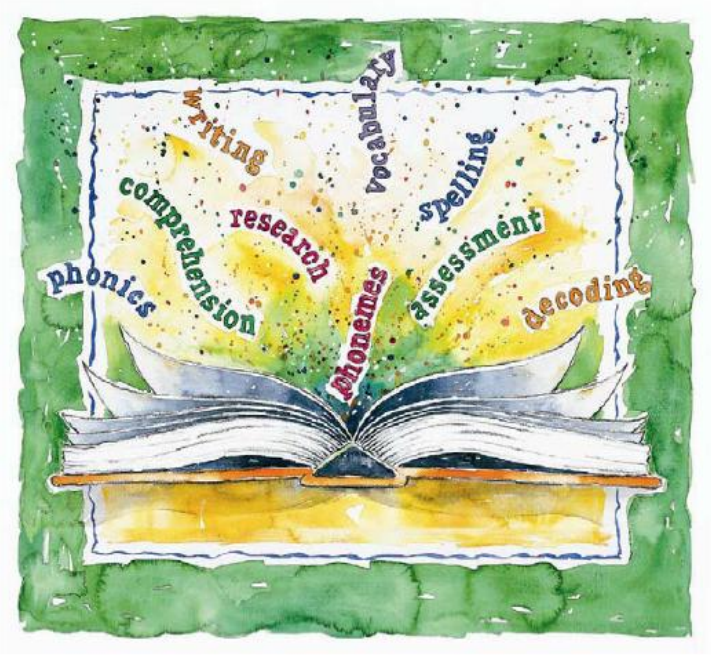
The California Reading First Year 4 Evaluation Report

December 2006



Prepared by:

Diane Haager, Ph.D.
Principal Investigator

and

Educational Data Systems
Renuka Dhar
Mark Moulton, Ph.D.
Susan McMillan, Ph.D.

Educational Data Systems

Executive Summary

Reading First is a federal initiative aimed at improving reading instruction in America. Authorized in 2001 as part of the No Child Left Behind Act, Reading First promotes the use of scientifically based reading practices in grades K-3. The initiative provides a significant amount of federal funding for improving reading instruction for large proportions of students experiencing academic difficulty and socio-economic disadvantage.

The Reading First program in California began in the 2002-2003 school year[1], 4 years ago. Its components include:

- Use of a state-approved reading program
- Access to training programs authorized by state legislation and based on research-based reading instruction: Assembly Bill (AB) 466 teacher and coach professional development and AB 75 principal professional development
- Access to assessment tools that test students' skills every six to eight weeks
- Hiring of reading coaches, expert teachers who support program implementation

Anecdotal evidence indicates that many non-Reading First schools have voluntarily been adopting some or all of these components over the same 4-year period, giving this evaluation study a relevance that extends beyond the Reading First population.

This report evaluates California's progress in implementation and achievement during the first four years of funding and provides information regarding program efficacy.

- Chapter 1 provides on overview of Reading First and its history, data sources, and the research design.
- Chapter 2 discusses demographic characteristics of the three cohorts of Reading First schools and how they compare to non-Reading First schools. It defines the "control group" we use as part of the design for judging program effectiveness.
- Chapter 3 provides the results of the implementation surveys administered to every Reading First teacher, coach, and principal in California. A Reading First Implementation Index statistic (RFII), computed for each school, is derived from these surveys.
- Chapter 4 provides the achievement results for all Reading First schools (high implementing and low implementing), as well as for a statistical control group and for non-Reading First schools.

[1] In this report, we generally refer to the "year" as that of the spring of the school year. For example, for the 2003-2004 school year would be referred to as "2004."

- Chapter 5 provides a qualitative analysis of interviews conducted with Local Education Agency (LEA) administrators, focus groups with school principals, and teacher, coach, and principal open-ended responses to the implementation survey.

In the Executive Summary we report six key findings and four policy recommendations regarding the Reading First program. These are supported by summary tables and trend-line charts.

Key Findings

As presented below, we find:

1. California's Reading First program is having a positive impact on overall student reading achievement in grades 2 and 3.
2. The Reading First program is having a positive impact on all implementing schools, including those with high percentages of English Learners or Socio-Economically Disadvantaged students, similar to the impact on all Reading First schools.
3. Based on analysis of the Reading First implementation surveys, the elements of Reading First program are being adequately implemented in Reading First schools, although there is room for improvement. The level of implementation of the Reading First program for each participating school is measured using the "Reading First Implementation Index" (RFII). [2]
4. The level of implementation of the Reading First program, per year and cumulatively across years, is a statistically significant predictor of school growth in reading achievement after controlling for other characteristics.
5. It takes one to two years for schools to begin fully implementing the Reading First program and about the same length of time to show an effect on achievement.
6. The process of achieving significant change in reading instruction is complex, as demonstrated by portraits of Reading First implementation derived from interviews, focus groups and open-ended survey questions.

Key Finding #1: California's Reading First program is having a positive impact on overall student achievement.

In this study, student achievement is measured at the school level using the California Standards Test (CST); the California Achievement Test, 6th Edition (CAT/6); the curriculum-embedded End-of-Year test

[2] The RFII, described in Chapter 3, is computed at the school level from surveys of all Reading First teachers, coaches, and principals.

