Based Research Support

Exhibit C.2 reports the composition, metric, specifications, and internal consistency of the survey outcomes. Exhibit C.3 includes information on the Reading First legislation and guidance that supports the survey outcomes.

Part 2: Classroom Instruction: The Instructional Practice in Reading Inventory (IPRI)

Background

To measure the impact of Reading First on classroom instruction, the RFIS team conducted classroom observations in both Reading First and non-Reading First (non-RF) classrooms. The primary instrument used to evaluate instruction was the Instructional Practice in Reading Inventory (IPRI). The RFIS Team was unable to identify an existing observational instrument that fulfilled all of the study requirements; consequently, the RFIS Team developed the IPRI specifically for the RFIS. The IPRI is designed to measure first- and second-grade teachers' use of instructional behaviors informed by scientifically based reading research (SBRR), as described in the National Research Council's report (Snow, Burns, and Griffin, 1998) and the National Reading Panel report (National Institute of Child Health and Human Development, 2000). In particular, the IPRI focuses on instruction in the five dimensions of reading instruction emphasized by SBRR (phonemic awareness, decoding/phonics, fluency, vocabulary, and comprehension). Exhibit C.4 gives specific examples of instructional activities associated with each of the five dimensions.

The development of the IPRI relied on several sources, including (1) research on the components of effective elementary grade reading instruction (e.g., Kamil, 2004; National Institute of Child Health and Human Development, 2000; Snow, Burns and Griffin, 1998; Stahl, 2004); (2) reviews of existing instruments (among the instruments reviewed were the following: *The Instructional Content Emphasis (ICE)* [Edmonds and Briggs, 2003]; *Foorman and Schatschneider direct observation system and instruments from the Center for Academic and Reading Skills (CARS)* [Foorman and Schatschneider, 2003]; *English Language Learner Classroom Observation Instrument (ELLCOI)* [Haager et al., 2003]; *Teachers' Instructional Practice (TIP)* [Carlisle and Scott, 2003]; *Utah's Profile of Scientifically based Reading Research* [Dole, et al., 2001]; *The Classroom Observation Record* [Abt Associates and RMC Research, 2002]; and *Observation Measure of Language and Literacy Instruction (OMLIT)*, developed by Abt Associates as part of the Even Start Classroom Literacy Interventions and Outcomes (CLIO) Study [Goodson et al., 2004]); and (3) research on the development of classroom observation instruments (Vaughn and Briggs, 2003).[2]

[2] For a comprehensive description of the development of the IPRI, see Dwyer et al., 2007.

Exhibit C.2: Composition, Metric, Specifications, and Internal Consistency of Survey Outcomes

Domain Survey Outcome …..Individual Items Comprising the Outcome (as applicable)	Survey, Item(s)	Metric	Outcome Specifications[1]	Internal Consistency[2]
Professional Development (PD) in SBRI				
1. Amount of PD in reading received by teachers …..Attended short, stand-alone training or workshop in reading (half-day or less) …..Attendance longer institute or workshop in reading (more than half-day) …..Attended a conference about reading (might include multiple short offerings)	Teacher, C1: a, b, d	Hours	Sum of hours across 3 items	0.22
2. Teacher receipt of PD in the five essential components of reading instruction …..In phonemic awareness …..In phonics …..In vocabulary …..In fluency …..In comprehension	Teacher, C4: a-s	0-5 scale	Each component (e.g., fluency) was scored dichotomously (1=teacher received PD in at least one of the topics listed, 0= teacher did not receive PD in any of the topics listed). Sum of 5 dichotomously scored components (1=addressed, 0=not addressed).	0.86
3. Teacher receipt of coaching	Teacher, C2: a	Proportion	Dichotomous variable (1=teacher received assistance from reading coach, 0=teacher did not receive assistance from reading coach or not available)	N/A
4. Amount of time dedicated to serving as K-3 reading coach	Reading Coach, B3	Percent	N/A	N/A
Amount of Reading Instruction				
5. Minutes spent on reading instruction per day …..Last week, approximately how many minutes per day did you devote to reading instruction? Reported separately for each of the five weekdays.	Teacher, B1	Minutes	Total number of minutes of reading instruction for last week divided by number of days with non-missing values.	0.99
Supports for Struggling Readers				
6. Availability of differentiated instructional materials for struggling readers[3] …..Use separate program materials in interventions …..Use core reading program with supplemental materials …..Use core reading program only …..Use reading materials written in ELLs' home language …..Use alternative materials designed for ELLs	Reading Coach/ Principal, E1: a-e	Proportion	Dichotomous variable (1=E1: a, b, d, or e are available for struggling readers, 0=only the core reading program is available for struggling readers, E1: c)	N/A

Exhibit C.2: Composition, Metric, Specifications, and Internal Consistency of Survey Outcomes (continued)

Domain Survey Outcome …..Individual Items Comprising the Outcome (as applicable)	Survey, Item(s)	Metric	Outcome Specifications[1]	Internal Consistency[2]
7. Provision of extra classroom practice for struggling readers (over the past month) …..Extra practice in the classroom with phonemic awareness …..Extra practice in the classroom with phonics …..Extra practice in the classroom with fluency …..Extra practice in the classroom with comprehension	Teacher, B8: b-e	0-4 scale	Sum of 4 dichotomously scored items (1=received, 0=did not receive)	0.77
Use of Assessments				
8. Use of assessments to inform classroom practice …..Use test results to organize instructional groups …..Use tests to determine progress on skills …..Use diagnostic tests to identify students who need reading intervention services	Teacher, B5: u, w, y	0-3 scale	Sum of 3 dichotomously scored items (1=central, 0=small or not part of reading instruction)	0.60

NOTES:

[1] Missing data rates ranged from 0.1 to 3.3 percent for teacher survey outcomes (RF: 0.1 to 1.0 percent; non-RF: 0.0 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).

[2] Internal consistency was calculated using Cronbach's raw alpha for survey outcomes other than single dichotomous outcome variables.

[3] This survey item was asked of both reading coaches and principals. In cases where reading coach survey data were not available, the study used principal responses for those schools (n=13).

EXHIBIT READS: The outcome variable "amount of PD in reading received by teachers" consisted of three individual items from the RFIS Teacher Survey, item C1. The sum of these three items represents the total the number of hours of professional development in reading attended by teachers in the form of short trainings/workshops, longer institutes/ workshops, and conferences. The internal consistency reliability was 0.22.

Sources: Reading First Impact Study Reading Coach, Principal, and Teacher Surveys.

Exhibit C.3: Reading First Legislative Support and Guidance for Survey Outcomes

Domain **Survey Outcome** …..Individual Items Comprising the Outcome (as applicable)	**Survey, Item(s)**	**Reading First Legislation[1] or Guidance[2]**
Professional Development (PD) in SBRI		
1. Amount of PD in reading received by teachers …..Attended short, stand-alone training or workshop in reading (half-day or less) …..Attendance longer institute or workshop in reading (more than half-day) …..Attended a conference about reading (might include multiple short offerings)	Teacher, C1: a, b, d	Legislation, Section 1202(c)(7)(A)(iv) Guidance, p. 7, C-3
2. Teacher receipt of PD in the five essential components of reading instruction …..In phonemic awareness …..In phonics …..In vocabulary …..In fluency …..In comprehension	Teacher, C4: a-s	Legislation, Section 1202(c)(7)(A)(iv)(I) Guidance, p. 7, C-3
3. Teacher receipt of coaching	Teacher, C2: a	Legislation, Section 1202(c)(7)(A)(iv)(III) Guidance, p. 7, C-3
4. Amount of time dedicated to serving as K-3 reading coach	Reading Coach, B3	Legislation, Section 1202(c)(7)(A)(iv)(III) Guidance, p. 7, C-1, C-3
Amount of Reading Instruction		
5. Minutes spent on reading instruction per day	Teacher, B1	Guidance, p. 7, C-1
Supports for Struggling Readers		
6. Availability of differentiated instructional materials for struggling readers …..Use separate program materials in interventions …..Use core reading program with supplemental materials …..Use core reading program only …..Use reading materials written in ELLs' home language …..Use alternative materials designed for ELLs	Reading Coach/ Principal, E1: a-e	Guidance, p. 7, C-2
7. Provision of extra classroom practice for struggling readers …..Extra practice in the classroom with phonemic awareness …..Extra practice in the classroom with phonics …..Extra practice in the classroom with fluency …..Extra practice in the classroom with comprehension	Teacher, B8: b-e	Legislation, Section 1202(c)(7)(A)(ii)(II)(ee)

Exhibit C.3: Reading First Legislative Support and Guidance for Survey Outcomes (continued)

Domain Survey Outcome …..Individual Items Comprising the Outcome (as applicable)	Survey, Item(s)	Reading First Legislation[1] or Guidance[2]
Use of Assessments		
8. Use of assessments to inform classroom practice …..Reading assessments are used to screen students for reading difficulties …..Diagnostic assessments are used to identify strengths and weaknesses of struggling readers …..Reading assessments are used to monitor student progress	Teacher, B5: u, w, y	Legislation, Section 1202(c)(7)(A)(i) Guidance, p. 7, C-1, C-2, C-4

NOTES:

[1] The legislation for Reading First is contained in the No Child Left Behind Act of 2001, ESEA, 2001, Title 1, Part B, Subpart 1.

[2] Guidance for the Reading First Program is provided by the U.S. Department of Education (2002).

EXHIBIT READS: The outcome variable "amount of PD in reading received by teachers", which consists of three individual items from the RFIS Teacher Survey (Question C1), was supported by both the Reading First legislation [No Child Left Behind Act of 2001, ESEA, 2001, Title 1, Part B, Subpart 1, Section 1202(c)(7)(A)(iv)] and the Guidance for the Reading First program (U.S. Department of Education, 2002, p. 7, C-3).

Sources: Reading First Impact Study Reading Coach, Principal, and Teacher Surveys.

Exhibit C.4: Examples of Instruction in the Five Dimensions of Reading Instruction

Phonemic awareness	The teacher is working with a group of four students. The teacher says, "Listen to me. The word is ***hat***. If I take away the /h/ sound at the beginning, I have the word ***at***. Then if I add a /b/ sound to the beginning I get ***bat***. Now you try. The word is ***sat***. If we take away the /s/ sound what word do we have?" [students respond orally]. "That's right, ***at***. Now add a /k/ sound to the beginning. What word? That's right, ***cat***."
	The teacher is working with a pair of students. He asks students to identify the final sound in each of a list of 10 words. The students respond orally to each prompt from the teacher: "***Crack***. What's the last sound in ***crack***? [students respond orally]. Good. Ok: ***Take***. What's the last sound? [students respond orally]. Ok, next: ***kite***. What's the last sound? [students respond orally]. How about ***flight***? [students respond orally]. That's right, /t/, /t/ is the last sound in ***flight***. "
Phonics	A group of 16 students has assembled in front of the classroom blackboard. The teacher writes the letters ***oi*** on the board and says, "Ok, now today we're going to be learning about words that have o, i in them. When you see these vowels together, they make the /***oy***/ sound. Here's an example." The teacher writes a sentence on the board: ***I want Roger to join my club.*** She underlines the letters oi in the word join. "This word is ***join***. 'I want Roger to ***join*** my club. See that ***oi***? What sound does ***oi*** make?" [students respond, some of them incorrectly]. "Ok, listen carefully. Not ***/eye/***… no, ***oi*** makes the ***/oy/*** sound. Everyone try that: ***/oy/***." [students in unison say /oy/]. "Ok, good, now what's this word [she points to **join**]?" The students pronounce **join** correctly. "Excellent, ok, let's try another one." She writes the word ***coin*** on the board. "Boys and girls, look at that ***oi*** in the word. Sound out this word for me."
	Six students are seated with a teacher. Each student has a set of individual magnetic letters and a metal tray. The teacher is asking students to form words that she dictates orally: "Ok, listen to the word, think about the sounds and what letters go with those sounds. Remember that we've been working with the ***/ō/*** sound and its spellings. We know that one way to spell that is with ***o, a***. Try to make the word using your letters. The first word is ***goat***. Use your letters to make the word ***goat***." Students assemble their letters and the teacher checks each student's work. "Good. Everyone used ***o, a*** to spell ***goat***. Ok, let's try another word: ***float***." Students form the word with their letters. "Ok, good! You're doing very well. Now, we also know another way to spell some words with the long ***/ō/*** sound. Remember the silent e rule? It makes the vowel say its name. So, to spell the word ***tote***, Arthur, tell me how we'd write ***tote***?"
Vocabulary	The teacher gives a definition for the word **swift** and uses it in a sentence: "Swiftly? Something that is swift is moving very fast, rapidly. So, remember when we learned about how fast cheetahs can run over land? Well, we might say, 'the cheetah ran ***swiftly*** across the ground, quickly catching up to the tiger.'"
	As they are reading a story in class, students come across the word **debating**, and the teacher discovers that they do not know what it means. The teacher defines **debating** by contrasting it with more familiar words (**chatting** and **talking**). The teacher says, "When two people are **debating** something, it means that they are talking about the reasons to do something and the reasons not to do something—so in our story, John and Sara are **debating** whether or not to go on a picnic. On the one hand, the weather is nice, but on the other hand they are thinking there may be a lot of ants. So they're **debating** what to do. **Chatting** is different than **debating**. When you're **chatting** with someone, you're usually not trying to decide something, you're just talking about things that aren't too serious. You chat more to enjoy the talking, not really to decide something together."

Exhibit C.4: Examples of Instruction in the Five Dimensions of Reading Instruction (continued)	
Fluency	Roberto is reading orally from a passage about parrots and their habitat. When he reaches the end of the second paragraph, the teacher asks Roberto to read that same passage aloud again. When Roberto finishes, the teacher asks him to read the passage out loud a third time.
	The teacher assigns four students to pairs and distributes a page-long excerpt from a story they have been reading in class that week. Each pair of students also has a one minute timer. "Ok, now you each have a partner, and I want you to time your partner reading this passage out loud. Readers, you try to read as far as you can in one-minute. Timers, you keep track of the time and tell your partner to stop reading when time runs out. Then circle the last word the reader got to in the passage."
Comprehension	A teacher pauses in the middle of a story about Shackleton's Antarctic Voyage to ask students to reflect on what they have just read and draw some inferences about how one character might be feeling. "What do you think the captain is feeling? Let's see. The story doesn't tell us exactly, but the story says the ship is starting to break apart. I'd certainly be very worried for myself and my crew if my ship were breaking apart! I bet the captain is really worried. Let's see… the story also says the captain 'furrowed his brow.' That means he made his forehead wrinkle or sort of frown. Some people do that when they're worried. That could be a sign that the captain is worried. He certainly has reason to be worried."
	The teacher introduces a comprehension strategy. "One thing you should always do when you read is constantly ask yourself questions about the story. Asking yourself questions is a strategy to help make sure you understand what you just read. Asking questions also helps you think about what might happen next. We're going to practice using this strategy. At the end of every paragraph today, we're going to come up with some questions and write them up here on the board. Some questions we'll be able to answer right away. But we might have other questions, too, and we'll need to read more of the story before we can find out how to answer those questions."

Overview of the IPRI

The IPRI observation instrument is a booklet containing a series of individual IPRI forms, each of which corresponds to a three-minute observation interval.[3] Observation data for a given reading block are collected via sequentially-ordered IPRI forms that span the entire observation period (e.g., a 60-minute observation would be recorded on 20 sequential forms, one for each successive three-minute interval). During each three-minute interval, observers record any of the teacher's instructional behaviors listed on the IPRI that occur during that interval. At the end of each three-minute interval (signaled by a pre-programmed vibrating wristwatch), observers turn to a new IPRI form and begin another three-minute interval, again recording the presence of targeted behaviors.

Within a given three-minute interval, a particular behavior is coded only once, regardless of how often that behavior occurs within an interval. Recurrences of that same behavior are coded in each subsequent interval. If behavior ***x*** occurs in interval ***n***, the observer circles the code for behavior ***x*** once during interval ***n***. If behavior ***x*** occurs in the next interval, ***n+1***, the observer circles the code for behavior ***x*** during interval ***n+1***.

[3] See Exhibit C.19 for a copy of the IPRI instrument.

Structure of the IPRI Instrument

Each IPRI form has four distinct parts: Part A, Part B, Part C, and Part D. Part A is divided into five color-coded sections that correspond to the five dimensions of reading instruction: phonemic awareness, decoding/phonics, fluency, vocabulary, and comprehension, respectively. Within each of these five sections are microcodes, specifically tailored to each of the five dimensions, which denote the following areas of interest:

- the size of the student grouping to which instruction is delivered;
- the use of any instructional support materials (e.g, manipulatives, pictures);
- the teacher's use of explicit instruction;
- the teacher's provision of practice opportunities for students; and
- the teacher's delivery of any corrective feedback or expansion of student responses.

For example, within the phonemic awareness row, the IPRI microcodes for grouping are "whole class, large group, small group, pair, or individual"; for the use of various types of instructional supports, "teacher manipulative or kinesthetic, student manipulatives, kinesthetics"; and for corrective feedback, "teacher pinpoints what student(s) did incorrectly with sound(s) and gives correct response with or without students." For the use of explicit instruction and the provision of practice opportunities for students, these areas of interest are often denoted by the combination of two or more microcodes. So, for example, if a teacher "demonstrates or models oral blending or segmenting with phonemes" *in conjunction with* "gives student(s) chance to practice oral blending or segmenting with phonemes," it would be counted as explicit instruction.

Part B of the IPRI contains codes to capture instruction or other activity outside the five dimensions, including:

- Oral reading by students;[4]
- Oral reading by teacher alone (without student accompaniment);
- Silent reading;
- Spelling;
- Written expression;
- Other language arts;
- Assessment;
- Non-literacy instruction;
- Non-instruction;
- Academic management;
- Transitions between activities;
- Interruptions to instruction for the purpose of managing student behavior.

[4] Oral reading under Part B is marked when the teacher has not clearly indicated the instructional purpose of the oral reading. If, however, oral reading is used to advance instruction in one of the five targeted dimensions of reading instruction (e.g., comprehension), then the oral reading is coded within the corresponding row in Part A of the IPRI.

Part C records teachers' instructional errors that are not subsequently self-corrected. Part D records whether the teacher worked with a different small group of students than in any previous part of the observation.[5]

Training and Inter-rater Reliability of Classroom Observers

Prior to each wave of data collection, field staff based in each of the RFIS sites attended a centralized, multi-day training on the IPRI and associated data collection protocols. The training curriculum included extensive practice coding a series of videotaped clips of real-time and unscripted classroom instruction that were filmed in RF and non-RF classrooms. The film clips were created specifically for the RFIS, and were edited to illustrate the codes included on the IPRI. Candidate observers conducted a live observation in a first or second grade classroom during the training session and received ongoing feedback, multiple opportunities for review, tutoring and other support throughout the training.[6]

One component of this training was that observers were required to pass two of three formal inter-rater reliability tests; each videotape used for reliability purposes was approximately 30 minutes in length. To calculate observers' percent agreement with the master coding of each reliability tape, the RFIS Team used a procedure that reduces inflation in inter-rater reliability estimates due to chance agreement (see Kelly, 1977, cited in Suen and Ary, 1989). The inflation due to chance agreement is especially severe when some events (or codes) occur infrequently, as is the case with the IPRI.[7] As a result, observers were credited only for codes that occurred at least once in the reliability tape. In sum, if a behavior occurred at all during a 30-minute tape, observers were credited (or penalized) for correctly coding instances of the behavior and for correctly abstaining from coding behaviors that did not occur. Observers were not credited for abstaining from, nor penalized for, marking behaviors that never occurred throughout the entire reliability tape.

For each potential observer, percent agreement with the master codes was calculated for each code individually; then agreement was aggregated across codes within the five sections in Part A and across codes within Part B. Finally, an aggregate overall percentage agreement across the five sections in Part A and codes within Part B was calculated. A report summarizing all of these measures of agreement (by individual code, by dimension, and overall) was prepared for each potential observer so that s/he (and the study team) could diagnose which codes had proven particularly troubling. Overall percent agreement was used to judge whether or not each observer had met the criterion for employment on the study. Only observers who successfully coded two of three videotaped reliability tests were hired. The mean overall percent agreement for observers was 88 percent (n=155 observers) in spring 2005 (for spring 2005 data collection). The mean overall percent agreement for observers was 90 percent (n=154 observers) in fall 2005 (for fall 2005 and spring 2006 data collection). The mean overall percent agreement for observers was 90 percent (n=130 observers) in fall 2006 (for fall 2006 and spring 2007 data collection).

---

[5] Minor changes were made to the IPRI after the spring 2005 data collection and prior to the fall 2005 wave of data collection; these changes included elaborating upon some micro-behaviors within each of the five dimensions.

[6] For a detailed description of the classroom observer training, see Dixon et al. (2007).

[7] During each observation interval, an IPRI form contains 142 possible codes; typically, only a small subset of the behaviors occur during a given interval. Thus, most of the possible codes are infrequent within a single interval. Including all 142 codes per interval in the calculation of percent agreement severely inflates inter-rater reliability.

Data Collection

Observations were conducted in first- and second-grade classrooms for two consecutive days during each classroom's designated reading block. During the 2004-05 school year, the RFIS conducted two days of classroom observation in spring 2005. In the following study year (2005-06), a second round of observations was added, so that observers conducted observations for two consecutive days in the fall, and then again for two consecutive days in the spring. Again in 2006-07, observers conducted observations for two consecutive days in the fall, and then again for two consecutive days in the spring. The increased number of observations reflects a decision by the National Center for Education Evaluation/Institute of Education Sciences at the Department of Education to collect more data, both in terms of the number of observations and in terms of when during the year data could be collected.

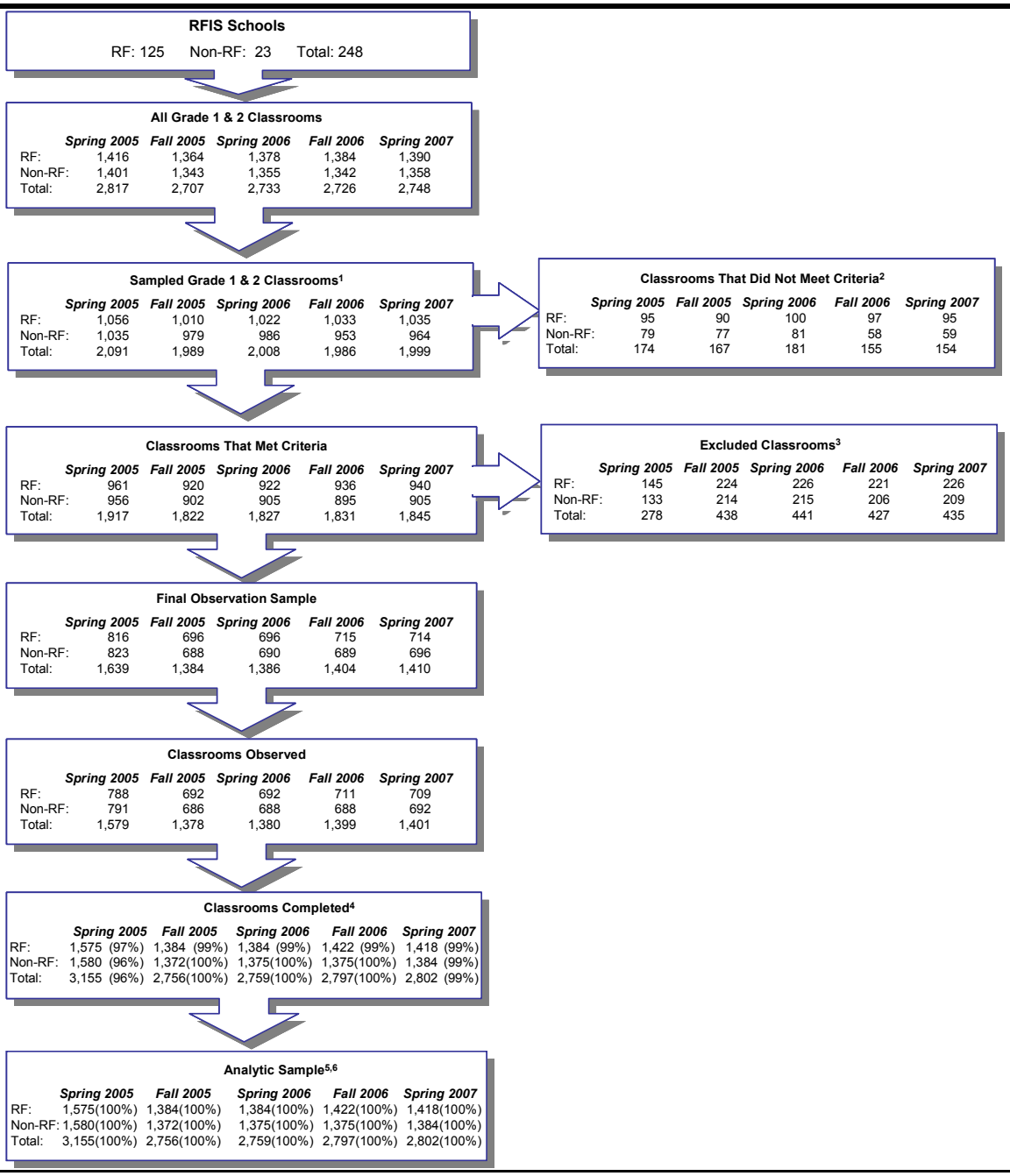
Observation scheduling was arranged by RFIS field supervisors via communication with each participating school's study liaison.[8] Observers coded during the entire scheduled observation period, even when teachers appeared to be offering non-reading-related instruction. In those instances when reading instruction appeared to continue beyond the scheduled reading block, observers observed for up to an additional 30 minutes. Throughout observations, IPRI observers followed the actions and behaviors of classroom teachers. In classrooms with more than one adult present, observers determined beforehand who was the official teacher of record and which adult would be delivering that day's reading instruction. The individuals responsible for delivering instruction were then followed for the observations whether or not they were the official teacher of record. Observations were rescheduled when the classroom teachers were absent or ill, although long-term substitutes replacing a teacher on an extended leave of absence (e.g., maternity, disability) were observed.