(EOY); and the Reading First Achievement Index (RFAI).[3] To support this finding, we present trends in achievement data from 2002 to 2006 using groups of schools that have been in the program from two to four years. Program effect is found by correlating implementation of the program with school achievement gains on multiple assessment metrics. For the Executive Summary, we profile Grade 2 CST results, though results for other grade levels and measures are discussed extensively in the report. In Grade 2, students take the CST for the first time and this provides the earliest opportunity to examine the impact of the Reading First program on students' achievement relative to state standards. In Grade 2, we find:

1. Reading First schools have shown significant growth since their entry into the program.
    - "Proficient" and above. Schools in the Reading First program for four years show an average gain of 15.7 percentage points over four years in the percent of students scoring "Proficient" and "Advanced" on the Grade 2 CSTs. Thus, a Reading First school with 15 percent of its students scoring "Proficient" or "Advanced" in 2003 is likely to reach 30 percent in the 2006-2007 school year.
    - "Below Basic" and below. The number of students scoring "Below Basic" and "Far Below Basic" declined 15.7 percentage points over the same period as they moved into higher performance levels. Therefore, desired change is occurring for students at all performance levels.[4]
    - Mean Performance Level. Students are assigned to performance levels based on CST scores. On a scale of "1" (Far Below Basic) to "5" (Advanced), the average performance level per student has increased from 2.39 to 2.85 over four years, half a performance level category.
2. The state population of non-Reading First elementary schools had smaller gains than Reading First schools, especially for poorly performing students, though this particular observation requires caution since the two groups are not strictly comparable.

[3] The RFAI is a weighted index based on (1) the California Standards Test (CST), (2) a norm-referenced test given in grade 3 (CAT/6), and (3) the Reading First End-of-Year (EOY) tests. It serves as a metric for measuring progress of individual schools in reading achievement and the goals of Reading First. RFAI scores are available for all participating Reading First schools as of the 2003-2004 school year. This index has utility for measuring the impact of Reading First at the individual school level and at the district level. The official use of the RFAI as an index for determining which districts have made significant progress and warrant continued funding under Reading First has been approved by the State Board of Education.

[4] We do not report on trends in the "Basic" category because increases and decreases in students scoring "Basic" have no clear achievement implication.

- "Proficient" and above. During the same four year period, from 2003 to 2006, non-Reading First elementary schools grew 13.6 percentage points in percent "Proficient" and "Advanced", compared to 15.7 points for all Reading First schools and 16.4 points for high implementing Reading First schools.
- "Below Basic" and below. For the same period, the number of students in non-Reading First schools performing "Below Basic" and "Far Below Basic" declined 7.7 percentage points. That decline was 15.7 for Reading First schools and 16.0 points for high implementing Reading First schools – twice as much.

Table ES.1.0 and Figures ES.1.1 and 1.2 present Grade 2 CST gains and trend-lines for schools that have been in Reading First since school year 2002-2003. Included is the population of non-Reading First elementary schools, adjusted in the trend-line graphs to have the same starting point as the Reading First schools. The purpose of adjusting the starting point is to permit comparisons of rates and levels of change rather than of actual achievement levels. The tables break out high-implementing and low-implementing Reading First schools (see footnotes to Table ES.1.0 for how they are defined) and include statistics for a statistical control group.[5]

The tables and figures do not include statistics or trend-lines for students in the "Basic" category because increases and decreases in this category have no clear achievement implications. The tables and figures are only for schools that have been in the program for four years. Results for the remaining schools are provided in the report.

[5] The "statistical control group" is a construct defined by using multiple regression to hold the effects of school population characteristics stable while examining the independent effect of the Reading First program implementation statistic (RFII) on student achievement. For purposes of this discussion, the results of these analyses are referred to as a "statistical control group" because this approach is analogous to creating a hypothetical control group of schools that are exactly like the Reading First schools, in terms of student characteristics, but without the influence of the Reading First program. Please see Chapter 2 for a more complete discussion.

Table ES.1.0: CSTs, Grade 2, 2002 – 2006, Average Percent of Students in Cohort 1 Reading First Schools in the Program for Four Years and non-Reading First Schools Scoring at Various Performance Levels and Mean Performance Level per Student

	Reading First Schools (in the program 4 years)				
	All Reading First Schools	High Implementation Schools[1]	Low Implementation Schools[1]	Statistical Control Group[3]	All Non-Reading First Elementary Schools[4]
Number of Schools (2006)[2]	267	162	105	N/A	4,096
Number of Students (2006)	30,945	17,252	13,693	N/A	371,736
% Proficient & Above					
2002	15.5	16.0	14.7	15.5	37.7
2006	31.2	32.5	29.3	27.3	51.3
Gain[6]	**15.7**	**16.4**	**14.6**	**11.8**	**13.6**
% Below Basic & Far Below Basic					
2002	54.2	53.0	56.1	54.2	30.7
2006	38.5	37.0	40.8	41.6	23.0
Gain[6]	**-15.7**	**-16.0**	**-15.3**	**-12.6**	**-7.7**
Mean Performance Level Per Student[5]					
2002	2.39	2.42	2.35	2.39	3.07
2006	2.85	2.90	2.79	2.75	3.41
Gain[6]	**0.46**	**0.47**	**0.44**	**0.36**	**0.35**

1. High Implementation schools are those with an RFII greater than the mean RFII for 2004, which was 36. Low Implementation schools have an RFII less than 36.
2. The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.
3. Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First school.
4. Elementary Schools refers to all non-Reading First elementary schools in California.
5. Mean Performance Level is on a scale from 1 to 5, where: 1 = Far Below Basic, 2 = Below Basic, 3 = Basic, 4 = Proficient, 5 = Advanced.
6. The gain statistics do not exactly match the difference between the 2002 and 2006 means because they are calculated differently, as the mean of individual school gains rather than as a difference between the means of the 2002 and 2006 school populations.

Figure ES.1.1: CSTs, Grade 2, Cohort 1 Schools, in Reading First for Four Years, 2002 – 2006: Percent of Students Proficient and Advanced

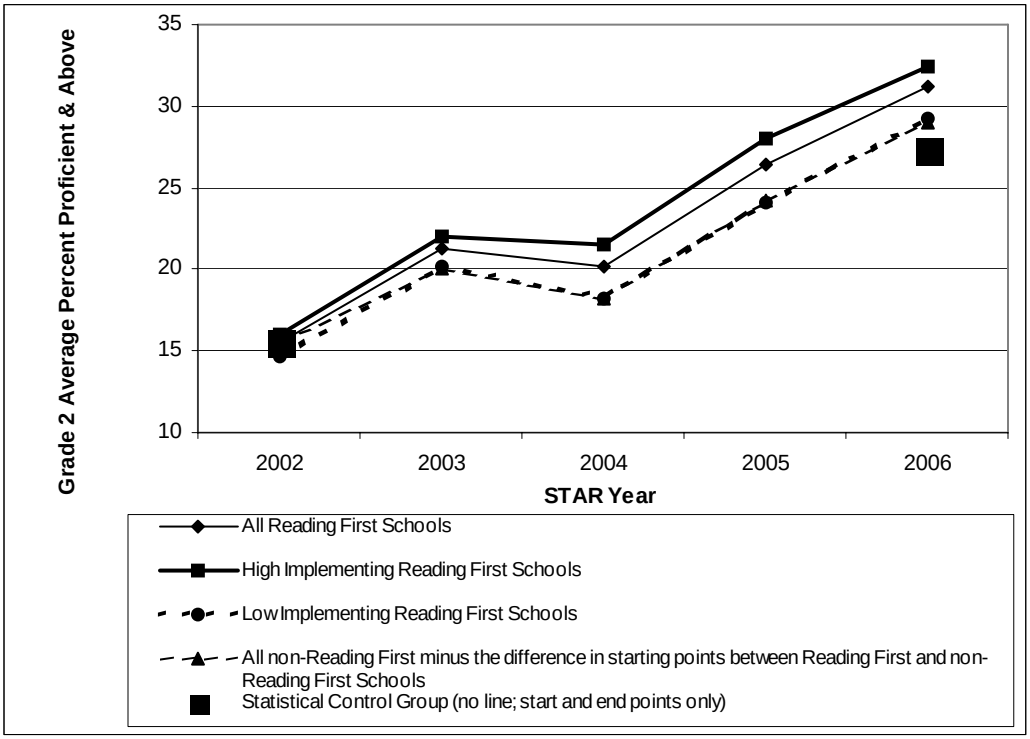
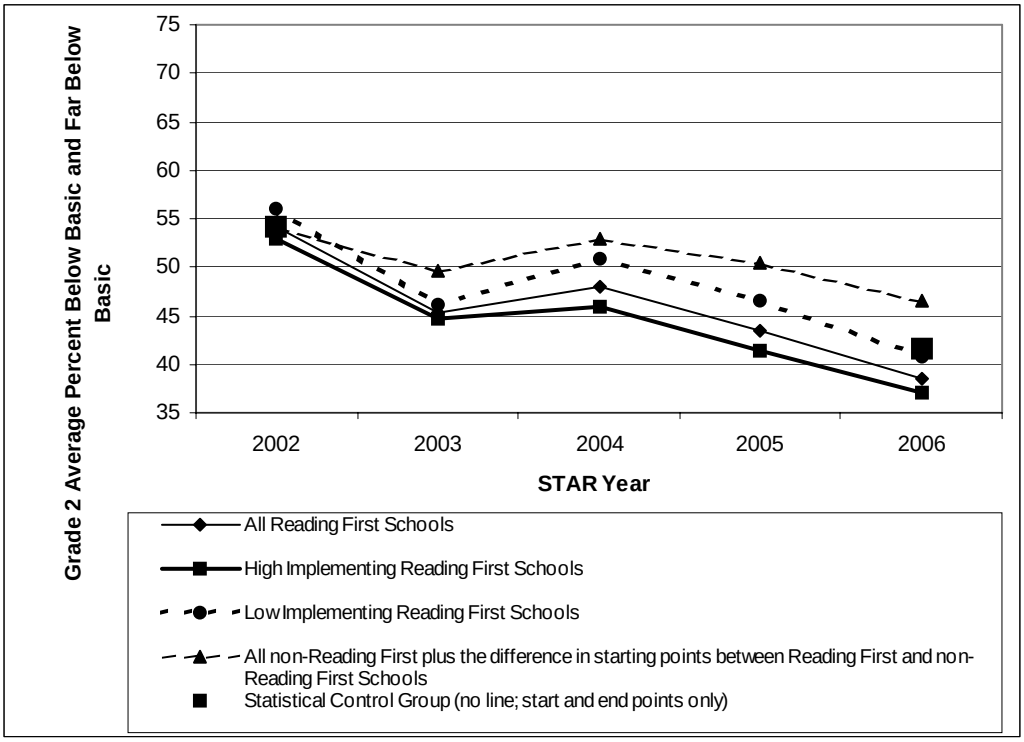


Figure ES.1.2: CSTs, Grade 2, Cohort 1 Schools, in Reading First for Four Years, 2002 – 2006: Percent of Students Below Basic and Far Below Basic



Key Finding #2: Reading 1st schools with high proportions of either Socio-Economically Disadvantaged students or English Learners showed similar patterns of reading achievement compared to the overall patterns among all Reading First schools.

The effect of implementing Reading First is statistically independent of whether schools have high percentages of English Learners (ELs) or Socio-Economically Disadvantaged (SED) students. In other words, the patterns of achievement are similar for schools with high-EL and high-SED populations as for all Reading First schools.

Table ES.2.0 shows Grade 2 CST gains for schools with a majority (>50%) of English Learners. Figures ES.2.1 and ES.2.2 provide trend-lines for schools with a majority of English Learners. The gain statistics are similar to those shown above for the whole Reading First population.