The 248 schools in the RFIS study sample allowed for observations in 2,091 classrooms in 2004-2005, 3,997 classrooms in 2005-2006, and 3,985 classrooms in 2006-2007. Of these, 1,917 classrooms met eligibility requirements for classroom observations in 2004-2005, 3,649 in 2005-2006, and 3,676 in 2006-2007. Classrooms were considered eligible to be in the study sample if they were not special education or English as a Second Language classes, if more than 75 percent of the students were in the target grades, and if the class was taught by the regular teacher or a long-term substitute.

Of the eligible classrooms, the RFIS selected a final observation sample of 1,639 classrooms in 2004-2005, 2,770 in 2005-2006, and 2,814 in 2006-2007. Classrooms were sampled within schools, if, within each site as a whole, the number of classrooms exceeded an average of three classrooms per grade. Each classroom in the sample was expected to be observed two times during each of the three waves of data collection. The RFIS completed 96 percent of the expected classroom observations in 2004-2005, and 100 percent in both 2005-2006 and 2006-2007. A flow chart of information on the RFIS IPRI sample and response rates is presented in Exhibit C.5.

[8] In schools that did not have a designated "reading block," the RFIS Team asked the school's study liaison when observers would be able to see typical reading, literacy, and/or language arts instruction in classrooms. In cases where reading instruction was delivered in two discrete blocks interrupted by other instruction or activities (e.g., lunch, recess, math instruction), field staff observed both blocks.

Exhibit C.5: IPRI Data Collection: School, Classroom, and Observation Sample Information



Notes:

1 The study conducted observations in all classrooms in schools if across the site as a whole, the average number of classrooms per grade per school was three or less. If for that site as a whole, the average number of classrooms per grade per school exceeded three, the study sampled classrooms for observations within schools.

2 At the beginning of each data collection wave, the study team contacted schools to obtain classroom rosters and indications of which classrooms were designated as regular classrooms, self-contained special education classrooms, or ESL classrooms. Those classrooms identified as special education or ESL were considered ineligible for observation.

3 Once the study team began to schedule and conduct observations, classrooms were excluded from the sample if the information that had been provided earlier about special education or ESL status was inaccurate, if the class was taught by someone other than the regular teacher or a long-term substitute, or if fewer than 75 percent of the students were in the target grade.

4 Each classroom was observed twice during each wave of data collection. Response rates are calculated by dividing the observations completed by two times the number of eligible classrooms selected into the sample.

5 All IPRI observations were included in the analytic sample. Percentages are calculated by dividing the analytic sample numbers by the number of observations completed.

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

During each data collection wave (spring 2005, fall 2005, spring 2006, fall 2006, and spring 2007), IPRI experts (from the training staff) served as quality control monitors for questions that arose in the field. Quality control monitors visited each site and accompanied a random selection of observers into scheduled classroom observations. The monitors reviewed the observation coding with the observers, addressing coding discrepancies and questions.[9] Throughout the data collection period, observers could direct questions to the monitors and to other RFIS staff. Questions and answers were aggregated and disseminated to all observers via an RFIS observer website and regular mailings.

Creation of Analytic Variables

To test whether or not instruction in RF classrooms differed from that in non-RF classrooms, the study team created eight measures of classroom instruction from the IPRI data. The number of measures was deliberately limited so that the analysis would be parsimonious, and would thereby restrict the number of statistical tests required. The measures were:

- Time spent in instruction in each of the five targeted dimensions of reading instruction separately:
  - phonemic awareness;
  - phonics/decoding;
  - vocabulary;
  - fluency;
  - comprehension;
- Time spent in instruction in the five dimensions combined;
- Proportion of instruction in the five dimensions that was highly explicit—that includes teacher modeling, clear explanations, and the use of examples;
- Proportion of instruction in the five dimensions that provided students with high quality practice opportunities—that includes, for example, teachers giving students the opportunity to practice word learning strategies (e.g., context, word structure, and meanings).

Before describing these measures in more detail, we first describe the transformation of raw interval data into more meaningful metrics.

Transformation of IPRI Observation Intervals Into Minutes

The IPRI contains multiple successive three-minute intervals, each of which could potentially record a large number of instructional behaviors, if the behaviors had indeed been observed. Each behavior on the IPRI is deemed to have occurred or not occurred in each observed interval (e.g., behavior was present [checked or coded] or not [unchecked]). Across the entire set of intervals comprising a classroom observation, the IPRI yields raw data in terms of the number (or proportion) of observed intervals in which a given behavior was observed. The raw data do not directly measure the duration of particular instructional activities or behaviors. In order to describe classroom instruction with a more interpretable metric, raw intervals were transformed into minutes of instruction via the process described below.

[9] Study protocols required observers to leave as is any codes marked during the observation. This procedure allowed the RFIS study team to collect a sample of paired observations for use in determining field-based reliability.

For each and every interval, observers recorded instruction in one of the five dimensions (hereafter referred to as "dimensions")[10] or in other activity/instruction not in one of the five dimensions (hereafter referred to as "non-dimension activities"). These latter activities are included in "Part B" described above. Consequently, every observation interval contains *at least* one of the following codes that categorizes the types of instruction the teacher provided during that interval:

- Phonemic awareness;
- Phonics/decoding;
- Vocabulary;
- Fluency;
- Comprehension;
- Oral reading by children;[11]
- Oral reading by teacher;
- Silent reading;
- Spelling;
- Written expression;
- Other language arts;
- Assessment;
- Non-literacy instruction;
- Non-instruction;
- Academic management; and/or
- Transitions between instructional activities.

The allocation of time within the three-minute intervals occurred at the broader level—that is, at the level of dimension and non-dimension activities. When only one dimension or non-dimension activity was observed in an interval, the conversion process was straightforward—all three minutes of the interval were assigned to the dimension or non-dimension activity observed.

When *two* activities were recorded in an interval, however, the process of converting intervals into minutes was less straightforward. From the raw data, there was no direct way to determine the proportion of the three minute interval that the teacher had devoted to each of the two recorded activities. Therefore, the study team developed an estimation process to allocate minutes of that interval to each of the two activities. The RFIS collected supplemental data on the actual duration of instructional activities recorded on the IPRI, and used those supplemental data to inform mathematical simulations of the outcomes of different estimation procedures.

---

[10] For purposes of calculating minutes of instruction in a particular dimension, the micro-level codes corresponding to aspects of instruction *within* each of the five dimensions were collapsed. For example, a teacher who had exhibited two different "phonics" codes within an interval was designated as having delivered phonics instruction within that interval.

[11] Note that the IPRI distinguishes oral reading for its own sake from oral reading in service to a larger instructional purpose. For example, oral reading that occurred to advance a lesson in comprehension was classified as being part of the overarching comprehension instruction and was not counted as oral reading for purpose of analysis. In contrast, oral reading that occurred outside the context of one of the five dimensions of reading instruction was classified for analytic purposes as Oral Reading.

Dividing the minutes of the interval equally. Initially, the RFIS Team considered allocating one-half of the three-minutes of an interval to each of the two activities observed. Under this procedure, if comprehension and phonics were observed in the same interval, for instance, then each would be assigned 1.5 minutes of the three-minute interval. Although this approach provides a good estimate of the true number of minutes spent in the two activities for intervals in which the two observed activities were of similar duration, for intervals in which activities were of unequal duration, however (e.g., one activity was 2.6 minutes and the other .4 minutes), this approach underestimates the amount of time in the longer activity and overestimates the amount of time spent in the shorter one.

Dividing the minutes of the interval according to their relative frequency of occurrence. The study team also explored an estimation method that allocates time to each of two activities within a given interval in direct proportion to the relative frequency with which the two activities occurred, on average, within the school in which the observation had been conducted. If, on average, comprehension was present in 30 percent of the intervals collected across all observations within a school, whereas fluency instruction was present in 10 percent of all intervals collected in the school, then comprehension was three times as likely to occur as fluency instruction. Then for each interval in which comprehension and fluency were the two activities recorded, comprehension would receive 75 percent of the three minutes (or 2.25 minutes) and fluency would receive 25 percent of the three minutes (one-third the amount of time as comprehension, or .75 minutes).

The RFIS Team used supplemental data on the true duration of instructional activities to simulate the precision of this estimate. The simulations suggested that the proportionally-weighted approach provided a close estimate of the true minutes spent in activities for intervals in which two activities were of unequal duration, but, conversely, it produced biased estimates of the true minutes spent in activities for intervals in which the two activities observed were of similar duration. Thus, the strengths and drawbacks of this approach were mirror opposites of those in the first approach (i.e., dividing the minutes equally among the two activities in an interval).

The RFIS Team decided that an average of the two estimations would minimize the biases introduced by using either of the two transformation approaches in isolation.

Dividing the minutes of the interval by taking the average of the equally and proportionally weighted approaches. For each interval with two instructional activities recorded, a three-step estimation process was used:

1. The minutes were allocated *equally* between the two activities (1.5 minutes to each).
2. The minutes were allocated *according to their relative frequency* of occurrence across all observations within school.
3. The *average* of the two estimates produced was calculated for each of the two activities.

Using the example cited above (an interval with only comprehension and fluency instruction, comprehension would be allocated 1.88 minutes, or the mean of the equally weighted and proportionally weighted approach [1.5 and 2.25, respectively]). Fluency would be allocated 1.12 minutes (the mean of 1.5 and .75 minutes).

Three or more activities occurring in the same interval. When three or more instructional activities were observed in a single interval, the three minutes of the interval were divided equally among the activities. This distribution strategy was followed rather than the estimation process used for two-activity intervals because the number of minutes assigned to any given activity type would be limited to one minute or less. Thus, the amount of bias introduced by using this estimation approach was likely to be small.

Analytic Variables

The study team constructed six variables based on the amount of time devoted to instruction in the five dimensions of reading instruction: one variable for the amount of time spent in each of the five dimensions separately, plus a sixth variable for the total amount of time spent in the five dimensions combined.

Also of interest were the degree to which instruction in RF and non-RF schools was highly explicit, and the degree to which instruction offered students meaningful opportunities to practice developing reading skills. To examine these outcomes, two additional variables were constructed: the percentage of instruction in the five dimensions in which at least one instance of highly explicit instruction occurred; and the percentage of instruction in the five dimensions in which at least one instance of high quality student practice occurred. These two variables are defined below.

Percentage of intervals of instruction in the five dimensions that included at least one instance of highly explicit instruction. "Highly explicit instruction" is defined differently in each dimension of reading instruction, based on research published in the National Research council report (Snow, Burns, and Griffin, 1998) as well as more recent research (e.g., Graves, Gerston and Haager, 2004; Gunn et al., 2002 for specific examples of highly explicit instruction in phonemic awareness, and Foorman and Torgesen, 2001, Graves, Gerston and Haager, 2004, for specific examples of highly explicit instruction in phonics). Exhibit C.6 lists the specific citations for examples of highly explicit instructional strategies for each of the five components of reading instruction targeted by the legislation.[12] The specific instructional strategies, or combinations of strategies used together, that were considered to be "highly explicit" are presented in Exhibit C.6. This variable was created by dividing the number of intervals that included one or more "highly explicit" instructional practices by the number of intervals that included instruction in one or more of the five dimensions.

Percentage of intervals of instruction in the five dimensions that included at least one instance of high quality student practice. "High quality student practice" is also defined differently in each dimension of reading instruction, based on research published in the National Reading Panel report (National Institute of Child Health and Human Development, 2000) as well as more recent research (e.g., Armbruster, Lehr and Osborn, 2003 for specific examples of high quality student practice in phonemic awareness, and Rasinski and Oswald, 2005, for specific examples of high quality student practice in phonics). Exhibit C.6 lists the specific citations for examples of high quality student practice for each of the five dimensions of reading instruction targeted by the legislation. The specific instructional strategies,

[12] No codes in the fluency dimension were classified as "highly explicit" instruction. Helping beginning readers build fluency inherently rests on providing students high quality practice opportunities, rather than delivering explicit instruction in how to read fluently. As a result, codes in the fluency section were used only in the construction of the high quality student practice variable.

Exhibit C.6: Composite of Classroom Constructs

Minutes spent in instruction in each of the five dimensions of reading instruction

Number of minutes spent in any teacher instruction or student practice activity on the IPRI that was in the five dimensions of reading instruction emphasized in Reading First:

- Phonemic awareness
- Phonics/decoding
- Vocabulary
- Fluency
- Comprehension
- All five dimensions combined

Percentage of observation intervals with instruction in the five dimensions of reading instruction with one or more instance of highly explicit instruction

An observation interval was coded as containing instruction in the five dimensions of reading instruction and at least one instance of highly explicit instruction if one or more of the following teacher activities (or combination of activities) was observed during instruction in one of the four reading dimensions that included highly explicit instructional activities.

Phonemic Awareness:[13]

- Teacher demonstrates or models oral blending or segmenting with phonemes *in conjunction with*:
  - Giving students practice in oral blending or segmenting with phonemes
- Teacher demonstrates or models phoneme isolation *in conjunction with*:
  - Giving students practice in phoneme isolation
- Teacher demonstrates or models phoneme categorization/identity (same/different sounds in words) *in conjunction with*:
  - Giving students practice in phoneme categorization/identity
- Teacher demonstrates or models phoneme deletion, addition, or substitution *in conjunction with*:
  - Giving students practice in phoneme deletion, addition, or substitution
- Teacher contrasts two phonemes to pinpoint a target sound
- Teacher pinpoints what students did incorrectly and gives correct response

Phonics/decoding:[14]

- Teacher identifies words that contrast with or do not follow pattern or rule
- Teacher reminds students of pattern or rule and has students produce or repeat correct response, if a student makes a mistake
- Teacher describes, explains, or identifies, or asks students to describe, explain, or identify a sound-symbol pattern, decoding rule, or a word structure pattern or rule *in conjunction with*:
  - Showing students how to apply a rule or pattern to a whole word example, *and*
  - Giving students chance to practice decoding words
- Teacher describes, explains, or identifies, or asks students to describe, explain, or identify a sound-symbol pattern, decoding rule, or a word structure pattern or rule *in conjunction with*:
  - Showing students how to apply a rule or pattern to a whole word example, and
  - Giving students practice encoding words by manipulating or writing letters

[13] Ball and Blachman (1991); Bus and van Ijzendoorn (1999); Foorman et al. (1998); Graves et al. (2004); Gunn et al. (2002); Hatcher et al. (2004); McCutchen et al. (2002); Torgesen et al. (1999).

[14] Foorman et al. (1998); Foorman and Torgesen (2001); Graves et al. (2004).

Exhibit C.6: Composite of Classroom Constructs (continued)

Highly explicit instruction

Vocabulary:[15]

- Teacher goes beyond synonym with definition and/or examples
- Teacher pinpoints word meaning by giving contrasting examples
- Teacher pinpoints word meaning by clarifying or extending a partially correct student response
- Teacher pinpoints word meaning by clarifying or extending a partially correct student response with a synonym, definition, example, or contrasting example

Teacher uses a picture, object, or physical demonstration to illustrate word meaning in conjunction with any other vocabulary instructional behaviors including those above and the following:

- Teacher asks students to give meaning of word
- Teacher gives synonym
- Teacher asks students to apply understanding of word meaning
- Teacher gives students opportunity to practice word learning strategies (e.g., using context, word structure, or root meanings)

Comprehension:[16]

Before, during, or after reading a text passage, teacher describes or explains, or asks students to describe or explain one or more comprehension strategies by specifying:

- What the comprehension strategy is called, *and*
- Why the comprehension strategy is helpful, *and*
- When in the reading process the comprehension strategy is used

During or after reading a text passage, teacher shows how to apply strategy by modeling how to:

- Answer inferential questions based on text
- Make predictions based on text
- Summarize, retell, sequence text, or identify the main idea(s)
- Make text-to-text connections
- Generate own questions about text
- Answer own questions about text
- Review passage to check or clarify understanding
- Check accuracy of prediction or inference
- Work with story or expository structure

If a student response is incorrect or incomplete, teacher assists student in using strategy(ies)

[15] Brett et al. (1996); Graves et al. (2004); Kamil (2004); McKeown et al. (1985); Tomesen and Aarnoutse (1998).

[16] Crowe (2005); Kamil (2004); Mason (2004); O'Connor et al. (2002); Rosenshine et al. (1996).

Exhibit C.6: Composite of Classroom Constructs (continued)

Percentage of observation intervals with instruction in the five dimensions of reading instruction with one or more instance of high quality student practice

An observation interval was coded as containing instruction in the five dimensions of reading instruction *and* at least one instance of high quality student practice if one or more of the following teacher activities (or combination of activities) was observed during instruction in one of the five dimensions.

Phonemic Awareness:[17]

- Teacher gives students practice in oral blending or segmenting with phonemes while working with pairs or small groups
- Teacher gives students practice in phoneme isolation while working with pairs or small groups
- Teacher gives students practice in phoneme categorization/identity (same/different sounds in words) while working with pairs or small groups
- Teacher gives students practice in phoneme deletion, addition, or substitution while working with pairs or small groups

Phonics/Decoding:[18]

- Teacher gives students practice encoding words by manipulating or writing letters

Vocabulary:[19]

- Teacher gives students the opportunity to practice word learning strategies (e.g. context, word structure, and root meanings)

Fluency:[20]

- Teacher gives students the opportunity to repeat oral readings with same text that was modeled by a fluent reader

Comprehension:[21]

During or after reading a text passage, teacher gives students practice in applying strategy by having students:

- Generate own questions about text
- Answer own questions about text
- Review passage to check or clarify understanding
- Work with story or expository structure *in conjunction with*:
  - Using a text organizer for support
- Check accuracy of prediction or inference
- Justify their response with evidence

[17] Ambruster et al. (2003); National Institute of Child Health and Human Development (2000).

[18] Rasinski and Oswald (2005).

[19] Ambruster et al. (2003); National Institute of Child Health and Human Development (2000).

[20] Graves et al. (2004); O'Connor et al. (2002); Stahl (2004); Therrien (2004).

[21] Kamil (2004); Mason (2004); Reutzek and Hollingsworth (1991); Taylor et al. (2002).

or combinations of strategies used together, that were considered to be "high quality student practice" are presented in Exhibit C.6. This variable was created by dividing the number of intervals that included one or more instance of "high quality student practice" by the number of intervals that included instruction in one or more of the five dimensions.

Field Reliability of the IPRI

In each wave of data collection, experienced IPRI trainers were paired with a random sample of classroom observers to collect data necessary to measure the field-based reliability of the IPRI.[22] In contrast to determining the accuracy of an individual observer for purposes of training and hiring, the purpose of field-based inter-rater reliability (IRR) estimates is to assess the reliability of the instrument itself. Researchers often characterize the reliability of an observation instrument by estimating an intra-class correlation (ICC), defined here as the proportion of variance associated with observers relative to the total variance in the collected data. That is, the team sought to characterize the proportion of variance in the observation data due to each of three sources:

- inter-observer differences
- inter-classroom differences
- random measurement error

The RFIS Team used several approaches to attempt to capture the degree of error that can be attributed to observers themselves (as opposed to random measurement error or other forms of systematic measurement error). These approaches included: (1)(a) calculating a pseudo intraclass correlation (ICC) by running an unconditional Hierarchical Linear Model (HLM), and (b) correlating Observer A's and Observer B's codes across multiple intervals within an observation and then averaging these correlations across pairs of observers, and (2) calculating a generalizability coefficient within the generalizability framework (Brennan, 2001; Cronbach et al., 1972 as cited in Brennan, 2001; and Shavelson and Webb, 1991).

Using a Pseudo Intraclass Correlation to Describe Inter-rater Reliability

In the context of measuring inter-rater reliability of the IPRI based on paired field observations, consider the following model:

$$X_{cr} = \mu + \nu_c + \nu_r + \nu_{cr} \qquad (1)$$

In (1), X_{cr} is the outcome measure for classroom *c*, as rated by observer *r*; μ is the mean outcome across classrooms; and ν_c, ν_r, and ν_{c4} are independent error terms associated with the variance across classrooms, systematic measurement error introduced by the observers, and random measurement error; each with a mean of 0 and variances of σ_c, σ_r, and σ_{cr}. Using this model, we can define the proportion of the total measurement variance that is due to the systematic measurement error introduced by the observers ρ_1 and the proportion of the true variance across classrooms ρ_2 as follows:

[22] Half of the field observers were paired with co-observers in spring 2005. In the second year, field observers were paired with co-observers either in fall 2005 or in spring 2006; the majority of field observers were paired for observation once during the 2005-06 school year. In the third year, approximately half of the observers were paired with co-observers either in fall 2006 or spring 2007.

$$\rho_1 = \frac{\sigma_r^2}{\sigma_c^2 + \sigma_r^2 + \sigma_{cr}^2} \quad (2)$$

$$\rho_2 = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_r^2 + \sigma_{cr}^2} \quad (3)$$

(2) indicates the proportion of error that can be attributed to variation across individual observers (observers may vary in their skill at using the IPRI). An examination of (3) shows that to the extent that variance attributable to observers (σ_r^2) is low, the proportion of variance due to true variance across classrooms is high (assuming that random measurement error is small); as ρ_1 decreases, ρ_2 increases. Thus, the lower the ICC, the higher the reliability of the IPRI.

Ideally, intra-class correlations are calculated using a fully crossed design, such that each of a set of *R* observers observes each of C classrooms. In a fully-crossed design, the variance associated with individual observers can be estimated separately from the systematic error associated with individual classrooms. However, a fully-crossed design was not possible in the context of the RFIS, which used 150 observers to record instruction in approximately 1,400 classrooms during each round of data collection. Instead, joint observations were conducted in a sample of classrooms by two observers, one a master observer and the other a member of the field staff. No individual observed more than a small subset of the total number of classrooms. Thus, these data do not allow separate estimate variation due to rater or classroom alone.

The RFIS Team obtained pseudo-ICC estimates using field IRR samples as if they were fully crossed. Such estimates provide a biased estimate of the actual error due to observers, because they also include some of the error associated with inter-classroom differences; however, the estimates are conservative, attributing *more* error to observers than they would in a fully-crossed design. Therefore, if the pseudo-ICC estimates of inter-observer error are low, despite the fact that they include error associated with the individual classrooms, we can be confident that the true amount of error due to differences between observers is even lower—and thus that the IPRI is a reliable instrument.

Most study classrooms were jointly observed for about 30 three-minute intervals, although some joint observations covered fewer and others covered more intervals. In order to construct a fully balanced sample, for each wave of the field IRR samples, the study team (i) dropped classrooms that were observed for fewer than 25 intervals; and (ii) included only the first 25 observation intervals from classrooms that were observed for more than 25 intervals.[23] As a result, the reliability was calculated with 65 classrooms from spring 2005, 62 classrooms from fall 2005, 36 classrooms from spring 2006, 24 classrooms from fall 2006, and 37 classrooms from spring 2007 data collections to assess field-based IRR. (See Exhibit C.7.)

For each of the analytic variables created from IPRI data, the team calculated reliability estimates by estimating the variance terms in equations 2 and 3 (σ_c, σ_r, and σ_{cr}) and by running an unconditional HLM with the field IRR samples for each observation wave. Each HLM was a two-level model with observer

[23] The 25-interval threshold attempts to balance two sometimes competing constraints: (i) minimizing the number of classrooms that would be dropped due to lack of observations and (ii) maximizing the number of observation intervals that could be used to assess IRR.