Table ES.2.0: >50% English Learners, CSTs, Grade 2, Cohort 1 Schools, in Reading First for Four Years, 2002 – 2006: Average % of Students in Reading First and non-Reading First Schools Scoring at Various Performance Levels and Mean Performance Level per Student

Schools with more than 50% English Learners	Majority EL Reading First Schools (in the program 4 years)				
	All Reading First Schools	High Implementation Schools[1]	Low Implementation Schools[1]	Statistical Control Group[3]	All Non-Reading First Elementary Schools[4]
Number of Schools (2006)[2]	186	104	82	N/A	683
Number of Students (2006)	24,370	12,775	11,595	N/A	70,376
% Proficient & Above					
2002	14.2	15.0	13.2	14.2	18.3
2006	30.6	32.0	28.7	27.0	33.2
Gain[6]	**16.4**	**17.0**	**15.5**	**12.8**	**14.9**
% Below Basic & Far Below Basic					
2002	55.8	54.4	57.6	55.8	51.1
2006	39.2	37.4	41.5	42.0	36.6
Gain[6]	**-16.6**	**-17.0**	**-16.1**	**-13.8**	**-14.5**
Mean Performance Level Per Student[5]					
2002	2.35	2.39	2.30	2.35	2.48
2006	2.83	2.88	2.78	2.75	2.92
Gain[6]	**0.49**	**0.49**	**0.48**	**0.40**	**0.44**

1. High Implementation schools are those with an RFII greater than the mean RFII for 2004, which was 36. Low Implementation schools have an RFII less than 36.
2. The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.
3. Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First school.
4. Elementary Schools refers to all non-Reading First elementary schools in California.
5. Mean Performance Level is on a scale from 1 to 5, where: 1 = Far Below Basic, 2 = Below Basic, 3 = Basic, 4 = Proficient, 5 = Advanced.
6. The gain statistics do not exactly match the difference between the 2002 and 2006 means because they are calculated differently, as the mean of individual school gains rather than as a difference between the means of the 2002 and 2006 school populations.

Figure ES.2.0: Schools with Majority of English Learners, CSTs, Grade 2, Cohort 1 Schools, in Reading First for Four Years, 2002 – 2006, Percent of Students Proficient & Advanced

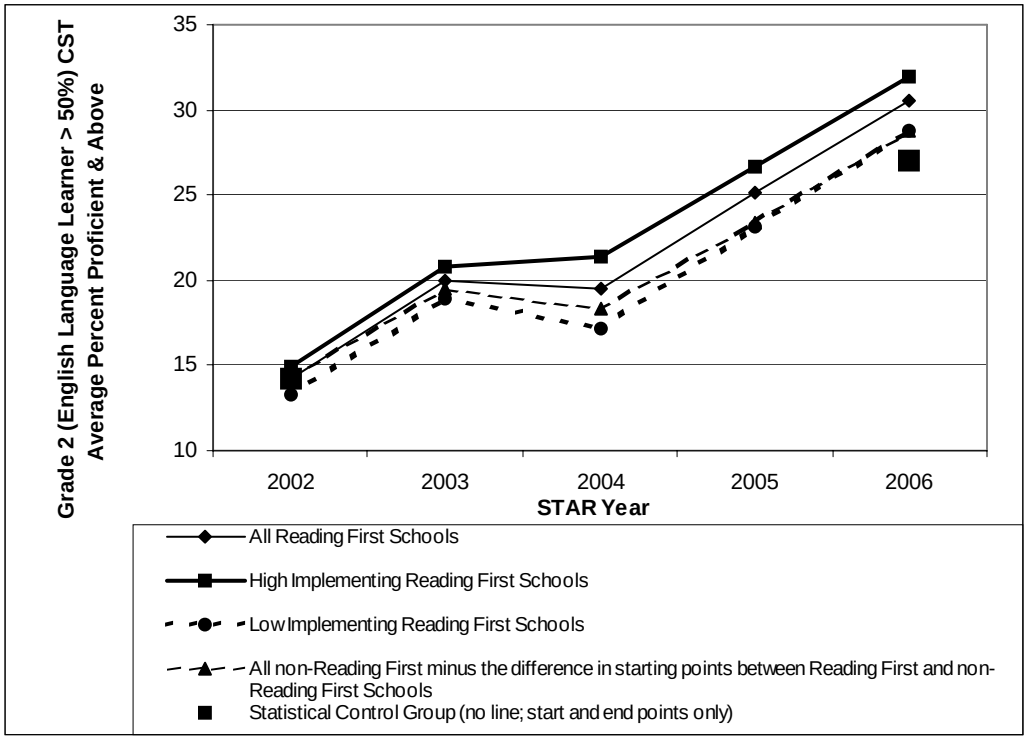
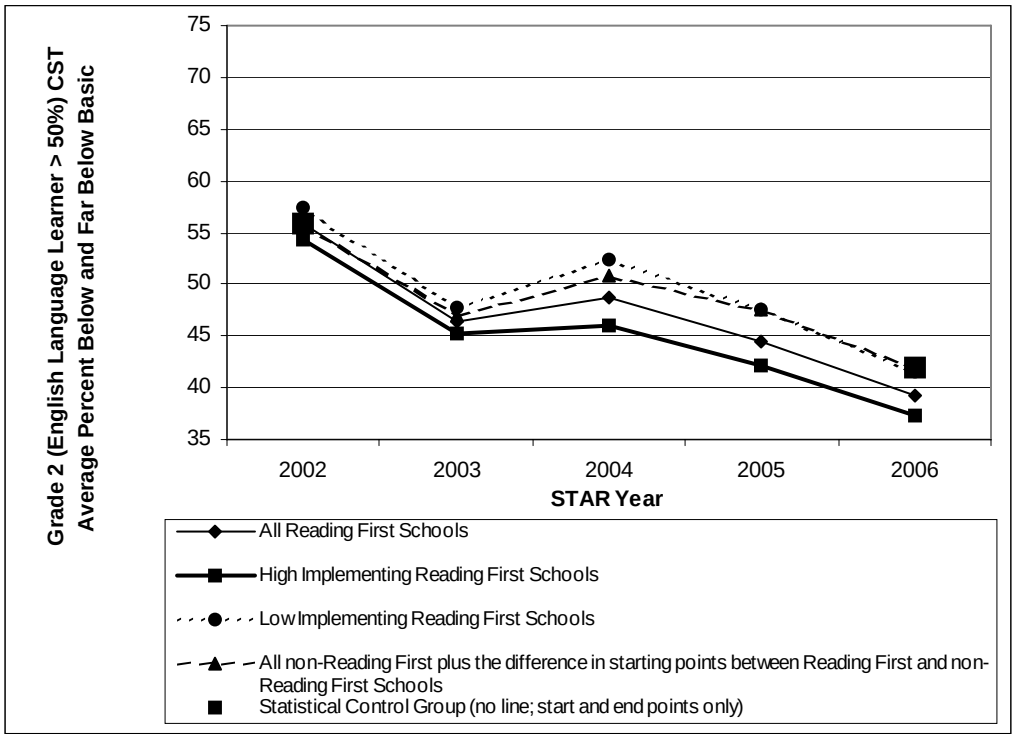