Exhibit C.7: Unconditional HLM Models to Estimate Pseudo-ICCs (ρ1) and True Variance Across Classrooms (ρ2)

	Spring 2005 (n=65)		Fall 2005 (n=62)		Spring 2006 (n=36)		Fall 2005 (n=24)		Spring 07 (n=37)	
Outcome	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2
Number of Minutes Spent on Phonics	0.046	0.930	0.025	0.959	0.059	0.914	0.175	0.774	0.149	0.805
Number of Minutes Spent on Comprehension	0.049	0.927	0.079	0.888	0.025	0.959	0.080	0.888	0.020	0.966
Number of Minutes Spent on Vocabulary	0.038	0.941	0.067	0.904	0.049	0.926	0.027	0.956	0.107	0.854
Number of Minutes Spent on Phonemic Awareness	0.111	0.849	0.25	0.684	0.030	0.952	0.072	0.897	0.040	0.938
Number of Minutes Spent on Fluency Building	0.170	0.779	0.069	0.901	0.075	0.893	0.031	0.951	0.019	0.968
Number of Minutes Spent on Five Dimensions Combined	0.061	0.912	0.096	0.868	0.058	0.915	0.080	0.888	0.024	0.961
Proportion of Intervals in the 5 Dimensions Containing Highly Explicit Instruction	0.281	0.654	0.327	0.604	0.375	0.551	0.045	0.933	0.395	0.530
Proportion of Intervals in the 5 Dimensions Containing High Quality Student Practice	0.265	0.670	0.274	0.662	0.303	0.632	0.135	0.822	0.439	0.482

NOTE:

The HLM model utilized for this analysis includes an intercept and three independent random error terms that are associated with the variance across classes, systematic measurement error introduced by the raters, and random measurement error. Definitions of ρ_1 and ρ_2 can be found in the text.

EXHIBIT READS: The proportion of variance due to differences between observers for Number of Minutes Spent on Phonics was .046 for the 65 co-observed classrooms from spring 2005. The proportion of variance due to differences between classrooms for Number of Minutes Spent on phonics was .930 for the 65 classrooms from spring 2005.

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

(A or B) nested within classroom, for each classroom that had been co-observed. Next ρ_1 and ρ_2 were calculated using these estimates. Corresponding results are presented in Exhibit C.7 and indicate that ICC-based reliability estimates (ρ_2) are consistent across the five observation waves, ranging from 0.868 to 0.961, for example, for the number of minutes spent on the five dimensions combined.

An alternative way of obtaining a pseudo ICC estimate is by simply correlating the two observers' codes *within* a given observation and *across* the multiple intervals with that observation, and averaging these correlations across the pairs of coders. Similar to the unconditional HLM model, using this method with the co-observation data attributes more error to the observers than it should. This method is also complicated when one observer reports that a specific IPRI code never occurred during an entire observation, but the other observer reports that the same code occurred (at least once); in this case, the correlation coefficient is not defined (these observations were not included in this analysis). In contrast, if both observers agreed that a particular IPRI code never occurred, we imputed the correlation coefficient to be one since these cases could be regarded as perfect agreement. Exhibit C.8 presents estimates of this pseudo ICC with the number of observations used for the calculations. As expected, these results are very similar to the ones from the unconditional HLM model in Exhibit C.7.

Using a Generalizability Coefficient to Measure Inter-rater Reliability

Recall that the previous approach of using a pseudo-ICC to measure the field-based reliability of the IPRI assumes that the field IRR samples are fully crossed. One way to account for the fact that the field IRR samples are not fully crossed and still be able construct an estimate of field based reliability is to calculate a generalizability coefficient using the generalizability framework. The generalizability framework can be defined as a "theory that liberalizes classical theory by employing ANOVA methods that allow an investigator to untangle multiple sources of error" to describe the reliability of a measurement (Cronbach, et al., 1972, as cited in Brennan 2001.)

In field IRR samples, each classroom (c) is observed by a different set of two observers (or raters, [r]) simultaneously during a number of intervals (i). In the generalizability framework, discussed in detail by Brennan (2001), this set-up could be regarded as a **G study (r: c) * i** design with n_c that were observed by 2 observers (n_r=2) for 25 intervals (n_i=25).[24] The main and interaction effects for this model can be depicted as:

Let X_{cri} denote the outcome (an IPRI item) recorded in classroom *c* by rater *r* at interval *i*. Utilizing the effects presented in Exhibit C.9, we can describe this outcome as follows:

$$X_{cri} = \mu + \nu_c + \nu_i + \nu_{r:c} + \nu_{ci} + \nu_{ri:c} \qquad (4)$$

[24] Note that here an interval is regarded as the object of measurement.

Exhibit C.8: Average Correlation Between Paired Observers' Codes Across Classrooms

Outcome	Spring 2005 Average Correlation	Spring 2005 N[1]	Fall 2005 Average Correlation	Fall 2005 N[1]	Spring 2006 Average Correlation	Spring 2006 N[1]	Fall 2006 Average Correlation	Fall 2006 N[1]	Spring 2007 Average Correlation	Spring 2007 N[1]
Phonics	0.869	65	0.815	60	0.835	32	0.840	23	0.866	35
Comprehension	0.866	65	0.890	62	0.841	35	0.915	23	0.885	35
Vocabulary	0.829	65	0.836	60	0.811	34	0.816	24	0.836	36
Phonemic Awareness	0.990	55	0.976	60	0.963	34	0.859	23	0.942	35
Fluency Building	0.946	50	0.963	55	0.955	36	0.950	23	0.915	30
Five Dimensions Combined	0.845	65	0.836	61	0.807	35	0.845	24	0.864	37
Highly Explicit Instruction	0.579	63	0.649	57	0.590	35	0.705	22	0.688	36
High Quality Student Practice	0.679	60	0.764	52	0.710	24	0.820	23	0.821	28

Notes:

[1] The effective N is shown for the calculation of the average correlation between observer and co-observer codes. Co-observations in which *only* one of the observers reported that the outcome of interest occurred in every interval (or did not occur in any of the intervals) are excluded from the analysis as for such cases, the correlation coefficient could not be calculated. Co-observations in which both of the raters reported that the outcome of interest occurred in every interval (or did not occur in any of the intervals) are included in the analysis with a correlation coefficient of 1.

EXHIBIT READS: The average correlation between paired observers' codes across classrooms for phonics was .869 in spring 2005 (n=65), .815 in fall 2005 (n=60), .835 in spring 2006 (n=32), .84 in fall 2006 (n=23), and .866 in spring 2007 (n=35).

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

Exhibit C.9: Main and Interaction Effects in a (r: c)*i Design		
Model	**Main Effects**	**Interaction Effects**
(r: c) * i	***i, c, r:c***	***ci, ri:c***

Here, μ is the grand mean in the population and ν terms represent the five main and interaction effects listed in Exhibit C.9 (*i, c, r:c, ci, ri:c)*. Using (5), one can decompose the total variance observed in the outcome into five independent variance components associated with the effects as follows:

$$\begin{aligned}\sigma^2(X_{cri}) &= \sigma^2(\nu_c) + \sigma^2(\nu_i) + \sigma^2(\nu_{r:c}) + \sigma^2(\nu_{ci}) + \sigma^2(\nu_{ri:c}) \\ &= \sigma^2(c) + \sigma^2(i) + \sigma^2(r:c) + \sigma^2(ci) + \sigma^2(ri:c)\end{aligned} \quad (5)$$

Using this general framework, a measure of the IRR for a single *random* rater ($n_r = 1$) observing a single *fixed* classroom ($n_c = 1$) can be calculated using a **D-*study* (R:C) *** i design. This design is sufficient if one wants to estimate a general IRR across all possible pairs of raters, such that the correlation between a pair of raters estimates the reliability of a single rater, and it is not necessary to generalize across all classrooms. Under a D-study, the IRR estimate is given by the generalizability coefficient, $E\rho^2$, defined in equation (6):

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} = \frac{\sigma^2_{\mathrm{i}} + \sigma^2_{\mathrm{ci}}}{\sigma^2_{\mathrm{i}} + \sigma^2_{\mathrm{ci}} + \sigma^2_{\mathrm{ri:c}}} \quad (6)$$

In (6), $\sigma^2(\tau)$ and $\sigma^2(\delta)$ denote the universe score variance and variance of the relative error respectively. Exhibit C.10 demonstrates the formulas that could be used to calculate the variance components of the generalizability coefficient $E\rho^2$. Technically, $E\rho^2$ can be interpreted as an intra-class correlation coefficient, which approximates the expected value of the squared correlation between the observed outcome and the universe ("true") outcome for a classroom. In this context, the universe outcome can be defined as the expected value of the mean outcomes for every instance of the measurement procedure (i.e., the mean of the outcomes coded by all possible sets of two observers) of a classroom. Alternatively, $E\rho^2$ can also be seen as the ratio of variance of the universe outcome to the variance of the observed outcome. The difference between the pseudo ICCs described earlier and the generalizability coefficient $E\rho^2$ is that $E\rho^2$ takes into account the fact that each classroom was observed by a different set of two observers during co-observations, whereas the former simply ignores this fact.

Exhibit C.11 presents estimates of the generalizability coefficient calculated using the five waves of the IPRI field IRR data. These estimates of reliability are slightly lower than the reliability estimates determined by calculating pseudo ICC estimates shown in Exhibits C.7 and C.8. One possibility for these estimates being slightly lower is that the generalizability coefficient accounts for the fact that the sample is not fully crossed.

Overall, the various methods of estimating IRR using observation and co-observation data provide consistent results. The reliability estimates for the five dimensions (phonics, comprehension, vocabulary, phonemic awareness, and fluency building) are consistent across all methods. The estimates for highly explicit instruction and high quality student practice measures are lower, a finding that might reflect the fact these measures attempt to capture micro behaviors that are harder for observers to recognize and code accurately.

Exhibit C.10: Calculating Variance Components for a (r: c)*i Design

α	df(α)	T(α)	SS(α)	MS(α)	$\hat{\sigma}^2(\alpha) \equiv \hat{\sigma}^2_\alpha$
i	n_i-1	$n_c n_r \sum_i \overline{X}_i^{\,2}$	T(i)- T(μ)	$\frac{SS(\alpha)}{df(\alpha)}$	$\frac{MS(i) - MS(ci)}{n_c n_r}$
c	n_c-1	$n_i n_r \sum_c \overline{X}_c^{\,2}$	T(c)- T(μ)		$\frac{MS(c) - MS(r:c) - MS(ci) + MS(ri:c)}{n_i n_r}$
r: c	$n_c(n_r-1)$	$n_i \sum_c \sum_r \overline{X}_{r:c}^{\,2}$	T(r: c)- T(c)		$\frac{MS(r:c) - MS(ri:c)}{n_i}$
ci	$(n_c-1)(n_i-1)$	$n_r \sum_c \sum_i \overline{X}_{ci}^{\,2}$	T(ci)-T(c)-T(i)+ T(μ)		$\frac{MS(ci) - MS(ri:c)}{n_r}$
ri: c	$n_c(n_r-1)(n_i-1)$	$\sum_c \sum_r \sum_i X_{cri}^{\,2}$	T(ri: c)- T(ci) – T(r:c)+ T(c)		$MS(ri:c)$

Notation:

α: any of the main and interaction effects

df(α): degrees of freedom for effect α

T (α): sum of squared mean scores for effect α

$T(\mu) = n_r n_c n_i \overline{X}^2$

SS (α): sum of squares for α

MS (α): mean squares for α

$\hat{\sigma}^2(\alpha)$: estimated variance component for effect α

X_{cri} : outcome of interest for class *c* as rated by rater *r* in interval *i*

Exhibit C.11: Generalizability Coefficients Estimated from the Co-Observation Data

Outcome	Spring 2005 (n=65) $E\rho^2$	Fall 2005 (n=62) $E\rho^2$	Spring 2006 (n=36) $E\rho^2$	Fall 2006 (n=24) $E\rho^2$	Spring 2007 (n=37) $E\rho^2$
Phonics	.859	.820	.807	.830	.852
Comprehension	.863	.881	.820	.911	.869
Vocabulary	.812	.769	.796	.821	.835
Phonemic Awareness	.802	.822	.792	.723	.582
Fluency Building	.706	.826	.827	.875	.683
Five Dimensions Combined	.841	.843	.799	.848	.886
Highly Explicit Instruction	.577	.610	.545	.691	.668
High Quality Student Practice	.625	.574	.443	.751	.551

EXHIBIT READS: The generalizability coefficients for Phonics are .859 for spring 2005, .820 for fall 2005, .807 for spring 2006, .830 for fall 2006, and .852 for spring 2007.

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

Part 3: Global Appraisal of Teaching Strategies (GATS)

The Global Appraisal of Teaching Strategies (GATS) is a 12-item checklist (see Exhibit C.20 for a copy of the GATS) adapted from the Checklist of Teacher Competencies (Foorman et al., 2006).[25] 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 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.[26]

For the first ten items on the GATS, the observer indicated how often the teacher demonstrated the behaviors targeted by each item, using a five point scale where 1 = all the time, 2 = more than half the time, 3 = half the time, 4 = less than half the time, and 5 = never. (A sixth "not observed" option was included in case the item did not apply during a particular observation; e.g., "takes advantage of opportunities to provide corrective feedback when student makes error" may not have been observed if no student errors occurred during the observation). For items 11 and 12, the observer also rated the teacher on a five point scale, but the descriptors of these scales varied to match the focus of the question.[27]

[25] B. Foorman graciously shared the Checklist of Teacher Competencies with the RFIS Study Team but played no role in the development of the GATS. See Foorman & Schatschneider (2003) and Foorman et al., (2006).

[26] The RFIS did not conduct inter-rater reliability analyses for the GATS data, because the study did not include GATS data in impact analyses, and study resources were focused on obtaining reliability data for observational data that were included in impact analyses.

[27] For item 11, the descriptors were 1 = true, 2 = mostly true, 3 = not sure, 4 = mostly not true, and 5 = not true. For item 12, the descriptors were 1 = all students, 2 = more than half of students, 3 = half of students, 4 = less than half of students, and 5 = none.

Factor analyses revealed that all 12 items loaded onto a single factor. Thus, each classroom was assigned a single GATS score, as follows:

- any item marked "not observed" was set to missing.
- the rating for each item was averaged across the two GATS observations for the classroom in each data collection wave.
- the 12 items were averaged together.

Before analyses, the assigned GATS score was subtracted from 6 to reverse the scale so that a higher rating corresponded to more frequent demonstration of the behavior.

Part 4: Student Time-on-Task and Engagement with Print (STEP)

The Student Time-on-Task and Engagement with Print (STEP) instrument was designed to capture information about student engagement during reading instruction as part of the Reading First Impact Study's (RFIS) classroom observation data collection. The STEP is focused on student behavior; it complements the Instructional Practice in Reading Inventory (IPRI) measure, which focuses on teacher behaviors. See Appendix C, Part 7, Exhibit C.21 for a copy of the STEP.

The STEP was designed to collect aggregate, not individual level, data on the percentage of students in classrooms during the scheduled reading block who are on-task and/or interacting with print. The STEP instrument combines a dichotomous "on-task/off-task" rating with additional indicators for student engagement with print.

The data collected with the STEP instrument do not measure the amount of time students are on-task or the amount of time students are engaged with print. Rather, across all students in the classroom, the STEP instrument yields data on the percentage of students who, at a particular point in time, are on-task and engaged with print.

During each wave of classroom observation data collection, one observer per school was assigned to collect student engagement data in each classroom being observed by IPRI observers. STEP observations took place during the reading block in each classroom. While each classroom was observed twice for the IPRI, each classroom was observed once for the STEP.

Each STEP observation consists of data on student engagement from three sweeps of a classroom. Specifically, for each sweep, at an interval of six minutes, an observer classifies every student in the classroom as either on- or off-task, and, if the student is on-task, whether the student is:

a) reading connected text (e.g., a paragraph, story, or longer passage); or

b) reading isolated text (letters, words, or sentences in isolation); and/or

c) writing; or

d) none of the above (i.e., not engaged with print).

A student can be marked as on-task without being engaged with print (but a student cannot be off-task and engaged with print). An on-task student can also be engaged with more than one type of print (e.g., the student is writing on a worksheet that contains isolated text, such as a list of words). The observer records student behavior for each student in each observed classroom three times.

Between sweeps, the observer waits until six minutes have elapsed before beginning the next sweep. After the third sweep, the observer moves on to the next classroom in the sample. The observation protocol is summarized in Exhibit C.12.

Exhibit C.12: Prototypical STEP Observation in One Classroom

Classroom A	Duration (minutes)	Sample Clock Time	Activity
Rest period 1	6	8:00-8:06	Observer waits for children to acclimate
Sweep 1	3	8:06-8:09	Observer records data on each student in classroom
Rest period 2	3	8:09-8:12	Observer waits
Sweep 2	3	8:12-8:15	Observer records data on each student in classroom
Rest period 3	3	8:15-8:18	Observer waits
Sweep 3	3	8:18-8:21	Observer records data on each student in classroom
Switch classes	6	8:21-8:27	Observer exits Classroom 1 and moves to next classroom
Total time per classroom	**27 min**	Time is approximate (travel time between classrooms may be shorter or longer than 6 minutes)	

NOTE:

The duration of a sweep varies depending on how long it takes the observer to record data on all students in the classroom, but never exceeds three minutes. Exactly six minutes separate the start of one sweep and the start of another.

Under certain circumstances, observers skipped a scheduled sweep. First, if at the time of a scheduled sweep, more than one-half of the students in the classroom were transitioning from one activity to another (e.g., students were rotating between activity "centers"), the observer skipped that sweep. Second, if at the time of a scheduled sweep, the whole class was listening to the teacher read aloud, and the students themselves did not have access to the printed text, the observer skipped that sweep.[28]

Data Collection and Response Rates for Fall 2005, Spring 2006, Fall 2006, and Spring 2007

The STEP was added to the classroom observation data collection battery beginning in fall 2005, reflecting a decision by IES staff (Institute for Education Sciences, U.S. Department of Education) overseeing the RFIS to augment the teacher-focused data collection (using the IPRI) with a student-focused measure. STEP observations were done in grade 1 and 2 classrooms in fall 2005, spring 2006, fall 2006, and spring 2007 by trained field staff who had successfully completed the requirements of the classroom observation training. As described above, during two consecutive days of classroom observations, STEP observations were completed once in each classroom, yielding one STEP record per classroom. For 2005-2006 a total of 2,715 STEP observations were completed in first and second grade

[28] These protocols were implemented because pilot-testing of the instrument revealed that on- and off-task judgments were difficult to make reliably under these two circumstances.

classrooms, which represents a 98 percent completion rate for expected observations. In 2006-2007, a total of 2,764 first and second grade observations were made, or 98 percent of expected observations. A flow chart of the sampling process and STEP response rates is presented in Exhibit C.13.

Analytic Variables

The RFIS Team focused on the percentage of students engaged with print as the primary analytic variable derived from the STEP data to be used in impact analyses. This variable was created for each classroom by first summing the number of students in each sweep who were on-task and who were either reading connected text, reading isolated text, or writing. The percentage of students engaged with print for each sweep was then calculated as the number of students engaged with print divided by the total number of students that the observer rated in the sweep (i.e., the number of students in the classroom at the time the sweep was conducted). The percentage of students engaged with print for each sweep was then averaged across the number of sweeps available for that classroom.[29]

STEP Reliability

For reasons of parsimony, results from the fall 2006 STEP training are presented below. Observers were trained on the STEP measure using a combination of still photographs and 3-second video clips of first and second grade students during reading instruction.[30] Trainees viewed five practice sequences, containing both still photographs and short video clips. A sixth sequence of video clips (hereafter, the "test tape") was used to assess the average inter-rater reliability of observers' judgments about student engagement.

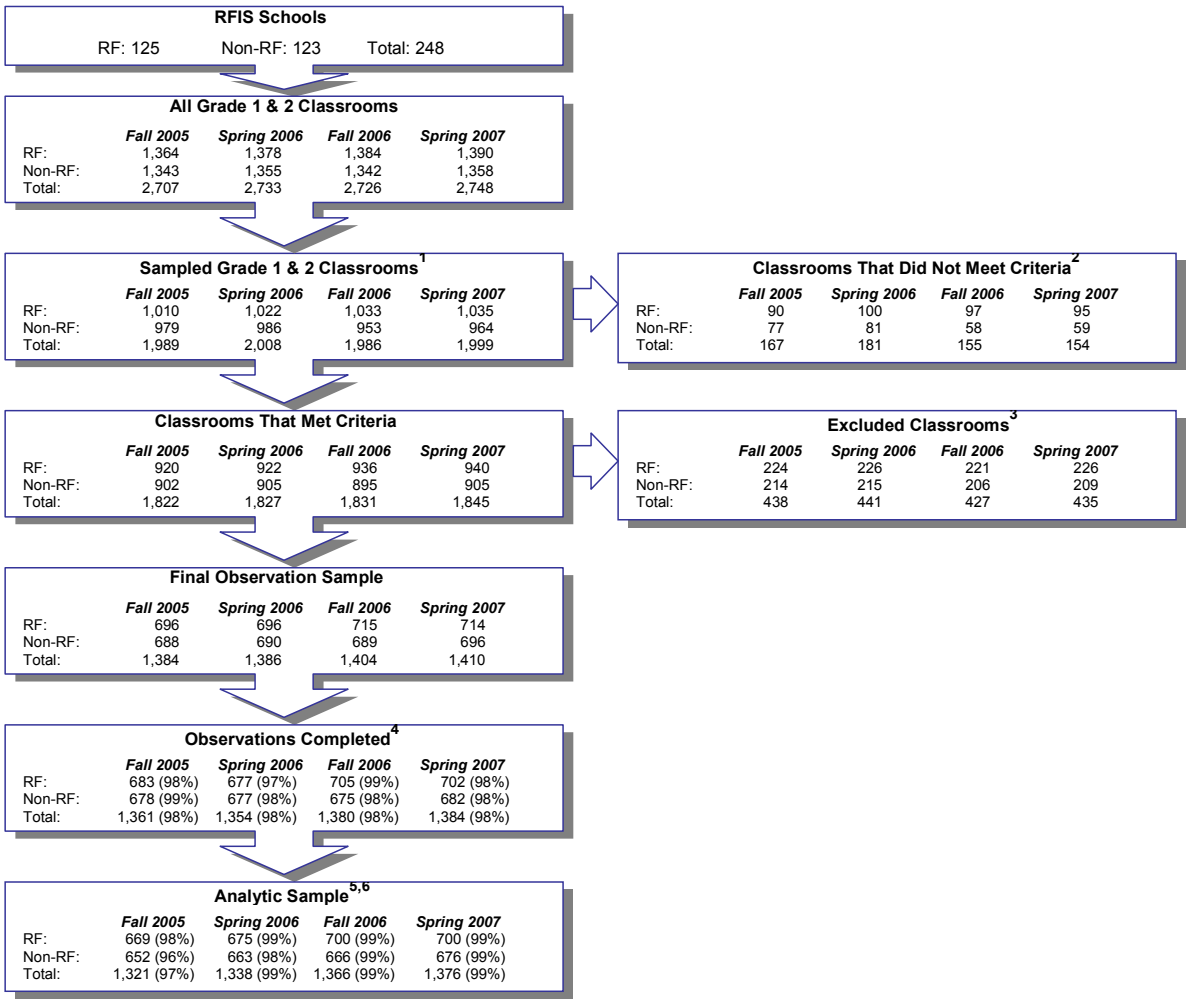
The test tape was designed to simulate a single "sweep," and it included three-second clips of 15 first- or second-grade students. Two master coders had viewed and scored the test tape to arrive at a set of master codes for each student on the tape.

Percent agreement was calculated for each trainee with the master codes for each code (i.e., On-Task, Reading Connected Text, Reading Isolated Text, Writing), and then a mean percent agreement was calculated across trainees for each code. Next, overall percent agreement was calculated by aggregating across codes.

[29] For the pooled analytic dataset (fall 2005, spring 2006, fall 2006, and spring 2007), 70 percent of classrooms had three sweeps of data; 23 percent had two sweeps of data; 5 percent had one sweep of data; and 1 percent was missing all three sweeps. (Numbers do not add to 100% due to rounding.)

[30] Classroom reading instruction was filmed in both Reading First and non-RF classrooms for the purpose of creating a training resource for the RFIS.

Exhibit C.13: STEP Data Collection: School, Classroom, and Observation Sample Information



Notes:

1 The study conducted observations in all classrooms in schools if across the site as a whole, the average number of classrooms per grade per school was three or less. If for that site as a whole, the average number of classrooms per grade per school exceeded three, the study sampled classrooms for observations within schools.

2 At the beginning of each data collection wave, the study team contacted schools to obtain classroom rosters and indications of which classrooms were designated as regular classrooms, self-contained special education classrooms, or ESL classrooms. Those classrooms identified as special education or ESL were considered ineligible for observation.

3 Once the study team began to schedule and conduct observations, classrooms were excluded from the sample if the information that had been provided earlier about special education or ESL status was inaccurate, if the class was taught by someone other than the regular teacher or a long-term substitute, or if fewer than 75 percent of the students were in the target grade.

4 Response rates are calculated by dividing the number of observations completed by the eligible classrooms selected into the sample.

5 Classrooms were dropped from the analytic sample if, for all three sweeps, the class was in transition between activities or the entire class was listening to a story. Of the 78 classrooms (1%) for which this was the case, 23 were Reading First and 55 were non-Reading First, which corresponds to 1% of Reading First classrooms and 2% of non-Reading First classrooms in the pooled analytic sample.

6 Percentages are calculated by dividing the analytic sample numbers by the number of observations completed.

SOURCE: RFIS Student Time-on-Task and Engagement with Print, fall 2005, spring 2006, fall 2006, and spring 2007

As shown in Exhibit C.14, observers achieved an average of 89 percent agreement across all codes appearing in the test tape. Seventy-five percent of the observers scored at least 86 percent overall agreement. Observers had the lowest average agreement about whether or not a student was Reading Isolated Text (77 percent), and they achieved the highest level of agreement when judging that a student was Writing (96 percent).[31] These differences reflect the fact that the video cameras could zoom in and capture students' expressions more effectively than they could discern the specific types of text with which students were engaged. During actual data collection, observers could move around the classrooms to determine whether students were engaged with specific types of text.

Exhibit C.14: Percent Correct by Code and Overall for STEP Reliability Tape, Fall 2006

	Percent Agreement				
		Student Is …			
	On Task	Reading Connected Text	Reading Isolated Text	Writing	Overall
Mean	92	92	77	96	89
Minimum	60	67	50	75	73
25th percentile	87	92	67	92	86
50th percentile	93	92	75	100	90
75th percentile	100	92	83	100	92
Maximum	100	100	100	100	100

NOTES:

The number of observers tested on this tape is 130.

EXHIBIT READS: Observers in the fall 2006 training achieved an average of 92 percent agreement on whether a student was on-task; 92 percent agreement on whether a student was reading connected text; 77 percent agreement on whether a student was reading isolated text; 96 percent agreement on whether a student was writing; and 89 percent agreement across all codes appearing in the test tape.

Part 5: Reading Achievement

At the heart of this evaluation is a question about the impact of Reading First on the reading achievement of students. The RFIS had initially planned to use a battery of tests to assess students' reading skill across the components of reading instruction targeted in the legislation (phonemic awareness, phonics, fluency, vocabulary, and comprehension), but when the study's design shifted to RDD, with a much larger number of schools, the planned data collection activities also changed. The RFIS Team, working with its Technical Work Group and staff from the National Center for Education Evaluation/Institute of Education Sciences at the Department of Education, focused its efforts on identifying a single test of reading comprehension. In the spring of 2007, for first graders only, the RFIS added the Test of Silent Word Reading Fluency (TOSWRF, Mather et al., 2004) to assess students' decoding skills, which are distinct from, although related to, reading comprehension.

[31] In fall 2005, similar results were obtained from the previous group of observers. They achieved, on average, 87 percent agreement across all codes appearing in the test tape. Seventy-five percent of the trainees scored at least 84 percent overall agreement. Trainees had the lowest average agreement on the Reading Isolated Text code (75 percent), and the highest level of agreement (95 percent) on the Writing code. (The test tape featured only one student who was engaged in Writing.)

Reading Comprehension

Reading Comprehension Instrument Selection

Stanford Achievement Test—10th Edition (SAT 10). The team's priorities in selecting a test for this study included, first, finding a test that directly measured skills related to text comprehension. Other factors included: ease and appropriateness of administration to groups or entire classrooms of students—including appropriateness for fall first grade; modest time demands; use of a norm-referenced test; and consistent reliability and validity. The team also sought a measure that had already been widely used in large-scale studies, and therefore would be more likely to be credible in the research community.

At the outset of the test selection and review process, the team identified 47 assessments of text comprehension that either had been proposed for use by states in their Reading First schools or had been proposed for use in other Department of Education-sponsored evaluations involving preschool and the early elementary grades. From this pool of tests, we identified six test batteries with subtests of reading comprehension that could be group-administered and were valid for fall of first grade.[32] The six test batteries included:

1. ITBS Total Core Battery Reading Subtest;
2. Terra Nova/CTBS Basic Battery Reading Subtest;
3. Gates/MacGinitie Reading Test-3 (GMRT);
4. GRADE (Group Reading Assessment and Diagnostic Evaluation);
5. Stanford Achievement Test—10th Edition (SAT 10); and
6. Stanford Reading First.

Five of the six tests have reliability coefficients reported in published manuals of close to 0.90 for the majority of subtests. Because the reliability for Terra Nova Grade 1 was 0.76, and data were not available for the other grade levels, that test was eliminated from consideration. The Stanford Reading First Test was also eliminated, because it had been normed on a relatively small sample according to a conversation with a Harcourt representative in 2004 (< 400 students across several grade levels), whereas the remaining five tests had been normed on samples of 1,000 or more students.

Next, the team reviewed two related aspects of the tests: the number of items and amount of time required. The number of items varies considerably—from approximately 30 to 80, with fewer items typically required for grade 3 tests (although the amount of time required per item increases by grade level). The tests also vary in amount of time required, from 50 minutes for the Stanford Reading First at all three grade levels to 95 minutes for the GRADE in grade 1. The amount of time required was a consideration, but not the deciding factor. The final consideration was the relative frequency of use for the four remaining assessments in schools in the study sample. Of the states and districts that (in Summer 2004) administered standardized reading assessments to children in grades 1, 2, and 3, more used the SAT 10 than any other test (although none did so in fall of grade 1). The study consequently chose the SAT 10 because it both met all the criteria above and because its use might allow the study to collect extant data, which would reduce the testing burden on students and schools. (Where extant data were not available, the study would administer the SAT 10.)

[32] See published manuals (Hoover et al., 2003; CTB/McGraw-Hill, 2003; MacGinitie et al., 2000; Williams, 2001; Harcourt Assessment, Inc., 2004).

The specific properties of the SAT 10 are summarized in Exhibit C.15.

Exhibit C.15: Features of SAT 10: Reading/Listening Comprehension for Spring Administration

	Grade Level		
	Grade 1 Spring (Primary 1)	**Grade 2 Spring (Primary 2)**	**Grade 3 Spring (Primary 3)**
Number of Items	40	40	54
Time in Minutes	50	50	60
Test-Retest Reliability*	.91	.91	.93
Concurrent Validity	To SESAT-2:[1] .63 Form A to B: .87	To Primary 1: .69 Form A to B: .85	To Primary 2: .80 Form A to B: .83
N in Norming Sample	3,392	3,558	2,160

*Reliability is test-retest Kuder-Richardson formula 20 (KR 20)

[1] Stanford Early School Achievement Test.

SOURCE: Harcourt Assessment, Inc. (2004)

Decoding

Test of Silent Word Reading Fluency (TOSWRF). In the spring of 2007, the study team augmented the SAT 10 reading comprehension assessment with the TOSWRF for first grade students only. The TOSWRF is a short (three-minute) assessment that measures students' ability to identify words quickly and correctly. It assesses students' decoding skills, which are distinct from, although related to, reading comprehension.

The RFIS added the TOSWRF because decoding is an important reading sub-skill. Based on recommendations from reading researchers, the TOSWRF was selected from among several possible decoding measures. Key factors in selecting an assessment included both pragmatic data collection considerations, including the manner of administration (individual versus group) and length of administration, and psychometric properties of the assessment. It was important that the assessment be not only valid and reliable, but also group administered and brief, in order to limit changes to the data collection protocol in the last year of the study.

The TOSWRF was normed in 2001 and 2002 using a representative sample of 3,592 students residing in 32 states. Raw scores can be converted to standard scores, percentile ranks, and age and grade equivalents. Reliability coefficients for the TOSWRF are .90 or higher across four types of reliability: alternate form, test-retest, alternate form (delayed administration), and scorer differences. Although a relatively new test, initial validity of the TOSWRF has been established by correlating it with other measures of reading fluency, word identification, and comprehension. Corrected coefficients for criterion-prediction validity range from .42 to .78.[33]

[33] For specific tests used in validity analyses and for more information on the psychometric properties, see the *TOSWRF Examiner's Manual* (Mather, N., Hammill, D., Allen, E., and Roberts, R., 2004).

Data Collection and Response Rates

In six sites, the RFIS obtained SAT 10 data directly from state and/or district education officials. In 12 sites, the RFIS collected test data directly. In all sites, the RFIS obtained TOSWRF data directly. The SAT 10 student assessments were administered in grades 1, 2, and 3, at four timepoints: fall 2004, spring 2005, spring 2006, and spring 2007, while the TOSWRF was administered in grade 1 only at one timepoint (spring 2007). To conduct the testing, one site assessment coordinator was hired at each district (local), and that coordinator in turn hired a local team of test administrators. Since both the SAT 10 and the TOSWRF are standardized tests, the requirements of the test publisher for administration were followed. Site assessment coordinators also observed each test administrator in the classroom for quality control and technical assistance. In addition, staff from the home office visited districts during the testing for quality control purposes.

The study team collected classroom rosters prior to administration, and used these rosters to pre-label the student test booklets with the student ID and a strippable name label. Once the test booklet was complete, the test administrator stripped the name label from the booklet (for privacy purposes) and adhered it to a receipt sheet. The test administrator then delivered the completed booklets and the receipt sheet to the site assessment coordinator who was responsible for keeping track of who had been tested and who required make-up testing. A computerized field management system allowed the site coordinators to receive the booklets and also to print out a list by school and grade regarding which students needed makeup testing. Once testing was complete in the district, the site coordinator shipped the hardcopy test booklets to be processed.

In fall 2004, there were two main factors in maximizing SAT 10 response rates: obtaining parent permission at more than one timepoint, and administering make-up tests for students who missed the originally scheduled testing sessions. In the initial two weeks of student assessment, the RFIS assessed all students present in the classroom who had returned signed permission slips. Study staff worked with school liaisons prior to the scheduled assessment date to obtain as many permission slips as possible.

For those students who returned permission slips after the scheduled assessment day, or were absent, group make-up sessions were held at each school. Students were not eligible for the assessments if they were excluded from testing in accordance with their own school or district policies (generally because they received instruction primarily in a language other than English), and/or needed special accommodations (particularly an exam writer/scribe). The consent rates and resultant response rates were considerably lower than hoped in fall 2004 (75 and 70 percent, respectively, for Reading First and comparison schools). The RFIS obtained a waiver from participating districts and from the Abt Associates IRB to use passive consent in subsequent testing, which increased the effective SAT 10 response rates to 84 and 83 percent, respectively, for Reading First and comparison schools in spring 2005. During the 2005-06 school year, the effective SAT 10 response rate was 86% for both RF and non-RF schools. During the 2005-06 school year, the effective SAT 10 response rate was 86% for both reading first and non-reading first schools. In 2006-07, the effective SAT 10 response rate was 88% for reading first schools and 85% for non-reading first schools. The TOSWRF response rates were 87% for RF schools and 85% for non-RF schools. A flowchart presenting student assessment sample information by school year in the 12 sites in which the RFIS collected test data directly is presented in Exhibit C.16.

Exhibit C.16: Student Assessment Data Collection: Sample School and Student Information

SAT 10

Panel 1: Data Collected Directly by the RFIS (12 sites)

RFIS Schools

RF: 91 Non-RF: 89 Total: 180

All Students in Grades 1-3

	Fall 04	Spring 05	Spring 06	Spring 07
RF:	18,855	19,157	18,379	18,506
Non-RF:	19,409	19,627	19,764	19,901
Total:	38,266	38,784	38,143	38,407

Sample Students in Grades 1-3[1]

	Fall 04	Spring 05	Spring 06	Spring 07
RF:	18,855	19,157	14,616	15,724
Non-RF:	19,409	19,627	14,689	15,794
Total:	38,266	38,784	29,305	31,518

Students Not Eligible to be Tested[2]

	Fall 04	Spring 05	Spring 06	Spring 07
RF:	1,007	1,021	377	440
Non-RF:	1,226	1,247	372	445
Total:	2,303	2,268	749	885

Students Eligible to be Tested

	Fall 04	Spring 05	Spring 06	Spring 07
RF:	17,778	18,136	14,239	15,284
Non-RF:	18,183	18,380	14,317	15,349
Total:	35,961	36,514	28,556	30,633

Eligible Students Not Tested[3]

	Fall 04	Spring 05	Spring 06	Spring 07
RF:	5,052	3,260	2,304	2,220
Non-RF:	5,272	3,734	2,180	2,299
Total:	10,324	6,994	4,484	4,519

Students Eligible to be Tested

	Fall 04	Spring 05	Spring 06	Spring 07
RF:	12,726 (72%)	14,876 (82%)	11,935 (84%)	13,064 (85%)
Non-RF:	12,911 (71%)	14,646 (80%)	12,137 (85%)	13,050 (85%)
Total:	25,637 (71%)	29,522 (81%)	24,072 (84%)	26,114 (85%)

Panel 2: Data Collected Directly by the RFIS (12 sites) and Extant Data (6 sites)

Overall Tested[4,5]

	Spring 05	Spring 06	Spring 07
RF:	22,672 (85%)	19,273 (86%)	19,903 (88%)
Non-RF:	21,097 (82%)	17,227 (86%)	17,272 (85%)
Total:	43,769 (83%)	36,500 (86%)	37,175 (85%)

Analytic Sample[6,7]

	Spring 05	Spring 06	Spring 07
RF:	22,581 (100%)	19,273 (100%)	19,902 (100%)
Non-RF:	20,971 (99%)	17,225 (100%)	17,270 (100%)
Total:	43,552 (100%)	36,498 (100%)	37,172 (100%)

Notes:

The information presented in the top panel represents the 12 sites in which the RFIS collected all test data directly. The information in the bottom panel includes data from those 12 sites plus data from six sites for which the RFIS obtained student test data from state and/or district education officials.

[1] In 2004-05, the study administered SAT 10 tests to all students in grade 1-3 classrooms. In 2005-06 and 2006-07, the study administered SAT 10 tests to all students in grade 1-3 classrooms if across the site as a whole, the average number of classrooms per grade per school was three or less. If for that site as a whole, the average number of classrooms per grade per school exceeded three, the study sampled classrooms within schools and tested students in those classrooms. Students in classrooms that were self-contained special education classrooms or classrooms in which instruction occurred in languages other than English were ineligible for testing.

[2] Students were not eligible for assessments if excluded from testing in accordance with their own school or district policies (e.g., because they received instruction in a language other than English), and/or they needed special accommodations beyond those that could be provided through additional time in a group administered testing situation.

[3] Eligible students were not tested if they were absent at the time the test was given and could not be rescheduled, they had transferred out, they had refused to take the test, or the RFIS did not have consent for them to participate in the study.

[4] Boxes in Panel 2 do not include data from fall 2004 because those data were used only to construct a pretest covariate at the school level.

[5] Response rates for the six sites for which the RFIS obtained student test data from state and/or district education officials were estimated.

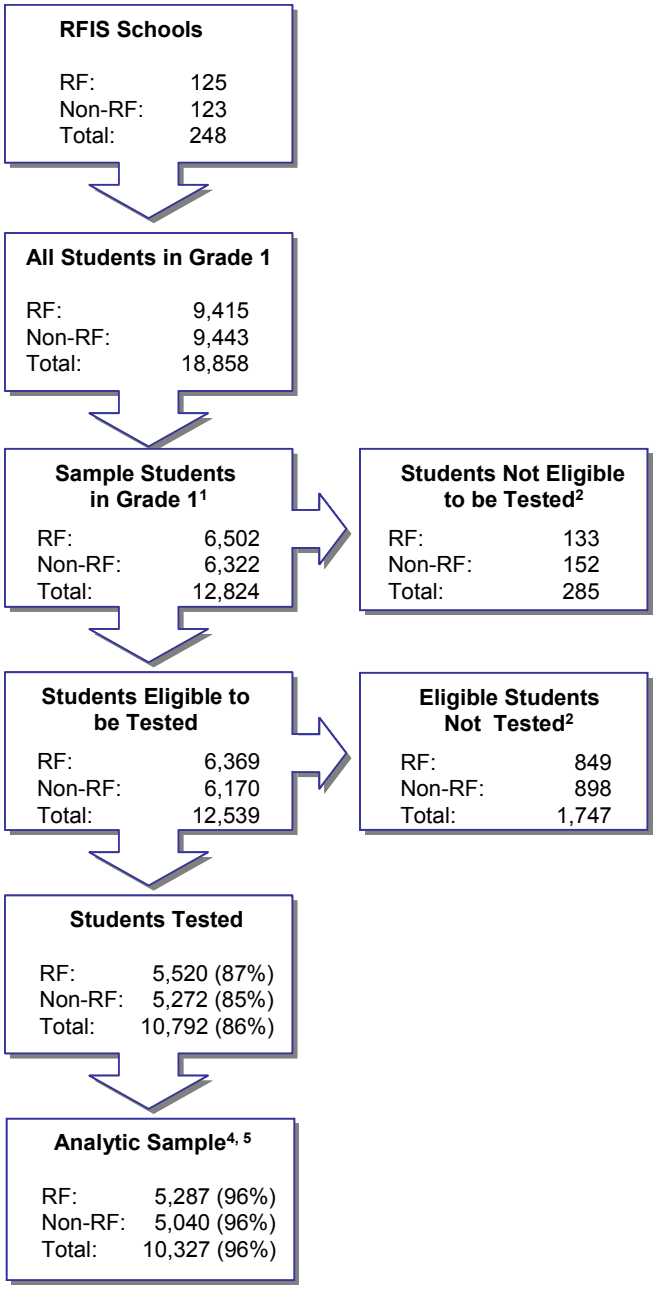
[6] Students test scores were excluded from the analytic sample if they were deemed invalid according to SAT 10 scoring guidelines. For the pooled sample (spring 2005, spring 2006, spring 2007), this amounted to 222 student test scores (0.2%). Of the missing scores, 92 were Reading First and 130 were non-Reading First, which corresponds to 0.1% of Reading First test scores and 0.2% of non-Reading First test scores.

[7] Percentages are calculated by dividing the analytic sample numbers by the number of students tested.

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.

Exhibit C.16: Student Assessment Data Collection: Sample Information (continued)

TOSWRF (Spring 2007 Only, Grade 1 Only)



Notes:

[1] In 2006-07, the study administered the TOSWRF to students in those grade 1 classrooms where observations were conducted. Students in classrooms that were self-contained special education classrooms or classrooms in which instruction occurred in languages other than English were ineligible to be tested.

[2] Students were not eligible for assessments if they were excluded from testing in accordance with their own school or district policies (generally because they received instruction primarily in a language other than English), and/or they needed special accommodations beyond those that could be provided through additional time in a group administered testing situation.

[3] Eligible students were not tested if absent at the time the test was given and could not be rescheduled, they had transferred out, they had refused to take the test, or the RFIS did not have consent for them to participate in the study.

[4] Student test scores were excluded from the TOSWRF analytic sample if birth dates were missing, or out of range, or if students did not follow test instructions. Of the 465 test scores (5%) excluded from the analytic sample, 233 were Reading First and 232 were non-Reading First, which corresponds to 4% of Reading First scores and 4% of non-Reading First scores in the analytic sample.

[5] Percentages are calculated by dividing the analytic sample numbers by the number of students tested.

SOURCES: RFIS TOSWRF administration, spring 2007.

In the 2004-05 school year, the study team endeavored to test all students within grades 1, 2, and 3 in the participating schools using the SAT 10. However, the fact that some schools had as many as 10 or 12 classrooms per grade level led the study team to sample classrooms within grades in subsequent testing, such that the team assessed an average of three classrooms per grade per school in spring 2006 and spring 2007 (in spring 2007, this approach was also used for the TOSWRF). Note that the RFIS tested all students as required by local policy in those schools that routinely administered the SAT 10 reading comprehension as part of state- or district-standardized assessment. In all sites, SAT 10 and TOSWRF testing procedures were equivalent for Reading First and for comparison schools. Some sites required classroom teachers to administer tests; other sites relied upon RFIS staff to administer assessments. In the latter sites, the RFIS Team worked with district officials to carry out testing in accordance with local guidelines.

In the spring of 2007, the TOSWRF was administered to 10,792 first grade students, with an average response rate of 86% across Reading First and non-Reading First classrooms. Analyses were conducted using standard scores, which are reported in tables. Conversions to percentiles and grade equivalents are also provided in tables for reference, however, the publisher of the TOSWRF discourages the use of grade equivalents due to their unreliability.

Part 6: Data Collection Instruments

Exhibit C.17: Reading Coach Survey

OMB Number: 1850-0797
Expiration Date: 10/31/07

Reading First Impact Study
Reading Coach Survey

Abt Associates has been commissioned by the Institute of Education Sciences (IES) at the U.S. Department of Education to conduct an independent national evaluation of the Reading First program. Your school is one of only 250 elementary schools that have been selected to participate in this study, so your participation is extremely important. The study includes two kinds of schools: half of which have received Reading First funding, and half of which have not received Reading First funding. We know that teaching children how to read is important in every single school, and that is why we are asking teachers, staff, and administrators in both kinds of schools to describe the reading instruction in their schools and classrooms. Your responses will help inform the U.S. Department of Education, Congress, policymakers, practitioners, and researchers about how reading instruction is implemented in schools and what strategies schools use to provide high-quality, evidence-based reading instruction in first, second, and third grades.

Additional Information

The survey will take you approximately 30 minutes to complete. All responses to the survey will be kept confidential. All individual identifying information will be used only by persons on the research team. Information such as school location (state), participants' general job titles, grades they teach, and gender will be included in the study data files to be submitted to the Department of Education. However, participants' names will be stripped from all analysis data files and data files to be submitted to the Department of Education. We will not report any data about individual classrooms—all information will be reported at the grade and school levels. Neither your school nor your district will have access to any of the completed surveys at any time.

Thank you for your cooperation with this survey!

Instructions

Unless otherwise noted, your responses should reflect your experiences during the 2006-2007 school year in the school to which this survey was sent.

- Please complete all questions; each question includes directions for recording your answer.

- You are sometimes told to skip over some questions in the survey. When this happens, you will see an arrow with a note that tells you what question to answer like this:

  ❑1 Yes
  ❑2 No → Skip to E4

- If you have any questions about how to complete the survey, please call: 1-866-421-6982. This is a free call and will connect you with our expert interviewers who can assist you.

A. Your Background and Experience

A **reading coach** is a staff member whose primary role is to provide ongoing training and support to school staff in the delivery of effective reading instruction.

	Enter # below…
A1. Including this year, for how many years have you been the K–3 reading coach for this school? (If less than one year, enter 1.)	_____ years 10-11/
A2. Including this year, for how many years have you worked at this school in any capacity? (If less than one year, enter 1.)	_____ years 12-13/
A3. Including this year, how many years of classroom experience do you have, as either a teacher and/or reading coach? (If less than one year, enter 1.)	
a. Number of years experience as a reading coach	_____ years 14-15/
b. Number of years experience as a teacher	_____ years 16-17/

B. Coach Responsibilities

	Enter # below…
B1. This school year, for how many schools do you serve as the reading coach (including this school)?	_____ Schools 18-19/
B2. This school year, for how many teachers do you serve as the reading coach (include all teachers in all schools)?	_____ Teachers 20-21/
B3. Approximately what percentage of your time do you spend as the K–3 reading coach for this school?	_____ % 22-24/

B4. How central is each of the following activities to your work this year (since July 1st) at this school?

Please rate the activity a "1" if you do not do the activity or if it is not at all central to your role as the literacy coach. Rate the activity a "5" if it is absolutely central or critical to your work.

Activity	Check only one box for each item… Do not do or not at all central				Somewhat central		Absolutely central
a. Administering/coordinating reading assessments	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 25/		
b. Compiling reading assessment data for teachers	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 26/		
c. Facilitating grade level meetings	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 27/		
d. Participating in school leadership team meetings	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 28/		
e. Facilitating or coordinating family literacy activities	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 29/		
f. Ordering/managing reading instruction materials	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 30/		
g. Participating in professional development provided by th district, state or other consultants	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 31/		
h. Providing sub time for teachers to observe other more experienced teachers	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 32/		
i. Providing direct reading instruction to students	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 33/		
j. Providing training/professional development in reading materials, strategies, and assessments	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 34/		
k. Coaching staff on a range of topics (note: specific coachi activities are asked about in the next item)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 35/		
l. Organizing professional development for K-3 teachers	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 36/		
m. Coordinating activities and meetings between classroom and special education teachers	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 37/		
n. Coordinating activities and meetings between classroom teachers and English Language Learner (ELL) staff	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 38/		
→ Check if no ELLs ☐ 1 39/							
o. Other (Please specify): _____ 41-42/	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 40/		

English language learner (ELL) indicates a student who is in the process of acquiring English and **has a first language other than English**. Other common related terms include language minority or limited English proficient (LEP) students, students in English as a second language (ESL), or students in classes for English for speakers of other languages (ESOL).