Figure ES.2.1: Schools with Majority of English Learners, CSTs, Grade 2, Cohort 1 Schools, in Reading First for Four Years, 2002 – 2006, Percent of Students Basic and Far Below Basic



Key Finding #3: Based on analysis of the Reading First implementation surveys, the elements of Reading First are being adequately implemented in Reading First schools, although there is room for improvement. The level of implementation of the Reading First program for each participating school is measured using the "Reading First Implementation Index" (RFII). [6]

To assess implementation, we administered a survey to every Reading First teacher, coach, and principal in a Reading First school in the springs of 2004, 2005, and 2006 (approximately 20,000 surveys each year). In addition, we conducted focus groups and interviews with a sample of school principals and LEA administrators and collected extensive qualitative data. The survey data were analyzed to produce an implementation statistic for each school on a 0-100 scale called the Reading First Implementation Index, or RFII.

We find that the basic elements of Reading First are being adequately implemented in Reading First schools, but it continues to be by a relatively slim margin. There is much room for improvement.

- On average, schools were considered to be "adequate" or better on about 62% of the questions in the survey as rated by Reading First teachers.
- Schools were considered to be "more than adequate" on about 39% of the questions. This corresponds to a 2006 mean RFII of 39. The 2005 mean RFII was 36, as it was for 2004.

Key Finding #4: The level of implementation of the Reading First program, as measured by the "Reading First Implementation Index" (RFII), is a statistically significant predictor of school growth in academic achievement after controlling for other characteristics.

Table ES.1.0 above reports that schools that are "high implementers" over four years of Reading First show a 4.6 percentage point advantage over a "comparable non-Reading First" statistical control group in moving students into the "Proficient" and "Advanced" performance levels. They show a 3.4 point advantage in moving students out of the "Below Basic" and "Far Below Basic" performance levels.

These findings are summarized using multiple regression, a statistical method for identifying which "predictor" variables best explain a specified outcome variable. In explaining the percent of students in 2006 who scored "Proficient" and above on the Grade 2 CST's, we find[7]:

- Effect of total Reading First implementation. A school's total Reading First implementation is equal to the sum of its implementation measures across the years it has been in the

[6] The RFII, described in Chapter 3, is computed at the school level from surveys of all Reading First teachers, coaches, and principals.

[7] We measure the effect of a predictor variable on achievement using the "standardized beta coefficient," which is the number of standard deviation units that the outcome variable can be expected to increase given a one standard deviation unit increase in the predictor variable.

program. The effect of total implementation on 2006 achievement is 0.21 (called a standardized beta coefficient). This means that for every one unit of increase in total implementation, a school can expect to see a statistically significant ($p<0.05$) increase of 0.21 units in the percent of students scoring "Proficient" and above.

  - In real terms, this 0.21 effect means that a Reading First school with average or above average program implementation can, after four years, expect to see an increase of 3-7 percentage points in the percent of students scoring "Proficient" and above as compared to the statistical control group. This effect manifests in Figure ES.1.1 as the increasing gap between the diverging trend-lines.

- Replication Across Achievement Metrics. Participation in Reading Firs is a significant predictor of reading achievement across various methods of measuring achievement outcomes. The variance explained by the regression models (the R-squared statistic) ranges from 0.33 for the Grade 2 CST's percent "Proficient" and above, to 0.63 for the RFAI.

The fact that implementation of the Reading First program is a statistically significant predictor of school achievement over time supports our finding that the Reading First program is effective.

Key Finding #5: It takes one to two years for schools to begin fully implementing the Reading First program and about the same length of time to show an effect on achievement.

Schools that have been in the program for three or four years have stable rising implementation measures. Schools in their first year tend to have low implementation statistics that increase in the second year. For example, schools that finished their second year of implementation in 2006 showed a 5-point jump in their RFII statistic relative to 2005. This represents a correction for a body of schools about which we registered concern in the 2005 (Year 3) Report.

Further, as is evident in Figures ES.1.1 and ES.1.2, high implementing schools have similar achievement effects as low implementing schools at the end of the first year of achievement. It is not until the second year of implementation that high implementing schools start out-performing low implementing schools.

Key Finding #6: The process of achieving significant change in reading instruction is complex, as demonstrated by portraits of Reading First implementation derived from interviews, focus groups and open-ended survey questions.