B5. When you **coach** K-3 staff, how central has each of the following activities been to your work this year (since July 1st)?

Please rate the activity a "1" if you do not do the activity or if it is not at all central to your role as the literacy coach. Rate the activity a "5" if it is absolutely central or critical to your work.

Coaching Activity	Check only one box for each item…				
	Do not do or not at all central		**Somewhat central**		**Absolutely central**
a. Giving demonstration lessons using core or supplemental materials	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 43/
b. Assisting teachers in using the core program	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 44/
c. Observing and providing feedback to teachers	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 45/
d. Assisting teachers in forming instructional groups	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 46/
e. Assisting teachers in designing strategies for addressing the needs of struggling readers	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 47/
f. Assisting teachers with monitoring the effectiveness of strategies addressing the needs of struggling readers	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 48/
g. Giving demonstrations on assessment administration and scoring	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 49/
h. Planning reading instruction with teachers	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 50/
i. Reviewing teachers' lesson plans and providing feedback	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 51/
j. Assisting teachers in interpreting assessment results	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 52/
k. Assisting teachers in designing strategies for addressing the needs of special education students	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 53/
l. Assisting teachers in designing strategies for addressing the needs of ELLs (see page 2 for definition of ELL)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 54/
→ Check if no ELLs ☐ 1 55/					

C. Reading Instructional Materials

C1. Which **core reading program** is being used to teach reading in each of grades K-3 **at this school**? We have provided a partial list of core reading programs below. Please check the core reading program used in each grade. If the core reading program you use is **not** included below, please check "other," and write in the name of the program and the publisher and/or developer (if applicable).

Note: Inclusion of a core reading program on the list below does not constitute an endorsement by the U.S. Department of Education.

A **core reading program** is one that provides a comprehensive program of instruction on a daily basis in all aspects of reading.

Publisher/ Developer	Core reading program	Check one per grade K 56-57/	1 58-59/	2 60-61/	3 62-63/
Addison Wesley	Unspecified or other (Please specify): _____ 64-65/	❑1	❑1	❑1	❑1
Harcourt	Collections	❑2	❑2	❑2	❑2
	Rigby Reading	❑3	❑3	❑3	❑3
	Signatures	❑4	❑4	❑4	❑4
	Trophies	❑5	❑5	❑5	❑5
	Unspecified or other (Please specify): _____ 66-67/	❑6	❑6	❑6	❑6
Heinemann	Fountas Pinnel units of study	❑7	❑7	❑7	❑7
	Unspecified or other (Please specify): _____ 68-69/	❑8	❑8	❑8	❑8
Houghton Mifflin	Horizons	❑9	❑9	❑9	❑9
	Invitation to Literacy	❑10	❑10	❑10	❑10
	Lectura	❑11	❑11	❑11	❑11
	Legacy of Literacy	❑12	❑12	❑12	❑12
	Nation's Choice	❑13	❑13	❑13	❑13
	Reading	❑14	❑14	❑14	❑14
	State Specific Edition	❑15	❑15	❑15	❑15
	Unspecified or other (Please specify): _____ 70-71/	❑16	❑16	❑16	❑16
McGraw-Hill	Open Court	❑17	❑17	❑17	❑17
	Reading	❑18	❑18	❑18	❑18
	Reading Mastery	❑19	❑19	❑19	❑19
	Spotlight on Literacy	❑20	❑20	❑20	❑20
	Unspecified or other (Please specify): _____ 72-73/	❑21	❑21	❑21	❑21
Saxon	Saxon Phonics	❑22	❑22	❑22	❑22
	Unspecified or other (Please specify): _____ 74-75/	❑23	❑23	❑23	❑23

C1. CONTINUED. Which **core reading program** is being used to teach reading in each of grades K-3 **at this school**? We have provided a partial list of core reading programs below. Please check the core reading program used in each grade. If the core reading program you use is **not** included below, please check "other," and write in the name of the program and the publisher and/or developer (if applicable).

Note: Inclusion of a core reading program on the list below does not constitute an endorsement by the U.S. Department of Education.

Publisher/ Developer	Core reading program	Check one per grade K	1	2	3
Scholastic	Literacy Place	❑24	❑24	❑24	❑24
	Unspecified or other (Please specify): _____ _76-77/	❑25	❑25	❑25	❑25
Scott Foresman	Literacy Works	❑26	❑26	❑26	❑26
	Reading	❑27	❑27	❑27	❑27
	State Specific Edition	❑28	❑28	❑28	❑28
	Unspecified or other (Please specify): _____ _78-79/	❑29	❑29	❑29	❑29
Sopris	Read Well	❑30	❑30	❑30	❑30
	Unspecified or other (Please specify): _____ _80-81/	❑31	❑31	❑31	❑31
Success for All	Success for All	❑32	❑32	❑32	❑32
	Unspecified or other (Please specify): _____ _82-83/	❑33	❑33	❑33	❑33
Voyager	Universal Literacy	❑34	❑34	❑34	❑34
	Unspecified or other (Please specify): _____ _84-85/	❑35	❑35	❑35	❑35
Wright Group	Unspecified or other (Please specify): _____ _86-87/	❑36	❑36	❑36	❑36
Other	Publisher/Developer: _88-89/ Program Title: _90-91/	❑37	❑37	❑37	❑37
	Publisher/Developer: _92-93/ Program Title: _94-95/	❑38	❑38	❑38	❑38
I use a core reading program developed by teachers or other school personnel		❑39	❑39	❑39	❑39
I do not use a core reading program		❑40	❑40	❑40	❑40

C2. Has your school made any of the following **changes** to your reading program that took effect at the beginning of the current school year (2006-2007)?

	Indicate 'yes' or 'no' for each item …	
	Yes	No
a. Adopted a new core reading program	☐ 1	☐ 2 96/
b. Added a new intervention program for struggling readers	☐ 1	☐ 2 97/
c. Added new supplementary materials	☐ 1	☐ 2 98/
d. Added new materials for ELLs (see definition of ELL on page 2)	☐ 1	☐ 2 99/
→ Check if no ELLs ☐ 1 100/		
e. Adopted new reading assessments	☐ 1	☐ 2 101/

D. Instructional Time

D1. Please indicate for which grades your school has a scheduled reading block.

> A **reading block** is the time period that is formally scheduled for teaching reading.

If yes, please indicate for how many minutes the reading block is scheduled.
Does your school have a reading block in:

	Yes	Scheduled number of minutes	No	
a. Kindergarten	☐ 1 →	☐☐☐	☐ 2	102/ 103-105/
b. First grade	☐ 1 →	☐☐☐	☐ 2	106/ 107-109/
c. Second grade	☐ 1 →	☐☐☐	☐ 2	110/ 111-113/
d. Third grade	☐ 1 →	☐☐☐	☐ 2	114/ 115-117/

E. Reading Interventions for Struggling Readers

E1. What methods has your school used to meet the needs of at-risk or **struggling readers**? For each method listed below, please check whether or not you use the method at your school.

A **reading intervention** is a program designed **for struggling readers** to be used only with struggling readers in addition to the core-reading program.

Methods for meeting needs of struggling readers		Check one for each item … Use this method	Not used
Materials	a. Use separate program materials in interventions	☐ 1	☐ 2 118/
	b. Use core reading program with supplemental materials	☐ 1	☐ 2 119/
	c. Use core reading program only	☐ 1	☐ 2 120/
	d. Use reading materials written in ELLs' home language (see definition of ELL on page 2)	☐ 1	☐ 2 121/
	e. Use alternative materials designed for ELLs	☐ 1	☐ 2 122/
	→ Check if no ELLs ☐ 1 123/		
Staff activities	f. A certified reading specialist provides additional direct instruction to struggling readers, individually or in small groups.	☐ 1	☐ 2 124/
	g. The classroom teacher provides additional direct instruction to struggling readers, individually or in small groups.	☐ 1	☐ 2 125/
	h. The classroom teacher provides additional opportunities for reading skill practice for struggling readers (e.g., partner reading, peer tutors, audio tapes, computer programs).	☐ 1	☐ 2 126/
	i. A certified specialist provides recommendations to classroom teachers on accommodations for struggling readers. (Indicate which type of specialist.)		
	A special education teacher	☐ 1	☐ 2 127/
	A bilingual/ESL teacher	☐ 1	☐ 2 128/
	Other (Please specify): _____ 130-131/	☐ 1	☐ 2 129/
	j. Trained aides or volunteers work with students under the direction of the classroom teacher during the school day.	☐ 1	☐ 2 132/
	k. Trained aides or volunteers work with students in a before or after school program.	☐ 1	☐ 2 133/
	l. Untrained aides or volunteers work with students under the direction of the classroom teacher during the school day.	☐ 1	☐ 2 134/
	m. Untrained aides or volunteers work with students in a before or after school program.	☐ 1	☐ 2 135/

F. Professional Development for Reading Coaches

F1. Below is a list of professional development topics for **reading coaches** in which you may have participated.

- In column A, identify any topics that were addressed in **reading coaches'** professional development activities **during the current school year, including summer 2006**.
- In column B, please identify the topics in which you need more professional development, whether or not this school's professional development activities have covered these topics.
- Please check all that apply in columns A and B.

Topics	CHECK ALL THAT APPLY	
	A. Topics addressed in professional development for reading coaches	**B. Topics in which you need more professional development**
a. How to use reading assessment data to guide instruction.	❑ 1 136/	❑ 2 137/
b. What are the types of assessments: screening, diagnostic, progress monitoring, and outcome?	❑ 3 138/	❑ 4 139/
c. How to use assessment data to form instructional groups.	❑ 1 140/	❑ 2 141/
d. How to provide constructive feedback to teachers.	❑ 3 142/	❑ 4 143/
e. How to establish credibility with teachers.	❑ 1 144/	❑ 2 145/
f. Essential components of scientifically based reading instruction.	❑ 3 146/	❑ 4 147/
g. What is the role of the reading coach in fostering change?	❑ 1 148/	❑ 2 149/
h. How to plan instructional interventions for struggling students.	❑ 3 150/	❑ 4 151/
i. Classroom management within the literacy block time.	❑ 1 152/	❑ 2 153/
j. How to conduct effective grade level meetings.	❑ 3 154/	❑ 4 155/
k. How to help teachers identify appropriate instructional materials.	❑ 1 156/	❑ 2 157/
l. How to help teachers make reading instruction systematic and explicit.	❑ 3 158/	❑ 4 159/
m. How to conduct demonstration lessons.	❑ 1 160/	❑ 2 161/
n. How to conduct classroom observations.	❑ 3 162/	❑ 4 163/
o. How to provide onsite professional development.	❑ 1 164/	❑ 2 165/
p. How to provide instructional supports for ELL students learning to read (see definition of ELL on page 2).	❑ 3 166/	❑ 4 167/
→ Check if no ELLs ❑ 1 168/		

G. Reading Instruction

G1. This item asks you to describe your school using the statements below. Please read each statement, and indicate whether the statement is a good description of your school on a scale from a "Very inaccurate" description of your school to a "Very accurate" description of your school.

In this school …	Check one box for each item… Very inaccurate ⟷ Very accurate					
a. K-3 teachers are knowledgeable about scientifically based reading instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	169/
b. K-3 teachers are motivated to improve reading instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	170/
c. Reading instruction in K-3 classrooms is aligned with the state reading/language arts content standards.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	171/
d. There is a school-wide focus on reading and language arts.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	172/
e. K-3 teachers are experienced with the core reading program.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	173/
f. K-3 teachers are experienced with supplemental reading materials.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	174/
g. K-3 teachers are experienced with reading intervention materials and strategies.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	175/
h. K-3 classrooms have ample, high quality instructional materials.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	176/
i. Teachers use a variety of instructional materials to fill in gaps in the core program.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	177/
j. The core reading program is aligned with scientifically based reading research.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	178/
k. Supplemental reading materials are aligned with scientifically based reading research.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	179/
l. Reading intervention materials are aligned with scientifically based reading research.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	180/
m. The reading coach has the support of the school principal.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	181/
n. K-3 teachers seek the assistance of the reading coach to improve their reading instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	182/
o. Sufficient time during the school day is allotted for reading instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	183/
p. Sufficient time during the school day is allotted for teacher planning.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	184/
q. K-3 teachers collaborate and plan for reading instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	185/
r. Sufficient time during the school day is allotted for professional development.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	186/
s. Reading assessments are used to screen students for reading difficulties.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	187/
t. Diagnostic assessments are used to identify strengths and weaknesses of struggling readers.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	188/
u. Reading assessments are used to monitor student progress.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	189/
v. Assessment data are used to group students for instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	190/
w. Assessment data are used to guide and/or modify instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	191/
x. The district provides direction concerning reading instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	192/
y. The state provides direction concerning reading instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	193/
z. K-3 teachers make an effort to involve parents in their children's reading instruction.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	194/

H. Out of School Reading

H1. What school-wide strategies do you use to encourage students to read outside of school?

Strategies to encourage students to read outside of school	Check only one box for each item … YES	NO
a. School-wide book-reading contests	❑ 1	❑ 2 384/
b. Grade-specific contests	❑ 1	❑ 2 385/
c. Book giveaways with parents about reading resources	❑ 1	❑ 2 386/
d. Use reading materials written in students' home language	❑ 1	❑ 2 387/
e. Relationship with community libraries	❑ 1	❑ 2 388/
f. School-wide book clubs	❑ 1	❑ 2 389/
g. Partnering students who live near each other to read to each other out of school	❑ 1	❑ 2 390/
h. Principal, reading coach or librarian awards/prizes for individual readers	❑ 1	❑ 2 391/
i. School-wide book celebrations	❑ 1	❑ 2 392/
j. School-wide rewards for reaching reading milestones	❑ 1	❑ 2 393/

Thank you for your cooperation and for taking the time to answer these questions. Please place the completed survey in the enclosed envelope, seal the envelope, and return to your school's Reading First Impact Study liaison.

If you have any questions about the survey, please call 1-866-421-6982 and leave a message, or send an e-mail to reading_impact@abtassoc.com

Abt ID // barcode

1-6/

Abt ID / barcode here

Reading First Impact Study
Teacher Survey

Abt Associates has been commissioned by the Institute of Education Sciences (IES) at the U.S. Department of Education to conduct an independent national evaluation of the Reading First program. Your school is one of only 250 elementary schools that have been selected to participate in this study, so your participation is extremely important. The study includes two kinds of schools: half of which have received Reading First funding, and half of which have not received Reading First funding. We know that teaching children how to read is important in every single school, and that is why we are asking teachers, staff, and administrators in both kinds of schools to describe the reading instruction in their schools and classrooms. Your responses will help inform the U.S. Department of Education, Congress, policymakers, practitioners, and researchers about how reading instruction is implemented in schools and what strategies schools use to provide high-quality, evidence-based reading instruction in first, second, and third grades.

Additional Information

The survey will take you approximately 30 minutes to complete. All responses to the survey will be kept confidential. All individual identifying information will be used only by persons on the research team. Information such as school location (state), participants' general job titles, grades they teach, and gender will be included in the study data files to be submitted to the Department of Education. However, participants' names will be stripped from all analysis data files and data files to be submitted to the Department of Education. We will not report any data about individual classrooms—all information will be reported at the grade and school levels. Neither your school nor your district will have access to any of the completed surveys at any time.

Thank you for your cooperation with this survey!

Instructions

Unless otherwise noted, your responses should reflect your experiences during the 2006-2007 school year in the school to which this survey was sent.

- Please complete all questions; each question includes directions for recording your answer.

- You are sometimes told to skip over some questions in the survey. When this happens, you will see an arrow with a note that tells you what question to answer like this:

  ❑1 Yes
  ❑2 No → Skip to E4

- If you have any questions about how to complete the survey, please call: 1-866-421-6982. This is a free call and will connect you with our expert interviewers who can assist you.

A. Your Background and Experience

A1. Including this year, how long have you been a teacher? (If less than one year, please enter '1')

	Enter # of years below…
a. Total number of years as a teacher	_____ **Years** 10-11/
b. Number of years teaching at this school	_____ **Years** 12-13/

A2. What grade(s) are you currently teaching? (Check all that apply)

- ❑$_1$ Kindergarten 14/
- ❑$_2$ First grade 15/
- ❑$_3$ Second grade 16/
- ❑$_4$ Third grade 17/
- ❑$_{95}$ Other (Please specify): _____ 18-19/ 20-21/ 22-23/ 24-25/

A3. How well do you feel your **pre-service teacher training** prepared you to teach each of the following dimensions of reading?

> **Pre-service teacher training** refers to training you received before you became certified and began teaching. For those who began their teacher career through an alternative certification or emergency certification program, and began teaching before they were certified, pre-service teacher training refers to the training you received to become fully certified.

Please choose a '1' if you were 'not at all prepared' to teach the dimension and a '5' if you were 'extremely well prepared.'

	Check only one box for each item…				
	Not at all prepared		**Somewhat well prepared**		**Extremely well prepared**
a. Phonemic awareness	❑$_1$	❑$_2$	❑$_3$	❑$_4$	❑$_5$ 26/
b. Decoding	❑$_1$	❑$_2$	❑$_3$	❑$_4$	❑$_5$ 27/
c. Vocabulary	❑$_1$	❑$_2$	❑$_3$	❑$_4$	❑$_5$ 28/
d. Comprehension	❑$_1$	❑$_2$	❑$_3$	❑$_4$	❑$_5$ 29/
e. Fluency building	❑$_1$	❑$_2$	❑$_3$	❑$_4$	❑$_5$ 30/

B. Instruction and Assessment in Reading

B1. **Last week**, approximately how many minutes per day did you devote to reading instruction? Include only reading instruction and not other language arts such as writing, spelling. Fill in the chart for each day last week with your best estimate of the number of minutes…

Monday	Tuesday	Wednesday	Thursday	Friday
# _____ min. / day 31-33/	# _____ min. / day 34-36/	# _____ min. / day 37-39/	# _____ min. / day 40-42/	# _____ min. / day 43-45/

B2. Has the average number of minutes you spend each day this year teaching reading decreased, remained the same, or increased from last year (2005-2006)? Please check one.

- ❑1 **I did not teach reading last year** 46/
- ❑2 Decreased
- ❑3 Remained the same
- ❑4 Increased

B3. How often **during this school year** is time regularly scheduled and formally set aside during the school day for Grade 2-3 teachers to:

	Check only one box for each item …					
	Not at all	**1-4 times**	**5-8 times**	**Once a month**	**Once a week or more**	**Occurs only informally, as needed**
a. Collaborate on reading lesson planning and instruction.	❑1	❑2	❑3	❑4	❑5	❑6 47/
b. Observe reading instruction in other classrooms.	❑1	❑2	❑3	❑4	❑5	❑6 48/
c. Use assessment data to plan instruction.	❑1	❑2	❑3	❑4	❑5	❑6 49/
d. Participate in coaching with or be coached about reading by other teachers.	❑1	❑2	❑3	❑4	❑5	❑6 50/
e. Be coached about my reading instruction by a reading coach (see below).	❑1	❑2	❑3	❑4	❑5	❑6 51/
f. Coordinate reading interventions for struggling readers with special education staff.	❑1	❑2	❑3	❑4	❑5	❑6 52/
g. Coordinate reading interventions for struggling readers with English language learner (ELL) staff (see below).	❑1	❑2	❑3	❑4	❑5	❑6 53/
→ Check if no ELLs ❑1						54/

A **reading coach** is a staff member whose primary role is to **provide ongoing training and support to classroom teachers** in the delivery of effective reading instruction. This assistance may include planning instruction, providing demonstration lessons, observing and providing feedback, using assessment results to guide instruction, etc.

English language learner (ELL) indicates a student who is in the process of acquiring English and **has a first language other than English.** Other common related terms include language minority or limited English proficient (LEP) students, students in English as a second language (ESL), or students in classes for English for speakers of other languages (ESOL).

B4. Please describe your use of the following reading instructional activities **this year**.

- **Check column A ONLY** if the instructional activity is one that you use frequently when you teach reading or one on which you rely heavily in your reading instruction.
- **Check column B** if you use the instructional activity, but it is a small part of your teaching, and not one you use frequently. It might be an activity that you use if there is time, but it is not one on which you rely heavily for your reading instruction.
- **Check column C** if the activity is not one you use in your reading instruction.

		Check only one box for each item…		
		A **Central to my reading instruction**	**B** **Small part of my reading instruction**	**C** **Not part of my reading instruction**
Reading Text	a. Students read texts that are easy to decode.	☐ 1	☐ 2	☐ 3 55/
	b. Students read to locate information.	☐ 1	☐ 2	☐ 3 56/
	c. Students read aloud unfamiliar text.	☐ 1	☐ 2	☐ 3 57/
	d. Students reread familiar stories.	☐ 1	☐ 2	☐ 3 58/
	e. Students read aloud together.	☐ 1	☐ 2	☐ 3 59/
	f. Students read silently.	☐ 1	☐ 2	☐ 3 60/
	g. I listen to students read aloud without correcting errors.	☐ 1	☐ 2	☐ 3 61/
	h. I listen to students read aloud and correct errors immediately.	☐ 1	☐ 2	☐ 3 62/
	i. Students confirm or revise predictions after reading.	☐ 1	☐ 2	☐ 3 63/
	j. Students generate their own questions about text material.	☐ 1	☐ 2	☐ 3 64/
	k. Students identify their comprehension break-downs and use fix-up strategies with a partner.	☐ 1	☐ 2	☐ 3 65/
	l. Students orally summarize main events in stories and informational texts.	☐ 1	☐ 2	☐ 3 66/
	m. Students use graphic and semantic organizers to track information.	☐ 1	☐ 2	☐ 3 67/
Work with sounds and words	n. Students decode multi-syllabic words in isolation.	☐ 1	☐ 2	☐ 3 84/
	o. I teach decoding skills while reading stories.	☐ 1	☐ 2	☐ 3 85/
	p. Students memorize sight words.	☐ 1	☐ 2	☐ 3 86/
	q. Students read irregularly spelled words and non-words.	☐ 1	☐ 2	☐ 3 87/
	r. Students practice reading high frequency words for automaticity.	☐ 1	☐ 2	☐ 3 88/
	s. Students use knowledge of root words, prefixes, and suffixes to decode new words.	☐ 1	☐ 2	☐ 3 89/
	t. Students work with prefixes and suffixes to change the meaning of words.	☐ 1	☐ 2	☐ 3 90/
	u. I stop students while reading and have them self-correct misidentified words.	☐ 1	☐ 2	☐ 3 91/
	v. Students use context clues to identify unknown words.	☐ 1	☐ 2	☐ 3 92/
	w. Students practice writing words as separate syllables.	☐ 1	☐ 2	☐ 3 93/

68-83/BLANK

94-107/BLANK

B4. CONTINUED. Please describe your use of the following reading instructional activities **this year**.

- **Check column A ONLY** if the instructional activity is one that you use frequently when you teach reading or one on which you rely heavily in your reading instruction.
- **Check column B** if you use the instructional activity, but it is a small part of your teaching, and not one you use frequently. It might be an activity that you use if there is time, but it is not one on which you rely heavily for your reading instruction.
- **Check column C** if the activity is not one you use in your reading instruction.

		Check only one box for each item…		
		A Central to my reading instruction	**B Small part of my reading instruction**	**C Not Part of my reading instruction**
Other Techniques	x. Students learn vocabulary through study of antonyms, synonyms, and homonyms.	☐ 1	☐ 2	☐ 3 108/
	y. Students learn vocabulary through study of word categories.	☐ 1	☐ 2	☐ 3 109/
	z. Students write vocabulary words in sentences.	☐ 1	☐ 2	☐ 3 110/
	aa. Students use dictionaries to find word meanings.	☐ 1	☐ 2	☐ 3 111/
	bb. I discuss new and unusual words before reading.	☐ 1	☐ 2	☐ 3 112/
	cc. Students answer questions in writing after reading stories.	☐ 1	☐ 2	☐ 3 113/
	dd. Students identify story structure and elements.	☐ 1	☐ 2	☐ 3 114/
	ee. Students read stories they have written to others.	☐ 1	☐ 2	☐ 3 115/
	ff. Students participate in literature extensions, (e.g., book talks, plays, readers' theaters).	☐ 1	☐ 2	☐ 3 116/
	gg. Students select books from the library for independent reading.	☐ 1	☐ 2	☐ 3 117/
	hh. Students are given time to read on their own for enjoyment.	☐ 1	☐ 2	☐ 3 118/
	ii. Class creates group stories.	☐ 1	☐ 2	☐ 3 119/

120-134/BLANK

B5. Please describe your use of the following teaching strategies and materials **this year**.

- **Check column A ONLY** if the item is one that you use frequently or one on which you rely heavily in your reading instruction.
- **Check column B** if you use the item, but it is a small part of your teaching, and not one you use frequently. It may be an approach you use if there is time, but it is not one on which you rely heavily.
- **Check column C** if the item is not one you use in your reading instruction.

		Check only one box for each item…		
		A **Central to my reading instruction**	**B** **Small part of my reading instruction**	**C** **Not Part of my reading instruction**
Instruction	a. Provide time in reading block for skill practice on own.	☐ 1	☐ 2	☐ 3 135/
	b. Provide materials for at-home practice of skills introduced in class.	☐ 1	☐ 2	☐ 3 136/
	c. Provide extra reading instructional time for struggling students.	☐ 1	☐ 2	☐ 3 137/
	d. Include writing opportunities in reading instruction.	☐ 1	☐ 2	☐ 3 138/
	e. Build spelling practice into reading instruction.	☐ 1	☐ 2	☐ 3 139/
	f. Develop reading skills through science and social studies.	☐ 1	☐ 2	☐ 3 140/
Grouping	g. Teach whole class reading lessons.	☐ 1	☐ 2	☐ 3 141/
	h. Work one-to-one with students on reading.	☐ 1	☐ 2	☐ 3 142/
	i. Work with small groups of students.	☐ 1	☐ 2	☐ 3 143/
	j. Group students based on skill levels.	☐ 1	☐ 2	☐ 3 144/
	k. Group students based on mixed abilities (cooperative groups).	☐ 1	☐ 2	☐ 3 145/
	l. Pair strong readers with those with weaker skills.	☐ 1	☐ 2	☐ 3 146/
Reading materials	m. Use core reading series.	☐ 1	☐ 2	☐ 3 147/
	n. Use supplementary reading materials for instruction in the following areas:			
	1. Phonemic awareness	☐ 1	☐ 2	☐ 3 148/
	2. Phonics	☐ 1	☐ 2	☐ 3 149/
	3. Fluency	☐ 1	☐ 2	☐ 3 150/
	4. Vocabulary	☐ 1	☐ 2	☐ 3 151/
	5. Comprehension	☐ 1	☐ 2	☐ 3 152/
	o. Use children's trade books.	☐ 1	☐ 2	☐ 3 153/
	p. Use books that are easy to decode.	☐ 1	☐ 2	☐ 3 154/
	q. Use books with patterned predictable language.	☐ 1	☐ 2	☐ 3 155/
	r. Use separate intervention materials for some students.	☐ 1	☐ 2	☐ 3 156/
	s. Use reading software/technology.	☐ 1	☐ 2	☐ 3 157/
	t. Use teacher-made materials.	☐ 1	☐ 2	☐ 3 158/

Supplementary Reading Materials provide additional instruction in a targeted area of reading to **all** students. **Do not** include materials that are used only with struggling readers. Include teacher-made materials, if applicable.

B5. CONTINUED. Please describe your use of the following teaching strategies and materials **this year**.

- **Check column A ONLY** if the item is one that you use frequently or one on which you rely heavily in your reading instruction.
- **Check column B** if you use the item, but it is a small part of your teaching, and not one you use frequently. It may be an approach you use if there is time, but it is not one on which you rely heavily.
- **Check column C** if the item is not one you use in your reading instruction.

		Check only one box for each item…		
		A **Central to my reading instruction**	**B** **Small part of my reading instruction**	**C** **Not Part of my reading instruction**
Assessments	u. Use test results to organize instructional groups.	☐ 1	☐ 2	☐ 3 159/
	v. Use informal reading inventories.	☐ 1	☐ 2	☐ 3 160/
	w. Use tests to determine progress on skills.	☐ 1	☐ 2	☐ 3 161/
	x. Use tests to determine who can benefit from the core reading series.	☐ 1	☐ 2	☐ 3 162/
	y. Use diagnostic tests to identify students who need reading intervention services.	☐ 1	☐ 2	☐ 3 163/
	z. Use screening tests to identify students who need a supplementary reading program.	☐ 1	☐ 2	☐ 3 164/
	aa. Conduct miscue analysis, analyzing errors students make while reading aloud.	☐ 1	☐ 2	☐ 3 165/

B6. What specific formal or informal assessments do you primarily use for placing and /or grouping students, determining student mastery of skills, and identifying core deficits of struggling students? If you use more than one assessment, please report only on the one that you use the most often. Please be as specific as possible when naming or describing the assessment(s).

- In column A enter the name of the primary assessment used for each purpose.
- In column B check the number of times the assessment is given during the school year.
- In column C check whether students are usually assessed individually, in small groups, or in a whole class.
- In column D check the average time that it takes to administer the assessment.
- In column E check whether the assessment is used for accountability purposes for the Reading First program, No Child Left Behind (NCLB), or another program. Please check all that apply in this column only.

	Check only one box per column for each item…			**Check all that apply…**
A. Primary purposes and names of assessments	**B. Number of times given per school year** *(Check one)*	**C. Students are assessed…** *(Check one)*	**D. Average time it takes to administer assessment** *(Check one)*	**E. Accountability purposes** *(Check all that apply)*
a. Placement and/or grouping students *(Check one):* 166/ 1☐ Assessment: _____ _____ 167-168/ 169-170/ 171-172/ 2☐ I do not use an assessment for this purpose	1☐ 1 2☐ 2 3☐ 3 4☐ 4 5☐ 5 or more 173/	1☐ Individually 2☐ In small groups 3☐ In whole class 174/	1☐ 1-15 *minutes* 2☐ 16-30 3☐ 31-45 4☐ 46-60 5☐ 61 or more 175/	1☐ Reading First 176/ 2☐ NCLB 177/ 3☐ Other 178/
b. Determining student mastery of skills *(Check one):* 179/ 1☐ Assessment: _____ _____ 180-181/ 182-183/ 184-185/ 2☐ I do not use an assessment for this purpose	1☐ 1 2☐ 2 3☐ 3 4☐ 4 5☐ 5 or more 186/	1☐ Individually 2☐ In small groups 3☐ In whole class 187/	1☐ 1-15 *minutes* 2☐ 16-30 3☐ 31-45 4☐ 46-60 5☐ 61 or more 188/	1☐ Reading First 189/ 2☐ NCLB 190/ 3☐ Other 191/
c. Identifying the core deficits of struggling students *(Check one):* 192/ 1☐ Assessment: _____ _____ 193-194/ 195-196/ 197-198/ 2☐ I do not use an assessment for this purpose	1☐ 1 2☐ 2 3☐ 3 4☐ 4 5☐ 5 or more 199/	1☐ Individually 2☐ In small groups 3☐ In whole class 200/	1☐ 1-15 *minutes* 2☐ 16-30 3☐ 31-45 4☐ 46-60 5☐ 61 or more 201/	1☐ Reading First 202/ 2☐ NCLB 203/ 3☐ Other 204/

B7. What materials are used with **English Language Learners (ELLs)** to whom you teach reading (see definition of ELL on page 2)?
(Check all that apply)

- ❑$_1$ **Do not teach ELLs →Skip to B8** 205/
- ❑$_2$ Core reading program materials in the native language of the ELL 206/
- ❑$_3$ ELL students use the same materials as other students 207/
- ❑$_4$ Core reading program materials, plus supplementary/intervention resources written in the ELL's native language 208/
- ❑$_5$ Core reading program materials, plus supplementary/intervention resources written in English especially for ELLs 209/
- ❑$_6$ Alternative core reading program materials in English geared toward the instructional level of the ELL 210/

B8. What additional supports have students who are struggling readers received **in the last month**? Check whether or not your students who are struggling readers received each of the supports during the past month.

Supports for Struggling Readers	Check only one box for each item… Received	Did not receive	
a. Diagnostic assessment to determine core deficits.	☐ 1	☐ 2	211/
b. Extra practice in the classroom with phonemic awareness.	☐ 1	☐ 2	212/
c. Extra practice in the classroom with decoding.	☐ 1	☐ 2	213/
d. Extra practice in the classroom with fluency.	☐ 1	☐ 2	214/
e. Extra practice in the classroom with comprehension.	☐ 1	☐ 2	215/
f. Extra instructional time.	☐ 1	☐ 2	216/
g. Placement in materials that supplement the core reading program.	☐ 1	☐ 2	217/
h. Placement in different level of core reading program.	☐ 1	☐ 2	218/
i. Placement in separate core reading program.	☐ 1	☐ 2	219/
j. Placement in special intervention program.	☐ 1	☐ 2	220/
k. Work with tutor on one-to-one basis.	☐ 1	☐ 2	221/
l. Work with reading specialist on one-to-one basis.	☐ 1	☐ 2	222/
m. Work with reading specialist in small group.	☐ 1	☐ 2	223/
n. Work with more advanced peer.	☐ 1	☐ 2	224/
o. Special materials for parents to provide practice.	☐ 1	☐ 2	225/

B9. What additional supports have students who are struggling readers *and* ELLs received **in the last month**? Check whether or not your students who are struggling readers and ELLs received each of the supports during the past month (see definition of ELL on page 2).

Check if no ELLs → ☐ 1 Skip to C1 226/

Supports for Struggling Readers who are ELLs	Check only one box for each item… Received	Did not receive	
a. English as a Second Language instruction.	☐ 1	☐ 2	227/
b. Provide reading instruction in home language.	☐ 1	☐ 2	228/
c. In classroom help in reading from ELL teacher.	☐ 1	☐ 2	229/

C. Professional Development in Reading for K-3 Teachers

C1. During the current school year, including summer 2006, did you participate in any professional development activities in reading?

230/

Yes ☐1
No ☐2 ⟶ Skip to C2

If yes, in how many of each of the following types of professional development activities **in reading** have you participated? Please count each activity only once. What is the total number of hours you spent in these activities?

First, write in the number of activities of each type in which you have been engaged. Then, write the total number of hours you spent in these activities. Mark 0 if you participated in none.

	Enter # and hours below…	
	# of Different workshops	**Total hours**
a. Attended short, stand-alone training or workshop in reading (half-day or less)	#_____ 231-232/	_____**Hours** 233-235/
b. Attended longer institute or workshop in reading (more than half-day)	#_____ 236-237/	_____**Hours** 238-240/
c. Attended a college course in reading (include any courses you are currently attending)	#_____ 241-242/	_____**Hours** 243-245/
d. Attended a conference about reading (might include multiple short offerings)	#_____ 246-247/	_____**Hours** 248-250/

C2. Below is a list of professional development activities that are often used to provide ongoing, direct support to teachers for teaching reading.

- In the first column, please indicate whether you have received any of the following types of assistance/support for teaching **during the current school year, including summer 2006**.
- If you did not receive that type of support, please indicate whether the support was available, but you did not receive it (column 2), or if it was not available at your school (column 3).

	Check only one box for each item …		
	Types of assistance I received this year	**Available, but I did not receive**	**Not available at my school**
a. Coaching or mentoring by reading coach in programs, materials, or strategies.	☐ 1	☐ 2	☐ 3 251/
b. Coaching or mentoring from fellow teacher.	☐ 1	☐ 2	☐ 3 252/
c. Peer study group or collegial circle for group study.	☐ 1	☐ 2	☐ 3 253/
d. Demonstrations in my classroom.	☐ 1	☐ 2	☐ 3 254/
e. Observations of other teachers.	☐ 1	☐ 2	☐ 3 255/
f. Diagnostic testing help from a reading coach or specialist for individual students.	☐ 1	☐ 2	☐ 3 256/
g. Intervention service help from a reading coach or specialist for individual students.	☐ 1	☐ 2	☐ 3 257/
h. Interpretation of assessment data.	☐ 1	☐ 2	☐ 3 258/
i. Grade level meetings devoted to reading.	☐ 1	☐ 2	☐ 3 259/
j. Using assessment data to determine topics that require additional instruction or practice.	☐ 1	☐ 2	☐ 3 260/

C3. During the current school year, including summer 2006, **approximately** how many of the **reading professional development activities** for **K-3 teachers**: (Please choose the category that most closely describes your professional development.)

	Check only one box for each item…				
	None	**One-Quarter**	**One-Half**	**Three-Quarters**	**All**
a. were also attended by the principal?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 261/
b. provide teachers options among which to choose?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 262/
c. provide a stipend?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 263/
d. provide follow-up activities?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 264/
e. include release time for participating teachers?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 265/
f. offer graduate college credits?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 266/
g. are held in a convenient location (e.g., activities held at school)?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 267/
h. use a team-based approach (joint training of people who work together)?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 268/
i. are given by trainers or facilitators who have a well-established reputation?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 269/

C4. Below is a list of topics that are often covered in professional development activities designed to provide teachers with new information about the **content of reading instruction**.

- In column A, identify the topics that were addressed in professional development activities in which you participated **during the current school year, including summer 2006**.
- In column B, please identify the topics in which you need more professional development, whether or not this school's professional development activities have covered these topics.
- Please check all that apply in columns A and B.

Professional development is defined as any activity in which a teacher has learned about reading or reading instruction. This includes school-based workshops, meetings with reading coaches, and meetings with a study group of other teachers.

		CHECK ALL THAT APPLY	
		A. Topics addressed in professional development	**B. Topics in which I need more professional development**
Phonemic Awareness	a. Building phonological awareness, e.g. rhymes, dividing spoken language into sentences, words, syllables	☐ 1 270/	☐ 2 271/
	b. Identifying, adding, deleting sounds in spoken words	☐ 3 272/	☐ 4 273/
	c. Blending phonemes to form words	☐ 1 274/	☐ 2 275/
Decoding	d. Teaching letter-sound correspondence	☐ 3 276/	☐ 4 277/
	e. Teaching letter patterns (blends, digraphs, diphthongs)	☐ 1 278/	☐ 2 279/
	f. Using syllable patterns to read words	☐ 3 280/	☐ 4 281/
	g. Teaching component parts: roots, prefixes, suffixes	☐ 1 282/	☐ 2 283/
Vocabulary	h. Teaching use of dictionary, thesaurus	☐ 3 284/	☐ 4 285/
	i. Direct teaching of vocabulary words and their meaning	☐ 1 286/	☐ 2 287/
	j. Antonyms and synonyms	☐ 3 288/	☐ 4 289/
Fluency	k. Teaching sight words	☐ 1 290/	☐ 2 291/
	l. Guided oral reading	☐ 3 292/	☐ 4 293/
	m. Encouraging expression while reading	☐ 1 294/	☐ 2 295/
Comprehension	n. Setting motivation/asking prediction/preview questions	☐ 3 296/	☐ 4 297/
	o. Constructing information about character, setting, and main events	☐ 1 298/	☐ 2 299/
	p. Summarizing main ideas in narrative and informational text	☐ 3 300/	☐ 4 301/
	q. Self-monitoring strategies	☐ 1 302/	☐ 2 303/
	r. Asking questions at different levels (literal, inferential)	☐ 3 304/	☐ 4 305/
	s. Strategies for organizing text structure, e.g. story maps	☐ 1 306/	☐ 2 307/

308-309/BLANK

C5. Below is a list of topics that are often covered in professional development activities that are designed to provide teachers with new information about **teaching strategies used during reading instruction**.

- In column A, identify the topics that were addressed in professional development activities in which you participated **during the current school year, including summer 2006**.
- In column B, please identify the topics in which you need more professional development, whether or not this school's professional development activities have covered these topics.
- Please check all that apply in columns A and B.

Topic		CHECK ALL THAT APPLY A. Topics addressed in professional development.	B. Topics in which I need more professional development
Teaching Strategies	a. How to use the core reading program	☐ 1 310/	☐ 2 311/
	b. How to use children's literature to teach reading	☐ 3 312/	☐ 4 313/
	c. How to use reading research to guide content of instruction	☐ 1 314/	☐ 2 315/
	d. How the core reading program incorporates research principles	☐ 3 316/	☐ 4 317/
	e. How to use the supplemental reading program(s)	☐ 1 318/	☐ 2 319/
	f. How to integrate reading and writing instruction	☐ 3 320/	☐ 4 321/
	g. Strategies for teaching reading to ELLs (see definition of ELL on page 2)	☐ 1 322/	☐ 2 323/
Grouping	h. Learning styles	☐ 3 324/	☐ 4 325/
	i. How to organize small group instruction	☐ 1 326/	☐ 2 327/
Assessment	j. How to diagnose reading problems	☐ 3 328/	☐ 4 329/
	k. How to administer assessments	☐ 1 330/	☐ 2 331/
	l. How to interpret and use assessment data to guide instruction	☐ 3 332/	☐ 4 333/
Struggling Readers	m. How to help struggling readers with decoding	☐ 1 334/	☐ 2 335/
	n. How to help struggling readers with vocabulary	☐ 3 336/	☐ 4 337/
	o. How to help struggling readers with comprehension	☐ 1 338/	☐ 2 339/
	p. How to motivate readers	☐ 3 340/	☐ 4 341/
	q. Strategies for teaching reading to students with diagnosed learning disabilities	☐ 1 342/	☐ 2 343/
Organization/ planning	r. How to use state/district content standards for curriculum planning and teaching	☐ 3 344/	☐ 4 345/
	s. How to align reading curriculum and instruction with state/district assessments	☐ 1 346/	☐ 2 347/
	t. How to work with parents	☐ 3 348/	☐ 4 349/
	u. Classroom management	☐ 1 350/	☐ 2 351/

C6. How well do you feel the **professional development activities in which you participated during the current school year (including summer 2006)** prepared you to teach each of the following dimensions of reading? Please choose a '1' if you were 'not at all prepared' to teach the dimension and a '5' if you were 'extremely well prepared.'

	Check only one box for each item …				
	Not at all prepared		**Somewhat well prepared**		**Extremely well prepared**
a. Phonemic awareness	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 352/
b. Decoding	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 353/
c. Vocabulary	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 354/
d. Comprehension	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 355/
e. Fluency building	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5 356/

D. Support for Teaching Reading

D1. The next set of statements is about your reading program. Please indicate the extent to which you agree or disagree with each statement.

	Check one box for each item …			
	Strongly Agree	**Agree**	**Disagree**	**Strongly Disagree**
a. I feel I need to make changes in the methods I use to teach children to read.	☐ 1	☐ 2	☐ 3	☐ 4 357/
b. Other faculty/staff members have helped me to understand the difficulties that some children have in learning to read.	☐ 1	☐ 2	☐ 3	☐ 4 358/
c. I have benefited from opportunities to learn more about methods for teaching reading.	☐ 1	☐ 2	☐ 3	☐ 4 359/
d. The children in my class are making satisfactory progress in learning to read.	☐ 1	☐ 2	☐ 3	☐ 4 360/
e. I do not have sufficient materials to teach reading effectively.	☐ 1	☐ 2	☐ 3	☐ 4 361/
f. I do not understand why some children learn to read easily while other children struggle to learn basic reading skills.	☐ 1	☐ 2	☐ 3	☐ 4 362/
g. The reading coach supports my efforts to teach reading effectively.	☐ 1	☐ 2	☐ 3	☐ 4 363/
h. I have a good understanding of how children acquire language and literacy skills.	☐ 1	☐ 2	☐ 3	☐ 4 364/
i. I wish I had more opportunities to discuss how to teach reading with other teachers.	☐ 1	☐ 2	☐ 3	☐ 4 365/
j. I know the current reading skill levels of all my students.	☐ 1	☐ 2	☐ 3	☐ 4 366/
k. I know how to assess the progress of my students in reading.	☐ 1	☐ 2	☐ 3	☐ 4 367/
l. I have changed my methods of teaching reading as a result of professional development in reading.	☐ 1	☐ 2	☐ 3	☐ 4 368/

E. Out of School Reading

E1. What strategies do you use in your classroom to encourage students to read outside of school?

Strategies to encourage students to read outside of school	Check only one box for each item… Yes	No
a. Classroom-wide reading contests	☐ 1	☐ 2 389/
b. School-wide reading contests	☐ 1	☐ 2 390/
c. Book giveaways or auctions	☐ 1	☐ 2 391/
d. Individual rewards (e.g., coupons, prizes) for reading a certain number of books or pages outside of school	☐ 1	☐ 2 392/
e. Class rewards (e.g., pizza party, "choice time") for reaching a class goal for out-of-school reading	☐ 1	☐ 2 393/
f. Communication with parents about reading resources	☐ 1	☐ 2 394/
g. Classroom relationship with community libraries	☐ 1	☐ 2 395/
h. Use of books or book series students can continue to read at home	☐ 1	☐ 2 396/
i. Partnering students who live near each other to read to each other out of school	☐ 1	☐ 2 397/
j. Classroom book clubs discussing books read outside of school	☐ 1	☐ 2 398/
k. Tape-reading students reading favorite parts of books	☐ 1	☐ 2 399/
l. Inviting community members to speak of their recreational reading	☐ 1	☐ 2 400/
m. Asking students to read to siblings and/or parents at home	☐ 1	☐ 2 401/
n. Using classroom resources for students to buy books through children's book publisher	☐ 1	☐ 2 402/
o. Having students write and illustrate their own books they bring home to read	☐ 1	☐ 2 403/
p. Sharing your own recreational reading with the class	☐ 1	☐ 2 404/
q. Providing time for students to retell stories they've read	☐ 1	☐ 2 405/
r. Providing opportunities for students to act out or role-play stories they've read	☐ 1	☐ 2 406/

E2. Do you assess students' out-of-school reading?

- ☐ 1 Yes→**Skip to E2a** 410/
- ☐ 2 No

E2a. If yes, how do you assess students' out-of school reading? (Check all that apply)

- ☐ 1 Reading logs 411/
- ☐ 2 Book reports 412/
- ☐ 3 Parent verficiation 413/
- ☐ 95 Other (Please specify): _____ 414-415/

Thank you for your cooperation and for taking the time to answer these questions. Please place the completed survey in the enclosed envelope, seal the envelope, and return to your school's Reading First Impact Study liaison.