- According to district and school administrators, the coaching model is integral to successful implementation and "buy-in" of school personnel. Reading/literacy coaches, on the "frontline" of instruction, are trusted allies of site administrators and teachers and play a

pivotal role in supporting implementation by providing ongoing feedback and professional development. Coaches were viewed as highly resourceful, competent, and collaborative.

- Reading First has brought coherence and focus to LEAs and individual schools, described by district administrators and school principals as a "paradigm shift." It takes over a year, often two or more years, for schools and LEAs to achieve coherence, and it often happens after initial struggles and resistance among staff. Having consistency of curriculum, training, structure of support, and assessments across grades within a school, and across schools within a district, gives teachers and administrators a common language and perspective for achieving program goals.
- The accountability and monitoring included in Reading First has had a positive impact on schools. Administrators and coaches work together to observe instruction and provide support when needed. The monitoring of student performance through curriculum-embedded assessments facilitates buy-in among staff and leads to teachers' authentic use of data to adjust instruction as needed.
- Despite some teachers' negative perceptions of the appropriateness of the curriculum for English Learners, the program has had a positive impact on EL student reading achievement and language development, and administrators view the program as beneficial for this population.

Policy Recommendations

Policy Recommendation 1

Continue to focus on full implementation of Reading First.

Review of school RFII measures and the qualitative data suggests that Reading First is being adequately implemented across the state in essential respects, in particular professional development, program materials, coaching, and instructional practices. Our findings show that high-implementing schools yield higher academic gains than low-implementing schools. State and local Reading First personnel should focus on schools with lower levels of implementation to ensure maximum benefit from participation in Reading First.

Policy Recommendation 2

Support participation in Reading First over multiple years. It takes at least two years of implementation to show significant achievement gains and achieve a coherent focus, even with extensive training and support. Continued support beyond the initial two years is essential to achieving significant results and

establishing the long-term institutional changes needed for Reading First instructional practices to continue after funding is discontinued.

Quantitative and qualitative data support the notion that continued program participation leads to continued improvement in teaching practices and student outcomes. Extended time (e.g., 6 years) will allow school personnel to gain depth of knowledge, refine their skills, and integrate program principles into the fabric of their school operations. Schools in the program since 2003 continue to show steady academic gains, supporting the premise that extended support and participation leads to continued improvement.

Policy Recommendation 3

Provide for a statewide data collection effort to facilitate accurate comparisons of student achievement across Reading First and non-Reading First schools, focusing on their use of the components that are required in Reading First. A statewide database of teacher and school data would confirm or refute the hypothesis that the statewide trend toward higher student proficiency is the result of the voluntary adoption of program elements that happen to be required as part of Reading First. Such data collection efforts would also inform policy over many years by facilitating future program evaluations.

Because Reading First schools agree to implement reading programs that are consistent with Reading First guidelines, the gains experienced by Reading First schools are likely the result of implementing the program. What, then, is causing the gains observed in non-Reading First schools over the same period? One possible hypothesis is that these gains are caused by non-Reading First schools adopting the same program elements that happen to be required in Reading First schools. The hypothesis that recent achievement gains in non-Reading First schools are the result of implementation of similar program elements should be confirmed or refuted.

Policy Recommendation 4

Continue to support the extensive and focused professional development provided to teachers, coaches and administrators. Continue to support the coaching model as a means for achieving instructional coherence and implementation of research-based instruction.

Reading instruction is complex and requires a high level of expertise among teachers and administrators. Intensive professional development coupled with ongoing support of implementation is a highly effective model for achieving change in schools, when it is continued and supported over time. According to comments collected from principals, administrators and teachers, the coaching model bridges an important gap among policy, administration and classroom instruction.

Table of Contents

List of Figures

List of Tables

List of Appendices

Chapter 1: Introduction

The federal initiative called Reading First was authorized in 2001 as part of the federal *No Child Left Behind Act*. Reading First, intended to improve reading outcomes in the nation, promotes the use of instructional practices based on scientifically based reading research in Grades K-3. On August 23, 2002, the State of California was approved to receive over $900 million over a six-year period. According to federal Reading First guidelines, continued funding for states depends on demonstrating "significant progress" toward the goal that all children learn to read on grade level by the third grade. California's successful application for federal Reading First funds met strict federal criteria by outlining a plan for assessments, instruction, materials, professional development, monitoring and oversight. This report evaluates implementation and academic achievement in California's Reading First program during the first four years of funding.

In four years of implementation, from 2002 to 2006, California received approximately $583 million for the Reading First program. California received approximately $133 million in Year 1 (2002-2003), $146 million in Year 2 (2003-2004), $152 million in Year 3 (2004-2005) and $152 million in Year 4 (2005-2006). With Reading First funds, California has established a system to provide training, assist school districts in acquiring supplemental curricular materials, monitor progress toward goals, and provide technical assistance to participating schools and school districts.

California's Reading First plan delineates the roles and operational procedures for personnel involved at the state and local levels. The State Board of Education (SBE), Office of the Secretary of Education (OSE), and the California Department of Education (CDE) direct the Reading First program in California. The Reading and Literacy Partnership Team, with membership broadly representing the interests of reading education in the state, serves an advisory role for Reading First. The California Technical Assistance Center (C-TAC) has responsibility for operating state and regional technical assistance centers. In addition, the Evaluation Advisory Group (EAG) was appointed to advise the external evaluator.