If you have any questions about the survey, please call 1-866-421-6982 and leave a message, or send an e-mail to reading_impact@abtassoc.com

Exhibit C.19: Instructional Practice in Reading Inventory (IPRI)

Instructional Practice in Reading Inventory (IPRI)

Part A. Dimensions of Reading | **Interval 1 Time: ___ : ___ AM PM**

DP Decoding with PRINT (Phonics)

Grouping		T. Instruction: The Teacher. . .		S. Student Practice: The Teacher . . .	
W	Whole class	T1	Describes, explains, identifies or asks S(s) to describe, explain or identify • *sound-symbol pattern* • *decoding rule* • *word structure pattern or rule*	S1	Gives S(s) practice decoding words (PRINT TO SOUND)
L	Lg grp			S2	Gives S(s) practice encoding words (manipulating letters or writing, not copying) (SOUND TO PRINT)
S	Sm grp				
P	Pair			S3	Asks student(s) to orally spell word(s)
I	Individual				
		T2	Shows how to apply a rule or pattern to whole word example(s)		
Support		T3	Identifies word(s) that contrast with or do not follow pattern or rule		
TM	Tchr Manipulatives	T4	***If S makes mistake***:		
V	Picture, Object		Reminds S of pattern or rule and has S produce or repeat correct response		
ST	Sentence(s)/Text			**GH**	**Grammar or Handwriting is part of lesson**
SM	Std Manipulatives	**PA**	**Includes Phonemic Awareness example**	**OR**	**Student Oral Reading is part of lesson**

CP Comprehension of Connected Text

Grouping		T. Instruction: The Teacher. . .		S. Student Practice: The Teacher . . .	
W	Whole class		***Before reading text:***		***During or after reading text passage***:
L	Lg grp	T1	Conducts pre-reading activity(ies)	S1	Asks students to answer literal recall questions about specific details in the text
S	Sm grp	T1a	Previews vocabulary prior to lesson (Go to Vocab)		
P	Pair		***Before, during, or after reading text passage:***		
I	Individual	T2	Describes or explains—or asks Ss to describe or explain—one or more comprehension strategies	S2	Asks students to identify or describe genre
		T2a	Specifies what the strategy is called		
Support		T2b	Specifies why the strategy is helpful		
		T2c	Specifies when in the reading process the strategy is used		
V	Picture, Object		***During or after reading text passage*: Shows how to apply strategy**		***During or after reading text passage*: Gives S(s) practice applying strategy**
O	Text Organizer		Using cues to support interpretation or make predictions about text:		Using cues to support interpretation or make predictions about text:
		T3a1	Pictures	S3a1	Pictures
C	Connected text	T3a2	Text cues (e.g., headers, captions)	S3a2	Text cues (e.g., headers, captions)
E	Expository	T3b	Answer inferential questions based on text	S3b	Answer inferential questions based on text
N	Narrative	T3c	Make predictions based on text	S3c	Make predictions based on text
CD	Can't determine	T3d	Summarize, retell, sequence text or identify main idea(s)	S3d	Summarize, retell, or sequence text or identify main idea(s)
		T3e	Make text-text connections	S3e	Make text-text connections
		T3f	Work with story or expository structure	S3f	Work with story or expository structure
		T3g	Use mental imagery to support interpretation of text	S3g	Use mental imagery to support interpretation of text
		T3h	Generate own questions about text	S3h	Generate own questions about text
		T3i	Answer own questions about text	S3i	Answer own questions about text
		T3j	Review passage to check or clarify understanding	S3j	Review passage to check or clarify understanding
		T3k	Check accuracy of prediction or inference	S3k	Check accuracy of prediction or inference
		T4	Teaches vocabulary during or after reading text passage (Go to Vocab)	S4	Asks S(s) to justify their response with evidence
		T5	***If S response is incorrect or incomplete***: Assists S in using strategy(ies)	S5	Sets up independent practice for Ss to apply comprehension strategy(ies) (student work product or response required)
		HD	**Helps student(s) Decode word(s)**	**GH**	**Grammar or Handwriting is part of lesson**
		OR	**Student Oral Reading is part of lesson**	**SR**	**Student Silent Reading is part of lesson**
		TOR	**Teacher orally reads as students listen**		

Interval 1

VD Vocabulary Development

Grouping		T. Instruction: The Teacher. . .		S. Student Practice: The Teacher . . .	
W	Whole class	T1	Asks S(s) to give meaning of word	S1	Asks S(s) to apply understanding of word meaning
L	Lg grp	T2	Gives synonym	S2	Gives S(s) opportunity to practice word learning strategy(ies) (e.g., context, word structure, root meanings)
S	Sm grp	T3	Goes beyond synonym with definition and/or examples		
P	Pair	T4	Pinpoints word meaning with contrasting examples		
I	Individual	T5	Pinpoints word meaning by clarifying or extending Ss' partially correct response:	**List vocabulary words:**	
Support		T5a	Extension/clarification includes synonym		
V	Picture, Object	T5b	Extension/clarification includes definition and/or example		
P	Physical Demo.	T5c	Extension/clarification includes contrasting example	**GH**	**Grammar or Handwriting is part of lesson**
M	Word Map	**HD**	**Helps student(s) Decode word(s)**	**OR**	**Student Oral Reading is part of lesson**

PA Phonemic / Phonological Awareness (Sounds, NO PRINT)

Grouping		T.	Instruction: The Teacher. . .	S.	Student Practice: The Teacher . . .
W	Whole class		***Demonstrates or models***:		***Gives S(s) chance to practice:***
L	Lg grp	T1a	Oral work with syllables	S1a	Oral work with syllables
S	Sm grp	T1b	Oral blending or segmenting with onset-rimes	S1b	Oral blending or segmenting with onset-rimes
P	Pair	T1c	Oral blending or segmenting with phonemes	S1c	Oral blending or segmenting with phonemes
I	Individual	T1d	Phoneme Isolation	S1d	Phoneme isolation
		T1e	Phoneme categorization/identity (same/different sound in words)	S1e	Phoneme categorization/identity (same/different sound in words)
		T1f	Phoneme deletion, addition, substitution	S1f	Phoneme deletion, addition, substitution
Support		T2	Contrasts two phonemes to pinpoint target sound		
MK	Tchr Manip or Kin	T3	Pinpoints what S(s) did incorrectly with sound(s) and gives correct response with or without students		
SM	Manipulatives				
SK	Kinesthetic	T4	Introduces printed letters corresponding to sounds		

FB Fluency Building With Connected Text

Grouping		T.	Instruction & Student Practice: The Teacher. . .
W	Whole class	T1	Sets up or prompts S(s) to practice repeated or timed readings with a listener
L	Lg grp		
S	Sm grp		***Listens to Ss practice repeated oral readings:***
P	Pair	T2a	With text that was not modeled
I	Individual	T2b	With same text that was modeled by fluent reader
Support			***Listens to Ss practice timed oral readings:***
(C)	Connected text	T3a	With text that was not modeled
W	Written record	T3b	With same text that was modeled by fluent reader

Part B. Other Instruction									
	OR	Oral Reading	SP	Spelling	AS	Assessment	TR	Transition	MB: Managing Behavior: ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
	SR	Silent Reading	WE	Written Expression			AM	Academic Mgmt	
	TOR	Teacher Oral Reading	OL	Other Language Arts	NL	Non-literacy instruction	NI	Non-instructional	

Part C. Instructional Errors	**Part D.** Small group changes	T working with new small group: YES

Exhibit C.20: Global Appraisal of Teaching Strategies

Classroom ID#:_____ AbtClassID#: _____

	Item The teacher: Check one for each item:	All the time	More than half the time	Half the time	Less than half the time	Never	Not observed
1	**Seems to be organized and has all the materials necessary for instruction easily accessible** Locates and introduces new materials smoothly; comfortable using equipment and materials	❑	❑	❑	❑	❑	❑
2	**Presents and delivers the lesson effectively (makes eye contact, varies voice, engaged)** Makes eye contact with individual students; uses voice for emphasis, to hold student attention; engaged in delivering instruction, neither bored nor distracted; not a distraction from focus on content	❑	❑	❑	❑	❑	❑
3	**Distributes opportunities for students to participate equally and broadly** Solicits responses from a variety of students not just the most active volunteers or same subset	❑	❑	❑	❑	❑	❑
4	**Manages and coordinates student responses effectively** Has students listen to each other, take turns, cooperate and be courteous	❑	❑	❑	❑	❑	❑
5	**Pays attention, acknowledges and responds to student responses or input** Listens actively. Offers verbal/nonverbal acknowledgment, takes opportunities to build on student response	❑	❑	❑	❑	❑	❑
6	**Takes advantage of opportunities to provide corrective feedback when student makes error** Notices when student makes error and takes the opportunity to correct it	❑	❑	❑	❑	❑	❑
7	**Provides feedback in a positive manner** Feedback is not condescending or harsh; feedback may be firm but is neutral or encouraging	❑	❑	❑	❑	❑	❑
8	**Manages student behavior effectively in order to avoid disruptions and to provide productive learning opportunities** Provides enough structure to minimize disruptions; quickly and firmly intervenes to redirect off-task behavior	❑	❑	❑	❑	❑	❑
9	**Maintains a classroom environment that minimizes distractions and is appropriate for learning** Structures activities so groups do not interfere with each other; monitors noise level, insures that students have sufficient physical space (e.g., seating) and resources to attend to instruction	❑	❑	❑	❑	❑	❑
10	**Clearly communicates classroom procedure(s) when student breaks a rule or does not follow proper procedure** If needed, reminds student of relevant class rule or procedure; clear about what student is expected to do	❑	❑	❑	❑	❑	❑
	The teacher: Check one:	**True**	**Mostly true**	**Not sure**	**Mostly not true**	**Not true**	
11	**Is generally motivated and keeps students actively involved** Appears goal-directed, motivated to deliver instruction, interested in holding students' attention	❑	❑	❑	❑	❑	
	The teacher: Check one:	**All Ss**	**More than half of Ss**	**Half of Ss**	**Less than Half of Ss**	**None**	
12	**Checks on or monitors individual students' progress** Examines students' independent work one by one, and/or checks on individual students' skill level via one-on-one conference (can be brief), or formal/informal assessment	❑	❑	❑	❑	❑	

Exhibit C.21: Student Time-on-Task and Engagement with Print (STEP) Instrument

BARCODE HERE BARCODE HERE

Classroom ID		AbtClassID	
School ID		School Name	
Observer ID		Observer Name	
Date			
Grade		Room number	
Total # of students			
Time observation began		Time observation ended	

Anything unusual?

Classroom combined with another? Check here if YES ❑ and complete below:			
❑ Both classrooms part of the study		❑ Only one classroom is part of the study	
ClassID observed		ClassID observed	
ClassID not observed			

Sources consulted:

Foorman, B.R. & Schatschneider, C. (2003). Measurement of teaching practices during reading/language arts instruction and its relationship to student achievement. In S. Vaughn and K.L. Briggs (Eds.), Reading in the classroom: Systems for the Observation of Teaching and Learning (pp. 1-30). Baltimore, MD: Paul H. Brooks.

Kim, A., Briggs, K.L., & Vaughn, S. (2003). The classroom climate scale. In S. Vaughn and K.L. Briggs (Eds.), Reading in the classroom: Systems for the Observation of Teaching and Learning (pp. 83-109). Baltimore, MD: Paul H. Brooks.

Vaughn, S. (2005). Personal communication (May, 2005).

Shanahan, T. (2005). Personal communication (May, 2005)

Version 3.2 8/15/06 RFIS 2006-2007 Field Version

Instructions:

STEP observations are focused on the students not on the teacher.

Enter the classroom and start your countdown watch (set for 3:00 minute intervals). Begin walking casually around the room looking at what students are doing during the acclimation phase. Let students become accustomed to you walking around the room. Do not impede teacher or student movement but do not sit in one location during the acclimation period. Do not block IPRI observer's view of the teacher. Do not interact with students. If a student makes eye contact with you, look away – do not smile or otherwise acknowledge the student. If a student talks to you, smile and respond firmly, "I'm working, I can't talk with you right now."

After 6 minutes (beginning with the second buzz of your watch), begin **Sweep 1**:
If the more than ½ the students in the class are in transition, circle Y at the top of the Sweep 1 column and wait for Sweep 2;
If the whole class is listening the to the teacher read aloud a story (not following along in their own texts), circle Y at the top of Sweep 1 column and wait for Sweep 2. Otherwise:

a) Select a student to be Pupil 1. Record whether or not P1 is on or off task.
b) If P1 is On Task, record whether or not P1 is Reading Connected Text, Reading Isolated Text, or Writing (or none of these three); **If P1 is Off-Task do not fill in any of the three remaining columns for that student.**
c) Go on to the next student (P2). Repeat above steps for P2, P3, etc.
d) When you have finished with all students in the classroom, draw a solid line underneath the last pupil. Only students who are in the classroom should be counted in any Sweep.

Six minutes (two watch buzzes) after the start of Sweep 1, begin **Sweep 2**. Do not try to observe children in the same order that you observed during Sweep 1. That is, do not try to locate the same individual student who was P1 in Sweep 1. Instead, move around the room in a systematic fashion to observe each child once per sweep.

Repeat for **Sweep 3**.

After Sweep 3 is complete, go to the next classroom. Be sure to use a new STEP for each classroom observed.

On-task behavior is any behavior in which a child appears to be:

- Engaged in some independent, paired, or group learning activity (regardless of what teacher has asked student to do);
- Paying attention to the teacher (if teacher is delivering instruction) or attending to work that is in front of him/her;
- Talking with other students about an instructional activity in which both students are engaged (e.g. working productively on a group project);
- Participating in a whole class routine such as the pledge of allegiance

Off-task behaviors include:

- Not paying attention to the teacher when appropriate
- Looking around or gazing at an activity in which student is not engaged; "blank stares"
- Crying or head down with eyes covered on desk.
- Going to get new materials or put old materials away; or, wandering aimlessly
- Conflict with students or teacher
- Playing, teasing, roughhousing with other students, distracting other students
- Play behavior (playing with non-literacy related board games, blocks, dolls, action figures, legos, etc.)
- Snack/meal times or transitions (e.g., lining up to use the rest room)

Reading Connected Text:
Eyes are on a book, story, passage, or child is turning to next page of story. Even if student is momentarily looking at pictures that accompany a story, code as reading connected text unless book has no text. However, if student is flipping quickly through a book without pausing to read words on the page, do not code as Reading Connected Text, even though student may be coded as OnTask.

Reading Isolated Text:
Working with flashcards, looking at letters or words in isolation, completing a worksheet with isolated letters, words, sentences. Reading isolated sentences not part of a coherent, connected passage.

Writing:
Student has pen, pencil, crayon or other writing implement in hand and is writing or copying text, either isolated or connected. If student is drawing pictures do not code "Writing."

Version 3.2 8/15/06 RFIS 2006-2007 Field Version

Student Time-on-Task and Engagement with Print (STEP)

SWEEP 1			
	Circle one:		
Half or more of Class in Transition?	Y	N	**If Yes, SKIP this SWEEP**
Whole Class Listening to Story?	Y	N	**If Yes, SKIP this SWEEP**

		Engagement with Print		
On Task?	**On Task?**	**Reading Connected Text**	**Reading Isolated Text**	**Writing**
01	Y N	❑	❑	❑
02	Y N	❑	❑	❑
03	Y N	❑	❑	❑
04	Y N	❑	❑	❑
05	Y N	❑	❑	❑
06	Y N	❑	❑	❑
07	Y N	❑	❑	❑
08	Y N	❑	❑	❑
09	Y N	❑	❑	❑
10	Y N	❑	❑	❑
11	Y N	❑	❑	❑
12	Y N	❑	❑	❑
13	Y N	❑	❑	❑
14	Y N	❑	❑	❑
15	Y N	❑	❑	❑
16	Y N	❑	❑	❑
17	Y N	❑	❑	❑
18	Y N	❑	❑	❑
19	Y N	❑	❑	❑
20	Y N	❑	❑	❑
21	Y N	❑	❑	❑
22	Y N	❑	❑	❑
23	Y N	❑	❑	❑
24	Y N	❑	❑	❑
25	Y N	❑	❑	❑
26	Y N	❑	❑	❑
27	Y N	❑	❑	❑
28	Y N	❑	❑	❑
29	Y N	❑	❑	❑
30	Y N	❑	❑	❑

SWEEP 2			
	Circle one:		
Half or more of Class in Transition?	Y	N	If Yes, SKIP this SWEEP
Whole Class Listening to Story?	Y	N	If Yes, SKIP this SWEEP

		Engagement with Print		
Pupil	On Task?	Reading Connected Text	Reading Isolated Text	Writing
01	Y N	❑	❑	❑
02	Y N	❑	❑	❑
03	Y N	❑	❑	❑
04	Y N	❑	❑	❑
05	Y N	❑	❑	❑
06	Y N	❑	❑	❑
07	Y N	❑	❑	❑
08	Y N	❑	❑	❑
09	Y N	❑	❑	❑
10	Y N	❑	❑	❑
11	Y N	❑	❑	❑
12	Y N	❑	❑	❑
13	Y N	❑	❑	❑
14	Y N	❑	❑	❑
15	Y N	❑	❑	❑
16	Y N	❑	❑	❑
17	Y N	❑	❑	❑
18	Y N	❑	❑	❑
19	Y N	❑	❑	❑
20	Y N	❑	❑	❑
21	Y N	❑	❑	❑
22	Y N	❑	❑	❑
23	Y N	❑	❑	❑
24	Y N	❑	❑	❑
25	Y N	❑	❑	❑
26	Y N	❑	❑	❑
27	Y N	❑	❑	❑
28	Y N	❑	❑	❑
29	Y N	❑	❑	❑
30	Y N	❑	❑	❑

SWEEP 3			
	Circle one:		
Half or more of Class in Transition?	Y	N	If Yes, SKIP this SWEEP
Whole Class Listening to Story?	Y	N	If Yes, SKIP this SWEEP

		Engagement with Print		
Pupil	On Task?	Reading Connected Text	Reading Isolated Text	Writing
01	Y N	❑	❑	❑
02	Y N	❑	❑	❑
03	Y N	❑	❑	❑
04	Y N	❑	❑	❑
05	Y N	❑	❑	❑
06	Y N	❑	❑	❑
07	Y N	❑	❑	❑
08	Y N	❑	❑	❑
09	Y N	❑	❑	❑
10	Y N	❑	❑	❑
11	Y N	❑	❑	❑
12	Y N	❑	❑	❑
13	Y N	❑	❑	❑
14	Y N	❑	❑	❑
15	Y N	❑	❑	❑
16	Y N	❑	❑	❑
17	Y N	❑	❑	❑
18	Y N	❑	❑	❑
19	Y N	❑	❑	❑
20	Y N	❑	❑	❑
21	Y N	❑	❑	❑
22	Y N	❑	❑	❑
23	Y N	❑	❑	❑
24	Y N	❑	❑	❑
25	Y N	❑	❑	❑
26	Y N	❑	❑	❑
27	Y N	❑	❑	❑
28	Y N	❑	❑	❑
29	Y N	❑	❑	❑
30	Y N	❑	❑	❑

Appendix D: Confidence Intervals

Appendix D presents 95 percent confidence intervals for main impacts in relevant metrics as well as effect sizes. Confidence intervals for estimated impacts are reported for reading comprehension, instructional outcomes, and student engagement with print. Data are reported across these areas for pertinent study years.

Exhibit D.1: Confidence Intervals for Estimated Impacts on Reading Comprehension and Decoding Skills: Spring 2005, 2006, and 2007*

	Metric			Effect Size		
	Impact	Standard Error	Confidence Interval	Impact	Standard Error	Confidence Interval
Reading Comprehension (SAT 10)						
Scaled Scores						
Grade 1	4.74	2.72	-0.63 – 10.11	0.10	0.06	-0.01 – 0.21
Grade 2	1.69	2.29	-2.83 – 6.21	0.04	0.05	-0.07 – 0.15
Grade 3	0.30	2.12	-3.88 – 4.48	0.01	0.05	-0.10 – 0.11
Percent reading at or above grade level						
Grade 1	4.22	2.58	-0.87 – 9.31	N/A	N/A	N/A
Grade 2	1.60	2.38	-3.11 – 6.30	N/A	N/A	N/A
Grade 3	-0.08	2.40	-4.81 – 4.65	N/A	N/A	N/A
Decoding Skills (TOSWRF)						
Standard Score						
Grade 1	2.51*	1.11	0.32 – 4.69	0.17*	0.07	0.02 – 0.31

NOTES:

*The TOWSRF was administered in spring 2007 only.

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.

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

A 95% confidence interval was used.

EXHIBIT READS: The estimated impact of the Reading First program for grade 1 on reading comprehension scaled scores was 4.74 points with a standard error of 2.72 scaled score points. The 95% confidence interval for the estimated impact ranged from -0.63 points to 10.11 points.

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 TOSWRF administration in the spring of 2007.

Exhibit D.2: Confidence Intervals for Estimated Impacts on Instructional Outcomes: Spring 2005, Fall 2005, Spring 2006, Fall 2006 and Spring 2007

	Metric			Effect Size		
	Impact	Standard Error	Confidence Interval	Impact	Standard Error	Confidence Interval
Panel 1	***(minutes)***					
Number of minutes of instruction in five dimensions combined						
Grade 1	6.92*	2.44	2.13 - 11.71	0.33*	0.12	0.10 - 0.56
Grade 2	9.79*	2.46	4.97 - 14.60	0.46*	0.12	0.23 - 0.69
Panel 2	***(percent)***					
Percentage of intervals in five dimensions with						
Highly Explicit Instruction						
Grade 1	3.29*	1.39	0.57 - 6.01	0.18*	0.08	0.03 - 0.33
Grade 2	3.00*	1.45	0.16 - 5.84	0.16*	0.08	0.01 - 0.30
High Quality Student Practice						
Grade 1	0.82	1.25	-1.63 - 3.27	0.05	0.07	-0.10 - 0.19
Grade 2	2.94*	1.25	0.50 - 5.39	0.16*	0.07	0.03 - 0.30

NOTES:

The complete Reading First Impact Study (RFIS) 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.

Impact estimates are statistically adjusted 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 *.

A 95% confidence interval was used.

EXHIBIT READS: The estimated impact of the Reading First program for grade 1 on the amount of time spent in instruction in the five dimensions was 6.92 minutes with a standard error of 2.44 minutes. The estimated impact was statistically significant. The 95% confidence interval for the estimated impact ranged from 2.13 minutes to 11.71 minutes.

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

Exhibit D.3: Confidence Intervals for Estimated Impacts on Time Spent in Instruction in the Five Dimensions: Spring 2005, Fall 2005, Spring 2006, Fall 2006, and Spring 2007

	Metric			Effect Size		
	Impact	Standard Error	Confidence Interval	Impact	Standard Error	Confidence Interval
Number of minutes of instruction in:						
Phonemic Awareness						
Grade 1	0.61*	0.28	0.06 - 1.16	0.23*	0.10	0.02 - 0.43
Grade 2	0.12	0.12	-0.12 - 0.36	0.10	0.10	-0.09 - 0.29
Phonics						
Grade 1	2.86*	1.44	0.04 - 5.68	0.21*	0.11	0.00 - 0.42
Grade 2	3.27*	1.17	0.97 - 5.57	0.31*	0.11	0.09 - 0.53
Vocabulary						
Grade 1	0.57	0.66	-0.72 - 1.86	0.09	0.10	-0.11 - 0.28
Grade 2	1.73*	0.82	0.13 - 3.34	0.20*	0.09	0.01 - 0.38
Fluency						
Grade 1	1.24*	0.61	0.05 - 2.43	0.20*	0.10	0.01 - 0.40
Grade 2	0.58	0.56	-0.51 - 1.67	0.11	0.10	-0.09 - 0.31
Comprehension						
Grade 1	1.78	1.53	-1.22 - 4.77	0.12	0.11	-0.08 - 0.33
Grade 2	4.01*	1.69	0.70 - 7.32	0.24*	0.10	0.04 - 0.44

NOTES:

The complete Reading First Impact Study (RFIS) 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.

Impact estimates are statistically adjusted 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 *.

A 95% confidence interval was used.

EXHIBIT READS: The estimated impact of the Reading First program for grade 1 on the amount of time spent in instruction in phonemic awareness was 0.61minutes with a standard error of 0.28 minutes. The estimated impact was statistically significant at the $p \leq .05$ level. The 95% confidence interval for the estimated impact ranged from 0.06 minutes to 1.16 minutes.

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

Exhibit D.4: Confidence Intervals for Estimated Impacts on Student Engagement with Print: Fall 2005, Spring 2006, Fall 2006, and Spring 2007

	Metric			Effect Size		
	Impact (percent)	Standard Error	Confidence Interval	Impact	Standard Error	Confidence Interval
Percentage of student engagement with print						
Grade 1	5.33	2.92	-0.40 – 11.05	0.18	0.10	-0.01 – 0.38
Grade 2	-4.75	2.91	-10.45 – 0.96	-0.17	0.10	-0.37 – 0.03

NOTES:

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

Impact estimates are statistically adjusted 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 *.

A 95% confidence interval was used.

EXHIBIT READS: The estimated impact of the Reading First program for grade 1 on the percentage of students engaged with print was 5.33 percentage points with a standard error of 2.92 percentage points. The 95% confidence interval for the estimated impact ranged from -0.40 percentage points to 11.05 percentage points.

SOURCE: RFIS Student Time-on-Task and Engagement with Print, fall 2005, spring 2006, fall 2006 and spring 2007.

Appendix E: Analyses of Impacts and Trends Over Time

Part 1: Additional Exhibits of Separate Impact Estimates for Each Follow-up Year and Pooled

Exhibits E.1 through E.3 present additional exhibits of separate impact estimates for each grade level for each follow-up year as well as pooled across applicable years, for the number of minutes of instruction in each of the five dimensions of reading and for the percentage of students engaged with print.

Part 2: Student Achievement Trends Over Time

This part of the appendix presents a brief discussion of student achievement results over time.

Part 3: Reading Achievement on State Tests

The final part of the appendix reports on the estimated impact of Reading First on statewide reading tests in the RFIS core sample, and discusses how these results compare to results on the SAT 10 reading comprehension subtest for the RFIS core sample.

Exhibit E.1: Estimated Impacts on the Number of Minutes of Instruction in Each of Five Dimensions of Reading in First Grade: 2005, 2006, and 2007, and Pooled

Number of minutes of instruction in:	Actual Mean With Reading First	Estimated Mean Without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (p-value)
Grade 1					
Phonemic Awareness					
Spring 2005	1.64	0.76	0.88*	0.33*	(0.004)
School year 2006	2.32	1.69	0.62	0.23	(0.102)
School year 2007	2.71	2.19	0.52	0.19	(0.244)
Pooled 3 years (Sp05, Sy06, Sy07)	2.32	1.71	0.61*	0.23*	(0.030)
Phonics					
Spring 2005	21.02	18.05	2.97	0.22	(0.141)
School year 2006	21.56	16.99	4.57*	0.34*	(0.012)
School year 2007	21.27	18.90	2.37	0.18	(0.223)
Pooled 3 years (Sp05, Sy06, Sy07)	21.32	18.45	2.86*	0.21*	(0.048)
Vocabulary					
Spring 2005	7.03	5.48	1.55	0.23	(0.072)
School year 2006	8.22	8.02	0.20	0.03	(0.827)
School year 2007	8.10	7.86	0.24	0.04	(0.794)
Pooled 3 years (Sp05, Sy06, Sy07)	7.92	7.35	0.57	0.09	(0.386)
Fluency					
Spring 2005	5.26	3.72	1.53	0.25	(0.180)
School year 2006	4.13	3.22	0.91	0.15	(0.165)
School year 2007	4.84	3.23	1.61*	0.26*	(0.041)
Pooled 3 years (Sp05, Sy06, Sy07)	4.67	3.43	1.24*	0.20*	(0.043)
Comprehension					
Spring 2005	24.29	22.19	2.10	0.15	(0.349)
School year 2006	23.27	20.85	2.42	0.17	(0.244)
School year 2007	22.01	20.80	1.21	0.08	(0.504)
Pooled 3 years (Sp05, Sy06, Sy07)	23.01	21.23	1.78	0.12	(0.247)

NOTES:

The complete Reading First Impact Study (RFIS) 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 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 per daily reading block in instruction in phonemic awareness for first grade classrooms in spring 2005 with Reading First was 1.64 minutes. The estimated mean amount of time without Reading First was 0.76 minutes. The impact of Reading First on the amount of time spent in instruction in phonemic awareness was 0.88 minutes (or 0.33 standard deviations), which was statistically significant (p=.004).