Local educational agencies (LEAs), or school districts, are required to use one of California's state-adopted reading programs: SRA/McGraw-Hill's *Open Court Reading 2000* or *2002* (OCR) or the Houghton Mifflin *Reading: A Legacy of Literacy 2003* (HM). California law (Proposition 227) mandates instruction in English for all students unless parents sign a waiver requesting bilingual instruction. In this study, classes taught in Spanish are referred to as "waiver" classrooms. In 2004, the State approved for use in Reading First schools two Spanish language reading programs that are Spanish versions of the State's adopted programs written in English that reflect research-based instructional principles: SRA/McGraw Hill's *Foro abierto para la lectura* and Houghton Mifflin's *Lectura: Herencia y futuro.*

Students receiving reading instruction in Spanish must participate in the statewide testing program in English (Grades 3 through 11). In the 2004-2005 school year, California's Reading First program began offering support and training for schools with waiver classrooms using the adopted Spanish language programs.

The Reading First program is designed to build state and local capacity in reading education. Through California's Reading First program, the state has established a network of support. Each year, the Reading First, California Technical Assistance Center, or C-TAC, housed at the Sacramento County Office of Education Reading Lions Center, has conducted three, one-day state meetings for district-level administrators; an annual "Superintendents' Summit;" two, two-day statewide professional development trainings for more than 1,300 Reading First coaches; four, one-day training of instructor sessions for ten modules appropriate for site administrators; distributed the progress monitoring assessments (Kindergarten, 6-8 Week Skills Assessments for Grades 1-3, and End-of-Year Assessments), and ensured the use of an online assessment reporting system by each Reading First LEA. The C-TAC also oversees the work of the eight Regional Technical Assistance Centers (R-TACs) that are geographically disbursed to county offices of education throughout the state. The R-TACs provide district-level technical assistance to Reading First LEAs on various aspects of implementation of the state adopted reading programs, professional development, assessments, the use of data for monitoring progress, and coaching strategies connected to the adopted instructional programs. C-TAC and the R-TACs work with local Reading First LEAs in California; the reader may visit http://www.calread.net for further information.

California has now completed four years of implementation of the Reading First program. Each year, additional schools have been included in cohorts receiving funding. The first year, 2002-2003, can be characterized as a start-up year because LEAs did not have a full year in which to implement. In November 2002, Cohort 1 districts received approval in the first round of competitive grant applications and the funding was disbursed to the school districts in February 2003 or after. Cohort 1 has been receiving funding and implementing the program for approximately three and one-half years. Districts in Cohort 2 were selected in a competitive grant process in the spring of 2003 and schools have received three years of funding. A third round of approvals in the spring of 2004 led to the selection of Cohort 3 districts. Each cohort has lost a few schools due to such factors as school closures or mergers, and districts may have replaced them with new schools. As of 2006, Cohort 1 consisted of 343 schools, Cohort 2 included 372, and Cohort 3 had 144 for a total of 859 schools in 110 districts. These are the schools included in Reading First in Year 4 and are included in this evaluation study.

Components of the California Reading First Plan

The California Reading First Plan is based on a series of Assurances that are implemented by the LEAs. CDE, C-TAC, and the R-TACs monitor and assist LEAs in implementing the Assurances. The Reading

First program is designed to ensure full implementation with fidelity to a comprehensive research-based reading program.

Following is a description of some of the key LEA Assurances and the components of Reading First that address them.

Vision Statement

Each LEA and participating school must articulate a vision statement that reflects the goals and objectives of Reading First, including the belief that all children can learn to read with adequate instruction.

Instructional Program

Each LEA is required to implement fully the district's state-adopted reading/language arts program for an uninterrupted 60 minutes per day in Kindergarten and 150 minutes per day in Grades 1-3, according to a district-approved pacing plan that outlines when each daily lesson is taught at each grade level in an academic year. This plan not only assures that students will complete the grade-level curriculum but also that implementation occurs systematically throughout every classroom at each grade in every Reading First school. Districts may select an alternative Spanish-language version for approved waiver classrooms where instruction occurs in Spanish. Districts may only use supplemental instructional or technology programs that support the scientifically-based reading instruction of the core program.

Professional Development

LEAs must assure that all K-3 and special education K-12 teachers in Reading First schools participate in professional development focused on the adopted reading program during each year of the program. During their first year of participation in Reading First, teachers attend a state-approved training as mandated in AB 466. This is a 5-day, 40-hour training that focuses on in-depth knowledge of the LEA's adopted research-based reading program. In districts that have been involved in Reading First for multiple years, teachers who are new to Reading First attend the AB 466 training. In these trainings, teachers learn about the design, content and instructional strategies included in the program. In subsequent years, the LEAs must provide advanced level professional development to provide in-depth knowledge of research-based instructional strategies. Teachers participate in 80 hours of district-sponsored follow-up professional development. For this, the LEAs are required to develop and provide advanced level professional development programs or contract with the state's available training curricula. Principals in their first year of assignment in Reading First schools must attend state-approved AB 75 training in reading curriculum for grades K-3.

Curriculum-Embedded Assessment

LEAs, as of 2004-2005, are required to develop and implement an assessment plan that includes the use of frequent, curriculum-embedded assessments for program monitoring that are distributed through the C-TAC (i.e., the 6-8 Weeks Skills Assessments). Teachers use these assessments to determine student progress and the efficacy of the delivery of instruction in either the English or Spanish version of the reading program. The assessments are aligned with the unit/theme content of the adopted reading program. For kindergarten, there are three assessments (fall, mid-program, and spring) on naming uppercase and lowercase letters, high frequency words, rhyming words, oral phoneme blending, oral phoneme segmentation, consonants and short vowel sounds, and phoneme replacement of medial vowels. For first grade, there are six assessments, with the first three covering spelling, word reading and writing; and the last three covering fluency, spelling, word reading, reading comprehension and writing. For grades two and three, there are six assessments covering fluency, comprehension, usage, spelling, vocabulary, and writing. The LEAs, also as of 2004-2005, are required to use an online assessment reporting system for these assessments. Teachers, administrators, and coaches use the data to make instructional adjustments and to identify groups of students who need extra assistance. Within the last three weeks of the academic year, the LEAs are required to administer the End-of-Year tests (EOY) that are also curriculum-based. For kindergarten, there are eight subtests that measure skills similar to those measured in the spring assessment; and for grades 1-3, there is an oral reading fluency assessment that is similar to, but separate from, the unit/theme assessments. The results of these assessments are used as part of the Reading First Achievement Index (RFAI; see Chapter 4 of this report). The LEAs' Internal Evaluation Reports also use these End-of-Year assessments along with the final unit/theme assessments and the STAR:CST results to provide information about students' reading performance and to evaluate the impact of Reading First at the school level.