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

Exhibit E.2: Estimated Impacts On the Number of Minutes in Instruction in Each of Five Dimensions of Reading in Second Grade: 2005, 2006, and 2007, and Pooled

Number of minutes of instruction in:	Actual Mean With Reading First	Actual Mean Without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (p-value)
Grade 2					
Phonemic Awareness					
Spring 2005	0.43	0.30	0.13	0.10	(0.386)
School year 2006	0.42	0.28	0.14	0.11	(0.271)
School year 2007	0.59	0.49	0.11	0.09	(0.621)
Pooled 3 years (Sp05, Sy06, Sy07)	0.49	0.37	0.12	0.10	(0.319)
Phonics					
Spring 2005	13.99	10.74	3.25*	0.31*	(0.044)
School year 2006	14.02	10.09	3.93*	0.37*	(0.013)
School year 2007	13.87	10.93	2.94	0.28	(0.067)
Pooled 3 years (Sp05, Sy06, Sy07)	13.92	10.65	3.27*	0.31*	(0.006)
Vocabulary					
Spring 2005	10.46	8.69	1.76	0.20	(0.117)
School year 2006	12.27	9.93	2.34	0.27	(0.058)
School year 2007	12.05	10.86	1.19	0.14	(0.275)
Pooled 3 years (Sp05, Sy06, Sy07)	11.79	10.06	1.73*	0.20*	(0.036)
Fluency					
Spring 2005	5.12	2.80	2.32*	0.42*	(0.013)
School year 2006	3.73	4.22	-0.49	-0.09	(0.475)
School year 2007	3.98	3.54	0.44	0.08	(0.523)
Pooled 3 years (Sp05, Sy06, Sy07)	4.14	3.56	0.58	0.11	(0.297)
Comprehension					
Spring 2005	28.33	22.84	5.49*	0.33*	(0.024)
School year 2006	29.69	24.85	4.85*	0.29*	(0.033)
School year 2007	28.08	26.24	1.84	0.11	(0.372)
Pooled 3 years (Sp05, Sy06, Sy07)	28.74	24.73	4.01*	0.24*	(0.019)

NOTES:

The complete Reading First Impact Study (RFIS) 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 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 per daily reading block in instruction in phonemic awareness for second grade classrooms in spring 2005 with Reading First was 0.43 minutes. The estimated mean amount of time without Reading First was 0.30 minutes. The impact of Reading First on the amount of time spent in instruction in phonemic awareness was 0.13 minutes (or 0.10 standard deviations), which was not statistically significant (p=.386).

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

Exhibit E.3: Estimated Impacts on the Percentage of Students Engaged with Print: 2006 and 2007, and Pooled

Construct	Actual Mean with Reading First	Estimated Mean without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (p-value)
Grade 1					
Percentage of students engaged with print					
School Year 2006	46.92	42.28	4.64	0.16	(0.216)
School Year 2007	48.68	43.24	5.44	0.18	(0.170)
Pooled (SY 2006, SY 2007)	47.84	42.52	5.33	0.18	(0.070)
Grade 2					
Percentage of students engaged with print					
School Year 2006	49.83	58.25	-8.42*	-0.29*	(0.029)
School Year 2007	51.13	52.14	-1.01	-0.04	(0.808)
Pooled (SY 2006, SY 2007)	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 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 pooled across the fall 2005 and spring 2006 STEP 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: For the 2006 school year, the actual average percentage of students engaged with print in first grade classrooms with Reading First was 46.92 percent. The estimated average percentage without Reading First was 42.28 percent. The impact of Reading First on the average percentage of student engagement with print was 4.64 percentage points (or 0.16 standard deviations), which was not statistically significant (p=.216).

SOURCE: RFIS Student Time-on-Task and Engagement with Print, fall 2005, spring 2006, fall 2006, and spring 2007.

Part 2: Student Achievement Trends Over Time

Exhibits E.4 and E.5 present student achievement trends over time for schools in the RFIS study sample. Data on mean SAT 10 scores are presented at three time points—spring 2005, spring 2006 and spring 2007—separately for Reading First and non-Reading First schools across the 248 schools in the 18 sites in the RFIS study sample.

For each year and grade, three mean scaled score values were calculated. **The Actual Mean with Reading First** value is simply that; it is the actual unadjusted mean for the Reading First schools in the study sample. The **Estimated Mean without Reading First** value represents the best estimate of what would have happened in Reading First schools absent Reading First funding. The Actual Mean for Non-Reading First schools value is the unadjusted mean for the non-Reading First schools in the study sample.[1]

The Estimated Mean without Reading First is the counterfactual and in the absence of Reading First represents the best estimate of what would have happened in the treatment schools—if they had not been selected as Reading First schools. The Actual Mean with Reading First and the Estimated Mean without Reading First values are identical to the values shown in the impact tables in Chapter 3 and Appendix E, Part 1. Calculation of the counterfactual accounts for each school's rating and prior achievement, both of which were generally higher in non-RF schools, as RF grants were awarded to schools with greatest need within each site. The Actual Mean for Non-Reading First schools value does not take into account either (1) the criteria (or rating) used to determine their RF status or (2) any pre-RF differences in student achievement.

In Exhibit E.4, the first row shows mean scaled scores on the SAT 10 for grade 1 in spring 2005. From left to right, the table displays the actual (or unadjusted) mean for RF schools (541.2), then the estimated mean in the absence of RF (538.9), and in the third column, the actual (or unadjusted) mean for non-RF schools, (542.5). Note that this exhibit does not display the estimated impact of Reading First, which is the presented in the main body of the report (i.e., 2.2 scaled score points, representing the difference between the values in columns 1 and 2).

Exhibit E.4 also includes the corresponding grade equivalent and national percentile for each scaled score mean value.[2] The remaining rows in the table show values for grade 2 and grade 3 (spring 2005), grades 1-3 (spring 2006), and grades 1-3 (spring 2007).

The scaled score means displayed in Exhibit E.4 are graphed in Exhibit E.5. Because the SAT 10 scaled score range is continuous across grades, all values can be shown on a single set of axes. For each grade, the vertical bars represent the average scaled score for RF schools (unadjusted), schools in the absence of RF (estimated), and non-RF schools (unadjusted); the first bar in each set of three represents the mean for spring 2005, the second bar in each set of three represents the mean for spring 2006, and the third bar in

[1] All means are weighted by the number of Reading First schools in each site, which is the same weighting scheme used for the impact estimates presented in the main body of this report.

[2] Calculations of mean values were done for scaled scores only. Average scaled scores for Reading First schools and non-Reading First schools were converted to grade equivalents and national percentiles. It is not appropriate to perform arithmetic calculations with grade equivalents or percentiles.

Exhibit E.4: Reading Comprehension Means: Spring 2005, Spring 2006, and Spring 2007

	Actual Mean with Reading First	Estimated Mean without Reading First	Actual Mean for Non-Reading First Schools
All Sites			
Spring 2005			
Grade 1			
Scaled Score	541.2	538.9	542.5
Corresponding Grade Equivalent	*1.7*	*1.7*	*1.7*
Corresponding Percentile	*43*	*41*	*44*
Grade 2			
Scaled Score	583.5	582.4	586.7
Corresponding Grade Equivalent	*2.5*	*2.4*	*2.5*
Corresponding Percentile	*38*	*38*	*41*
Grade 3			
Scaled Score	607.4	609.9	610.7
Corresponding Grade Equivalent	*3.2*	*3.3*	*3.4*
Corresponding Percentile	*38*	*39*	*40*
Spring 2006			
Grade 1			
Scaled Score	545.7	540.4	545.8
Corresponding Grade Equivalent	*1.8*	*1.7*	*1.8*
Corresponding Percentile	*46*	*42*	*46*
Grade 2			
Scaled Score	585.3	583.7	586.0
Corresponding Grade Equivalent	*2.5*	*2.5*	*2.5*
Corresponding Percentile	*40*	*38*	*40*
Grade 3			
Scaled Score	609.5	610.0	613.9
Corresponding Grade Equivalent	*3.3*	*3.3*	*3.5*
Corresponding Percentile	*39*	*39*	*43*
Spring 2007			
Grade 1			
Scaled Score	545.3	537.8	545.8
Corresponding Grade Equivalent	*1.8*	*1.7*	*1.8*
Corresponding Percentile	*46*	*40*	*46*
Grade 2			
Scaled Score	584.8	582.3	585.9
Corresponding Grade Equivalent	*2.5*	*2.4*	*2.5*
Corresponding Percentile	*39*	*38*	*40*
Grade 3			
Scaled Score	610.6	605.1	611.7
Corresponding Grade Equivalent	*3.4*	*3.1*	*3.4*
Corresponding Percentile	*40*	*36*	*41*

NOTES:

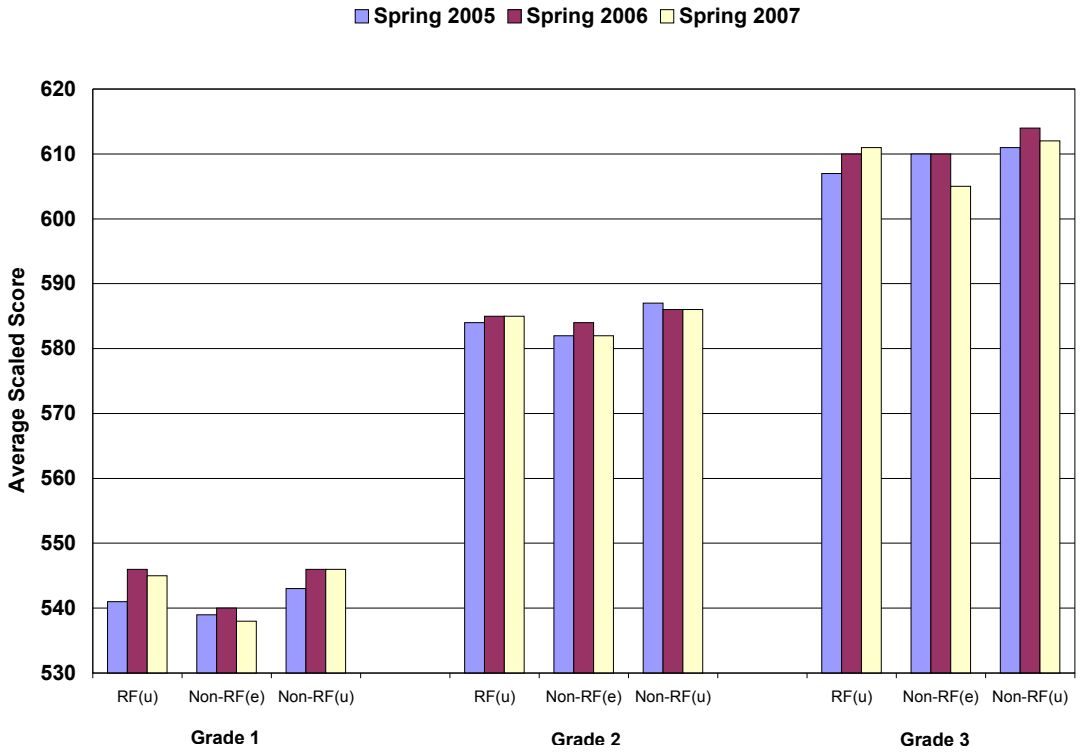
The complete Reading First Impact Study (RFIS) sample includes 248 schools from 18 sites (17 school districts and 1 state) located in 13 states. 125 schools are Reading First schools and 123 are non-Reading First schools. For grade 2, one non-RF school could not be included in the analysis because test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

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. The actual mean for non-Reading First schools is the observed average for non-Reading First schools in the study sample.

EXHIBIT READS: On average, for first-graders in the spring of 2005, the observed mean reading comprehension score with Reading First was 541.2 scaled score points. The estimated mean without Reading First was 538.9 scaled score points. The observed mean in non-Reading First schools was 542.5 scaled score points.

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 E.5: Reading Comprehension Means: Spring 2005, Spring 2006, and Spring 2007



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 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.

For each grade, the vertical bars represent the average scaled score for RF schools (unadjusted), schools in the absence of RF (estimated), and non-RF schools (unadjusted).

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).

each set of three represents the mean for spring 2007. Mean values for grade one are the first group of vertical bars, mean values for grade two are the middle group of bars, and mean values for grade three are the last group of bars.

Part 3: Reading Achievement on State Tests

The Reading First Impact Study (RFIS) also examined reading achievement patterns for the RFIS sample schools on state-mandated reading assessments in grade 3 only (the first grade level for which state assessments are required). The study team used extant data from the 2006 administration of state-specific assessments to address the following two questions:

1) What is the impact of Reading First on statewide reading tests in the RFIS core sample?

2) How do results for the RFIS core sample on the SAT 10 reading comprehension subtest compare with results for the RFIS core sample on statewide tests?

Data

The RFIS team used school-level reading performance test score data for elementary grades in all study schools. Most states use criterion-referenced tests and report performance as the percent of students (in a given grade at a given school) who scored at state-defined levels of proficiency (e.g., percent proficient, percent advanced). The extant data available (at the time of these analyses) are current through school year 2005-06, and include data from as early as 1997 for some states. These extant data were merged with the Common Core of Data (CCD), which includes salient demographic and other school-level information.

Analysis

As was the case for the main impact findings reported in Chapter 3, the analyses described below used the regression discontinuity design (RDD) to model achievement outcomes. To be eligible for these analyses, a state's data had to include: (1) at least one year of pre-Reading First reading/English Language Arts (ELA) assessment data, so that a pretest measure could be included in the analyses, (2) 2006 grade 3 reading/ELA assessment scores, and (3) school-level scores based on a percent proficient (or percent advanced, etc.) metric. Two of the 18 sites were excluded from these analyses because their states' data did not meet these three criteria. A total of 210 schools in 16 sites in 11 states were included in these analyses.

The outcome measure for these analyses is percent performing at grade level on the SAT 10. For the statewide reading tests, the outcome measure was percent proficient (on statewide tests); in some states, however, the percent proficient metric indicated ceiling effects, and therefore an alternative benchmark was used (e.g., percent advanced) to allow more meaningful comparisons to the SAT 10 metric.

Results

Exhibit E.6 shows the RDD impacts on state test scores and percent performing at grade level on the SAT 10 for the RFIS core sample. In both cases, the impact estimates are negative (-2.11 percentage points for the state test, -.70 percentage points for the SAT 10) and are not statistically significant.[3]

[3] In 2006, the 3rd grade impact estimate on the SAT 10 percent at grade level metric for the *entire* RFIS core sample (244 schools and 18 sites) was -0.9, p=0.80. (See Exhibit 3.3 for year by year SAT 10 results).

Exhibit E.6: Estimated Impacts of Reading First on Grade 3 State Reading/ELA Tests and SAT 10 Reading Comprehension Subtest: 2006

	State Reading/ELA Test *Percent Proficient*	SAT 10 Reading Comprehension *Percent at Grade Level*
Estimate	-2.11	-0.70
Standard Error	2.51	3.51
P-value	(0.401)	(0.843)

NOTES:

The sample includes 210 schools from 16 sites. 104 schools are Reading First and 106 are non-Reading First schools.

Site-by-site impact estimates are weighted by the number of RF schools in each site's core sample to yield the overall impact estimates.

Models include site indicators, interactions of the site indicators with the rating and pre-test variables, and percentage of male students.

EXHIBIT READS: The estimated average impact of Reading First on the percentage of grade 3 students performing at the proficient level on 2006 state reading assessments was -2.11 percentage points with a standard error of 2.51, which was not statistically significant (p=.401).

SOURCES: State reading/ELA test data. RFIS SAT 10 administration in the spring of 2006, 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).

Appendix F: Analysis of Student Exposure to Reading First

Variation in Impacts on Reading Comprehension Based on Student Exposure

Reading First is intended to provide schools with a complete instructional program from kindergarten through grade three. However, some students have not had the opportunity to be exposed to the full range of Reading First instructional practices and support services due to student mobility. As a result, the impact of the program for any one grade level may be muted by the mix of students with varying levels of exposure to the program.

To explore the potential effect of a full three years of exposure to Reading First, the analysis focuses on students in the study sample who were likely to have been enrolled in a Reading First school in grades one, two and three, specifically, those 3rd grade students tested by the study in both 2007 and 2005 who were enrolled in Reading First schools in both 2007 and in 2005. Those students' reading achievement scores are then compared with their 3rd grade counterparts in non-Reading First schools.[1] Impact estimates from this sample of students may provide an indication of the effect of Reading First for students who received three years of exposure to Reading First.

These results should be interpreted with caution, however, because the impact estimates may be biased if Reading First caused a difference either in student mobility rates or in the types of students who move from or stay at the same school. For example, if Reading First induced higher achieving students to remain in the same school while similar students were more likely to leave a non-Reading First school, positive impact estimates may be an artifact of differential mobility patterns rather than real improvements in reading comprehension for the "remaining" students. Consequently, before presenting the test score impact findings for students with three years of exposure to Reading First, these analytic issues must be discussed.

Analysis of the proportion of 3rd grade students from the 2007 data collection who were in the study sample in 2005 provides evidence that Reading First did not appear to affect overall mobility rates. Exhibit F.1 shows the observed percentage of 3rd grade students from the 2007 data collection effort who were in a school with the same treatment status in 2005,[2] by Reading First status and by site.

---

[1] In the spring of 2005, the study tested in all classrooms in study schools; in subsequent waves of testing, the study tested only a subsample of classrooms in those schools with more than an average of three classrooms per grade. Because not all classrooms per grade level were tested in both 2006 and 2007, we cannot use 2006 data to identify how many 3rd graders in our 2007 sample were in study schools in 2006.

[2] The percentages presented in Exhibits F.1 and F.2 do necessarily reflect actual mobility patterns. It is possible that some of these students were enrolled in the same school in 2005 and 2007 but in a different school in 2006. Also, the sample of students included in this analysis are those who were tested as 3rd graders in a given school in 2007 and were also enrolled in a study school with the same program status (Reading First or non-Reading First) in 2005. Because not all 3rd grade students were tested in all of the study schools in 2007, this sample may not encompass all students who remained in the same type of school for three years.

Exhibit F.1: Percentage of Third Graders in Same Treatment Status for Three Years by Site and Treatment Status

	Percentage of 3rd Graders in Reading First Schools in 2007 and in 2005	Percentage of 3rd Graders in Non-Reading First Schools in 2007 and in 2005	Overall Percentage of 3rd Graders in Same Treatment Status in 2007 and 2005
Site 1	64	41	52
Site 2	64	67	66
Site 3	53	54	53
Site 4	73	71	72
Site 5	42	41	42
Site 6	51	75	63
Site 7	59	63	61
Site 8	72	71	72
Site 9	55	61	59
Site 10	67	68	67
Site 11	70	54	61
Site 12	65	59	63
Site 13	51	54	52
Site 14	61	63	62
Site 15	63	74	69
Site 16	56	56	56
Site 17	66	65	66
Site 18	70	65	67
Overall	60	60	60

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 3 in 2007, one RF school could not be included in the analysis because test score data were not available.

This exhibit shows unweighted means for each site, unlike exhibits that present impact estimates in which means for each site are weighted by the number of Reading First schools in the site.

The percentages presented above do necessarily reflect actual mobility patterns. It is possible that some of these students were enrolled in the same school in 2005 and 2007 but in a different school in 2006. Also, the sample of students included in this analysis are those who were tested as 3rd graders in a given school in 2007 and were also enrolled in a study school with the same program status (Reading First or non-Reading First) in 2005. Because not all 3rd grade students were tested in all of the study schools in 2007, this sample may not encompass all students who remained in the same type of school for three years.

EXHIBIT READS: In Site 1, 64 percent of third grade students in the study sample in Reading First schools in 2007 were also in Reading First schools in 2005. Forty-one percent of third graders in the study sample in non-Reading First schools in 2007 were in non-Reading First schools in 2005. Overall, in Site 1, 52 percent of third grade students in the study sample in 2007 had the same treatment status in 2005.

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).

The average mobility rate across all sites is 60 percent, and it ranges from 41 to 75 percent. Exhibit F.2 shows the overall estimated differences in the percentage of students who were enrolled in schools of the same treatment status in 2005 and 2007. Panel 1 shows a regression adjusted difference, including adjustments for the regression discontinuity design. Panel 2 is not regression adjusted.

Exhibit F.2: Estimated Regression Adjusted and Unadjusted Impacts of Reading First on the Percent of Students With Three Years of Exposure to the Same Treatment Status, Spring 2005-Spring 2007

	Actual Mean with Reading First	Estimated Mean without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (p-value)
Panel 1					
Percent of Students With Three Years of the Same Treatment Status (%)					
Regression Adjusted	60.8	56.4	4.4	N/A	(0.278)
Panel 2					
	Actual Mean with Reading First	**Actual Mean for Non-Reading First Schools**			
Percent of Students With Three Years of the Same Treatment Status (%)					
Unadjusted	60.8	61.5	-0.7	N/A	(0.696)

NOTES:

The Three-Year Exposure sample includes 243 schools from 18 sites (17 school districts and 1 state) located in 13 states. 123 schools are Reading First schools and 120 are non-Reading First schools.

In panel 1, the value in the "Actual Mean with Reading First" column is the actual, unadjusted value for Reading First schools; the value in the "Estimated Mean without Reading First" column represents the best estimates of what would have happened in RF schools absent RF funding and is calculated by subtracting the impact estimate from the RF schools' actual mean values.

In panel 2, the value in the "Actual Mean with Reading First" column is the actual, unadjusted value for Reading First schools; the "Actual Mean for Non-Reading First Schools" is the actual, unadjusted value for non-Reading First schools in the study sample and is calculated by subtracting the impact estimate 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 average percent of third-grade students in Reading First schools that were in a Reading First school two years earlier is 60.8 percent. The estimated average percent of third-grade students in non-Reading First Schools in non-Reading First schools two years earlier is 56.4 percent. The impact of Reading First on the percent of third-grade students with three years of exposure to the same treatment status is 4.4 percentage points, which is not statistically significant (p=.278).

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).

- There was a 4.4 percentage point difference in the number of students who stayed in a Reading First versus a non-Reading First school for three years. The difference was not statistically significant, indicating that Reading First did not have a systematic impact on whether or not a student stayed in a school with the same treatment status over time. Also, although Exhibit F.1 indicates that there was variation in the observed difference between Reading First and non-Reading First schools on this percentage (ranging from +23 percentage points in site 1 to -24 percentage points in site 6), an F-test indicates that the variation is not statistically significant.

It is also important to assess the extent to which the Reading First program may have influenced the compositional mix of students who were enrolled in RF or non-RF schools, among schools whose RF or non-RF status did not change between 2005 and 2007. Because the study does not include pre-Reading First characteristics for students in the study sample, this question cannot be examined directly. As a result, the findings presented in this section should be interpreted with caution. Also, students who remain in schools with the same treatment status for three years likely differ along a number of important dimensions from students who do not, so the results of this analysis may have limited external validity.

Exhibit F.3 (which is identical to Exhibit 3.4) shows the estimated impacts of Reading First on the subsample of students who remained in either RF or non-RF study schools in both 2005 and 2007.

- The results indicate that the impact of the program on grade three students with three years of exposure was 4.3 scaled score points and was not statistically significant.

Exhibit F.3: Estimated Impacts of Reading First on the Reading Comprehension of Students With Three Years of Exposure: Spring 2005-Spring 2007

	Actual Mean with Reading First	Estimated Mean without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (P-value)
Students With Three Years of Exposure					
Grade 3, Spring 2007					
Reading Comprehension					
Scaled Score	613.6	609.3	4.3	0.11	(0.223)
Corresponding Grade Equivalent	*3.5*	*3.3*			
Corresponding Percentile	*43*	*39*			

NOTES:

The Three-Year Exposure sample includes 243 schools from 18 sites (17 school districts and 1 state) located in 13 states. 123 schools are Reading First schools and 120 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 pooled across the spring 2005 and 2006 SAT 10 test scores (by grade).

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 third-graders with three years of exposure to Reading First was 613.6 scaled score points. The estimated mean without Reading First was 609.3 scaled score points. The impact of Reading First was 4.3 scaled score points (or 0.11 standard deviations), which was not statistically significant (p=.223).

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).