Collaborative Teacher Meetings

All Reading First schools are required to hold regular grade-level meetings to provide an opportunity for teachers to work together to refine their skills in the implementation of the program. These meetings focus on studying assessment results from the curriculum-embedded assessments, conducting lesson study, co-planning delivery of specific parts of lessons, and adjusting instruction to better meet students' needs. School principals and reading coaches are encouraged to assist in facilitating and supporting these meetings.

District Commitment

Each LEA is required to conduct an internal evaluation on the effectiveness of its implementation of the Reading First program. As part of this effort, district personnel must make regular site visits to monitor

implementation and adherence to the purposes of Reading First. The LEA must assure that its administrators overseeing curriculum and instruction and Title I programs have established and reinforced policy guiding consistent implementation of the adopted instructional reading program, including the allocation of instructional time, the use of instructional materials and the use of selected assessments. In addition, district personnel must assure that the Reading First program is well coordinated with other programs such as Title I, English Language Acquisition and Special Education. Each LEA must have a district Reading First Leadership Team that meets regularly to advise and support the program.

Coaching

LEAs may use Reading First funds to provide reading coaches, content experts, and coach coordinators. The Reading First program encourages LEAs to utilize coaches or content experts solely to provide site-based support of the adopted reading program. The Reading First Assurances encourage the districts to ensure that coaches have adequate training and support to serve as peer coaches for reading instruction in grades K-3. Coaches are to spend a significant amount of time in classrooms, observing and supporting instruction, including conducting demonstration lessons. Four two-day Coach Institutes were offered by C-TAC during 2005-06. At the Coach's Institutes, delivered both in the southern and northern regions of California, coaches, content experts and district coach coordinators not only developed expertise in the adopted reading program at all K-3 grade levels, but also learned how to provide consultation and serve in a leadership role regarding reading instruction. C-TAC developed and provided instruction in eight new modules for a total of 48 modules since 2003-2004. For a description of each module, the reader may visit http://www.calread.net/coach_institutes. LEAs may request technical assistance from the R-TACs on how to support coaches.

Site Leadership

The site administrator's role is clearly defined by Reading First. The school principal (and assistant principal, if applicable) must support the full implementation of the school's adopted reading program. Duties include protecting the daily instructional time allocated to reading instruction, monitoring student assessment results, working with coaches to address any problems that may arise with implementation, and conducting classroom observations. Administrators must attend the state's 40-hour AB 75, Module 1 training program to become fully knowledgeable of the reading program. They are also required to participate in an additional 40 hours of aligned activities within a two-year period.

Program Coherence

Reading First schools must ensure that any supplemental programs or materials are fully aligned with the adopted reading program. All categorical programs such as English Language Acquisition, Title I, School Improvement, and Special Education programs, must be coordinated with the core program. If

supplemental materials are used, the schools must demonstrate that they are aligned with the core program.

State Leadership

The CDE has designated key personnel to oversee and facilitate the administration of grants, the contract with the external evaluator, and communications and legislation for the Reading First program. The SBE has several roles. It serves as the State Educational Agency for Reading First. It works collaboratively with the CDE and the governor's office to develop and approve policy decisions that affect the program's implementation. It approves the external evaluation report and the required federal reports. It also authorized the C-TAC to oversee the technical assistance work of the R-TACs and to provide direct technical assistance to the LEAs. This structure provides opportunities for consistent communication between the state leadership (CDE and SBE) and the Reading First LEAs.

Technical Assistance

The CDE annually provides grants to eight R-TACs housed in county offices of education (Alameda, Butte, Los Angeles, Sacramento, San Bernardino, San Diego, San Joaquin, and Santa Barbara). In these grants, funds are set aside to support the work that is assigned to the C-TAC. In addition, CDE contracts with the C-TAC to meet requirements of state Reading First legislation. These requirements include developing materials and assessments for training teachers, ensuring that professional development for teachers and instruction for pupils is consistent in quality and delivery, and providing assistance to the R-TACs. It is through this contract and the funds set aside by the R-TACs and provided to C-TAC that advanced levels of teacher and coach training materials are developed and available for the following year's professional development program options.

California Reading First Evaluation Study

The California Reading First Plan, approved by the U. S. Department of Education (USDE), included an annual evaluation to study the implementation of the program and to evaluate program outcomes. A Request for Proposals (RFP) for prospective vendors for this study was issued in the spring of 2003, and in June 2003, the SBE approved Educational Data Systems (EDS[8]) as the contractor for the Reading First evaluation study. The contract was finalized in November 2003. To date, EDS has completed the Year 1 Report including outcomes for the 2002-03 school year, the Year 2 Report, reporting outcomes for the 2003-04 school year and a cumulative two-year analysis, and the Year 3 Report, examining outcomes for the 2004-2005 academic year and cumulative effects from 2002-2005. This current report represents the Year 4 Report, and will include outcomes from the 2005-2006 academic year and cumulative effects.

[8] EDS is a registered trademark of Electronic Data Systems. However, in the context of this document, EDS refers exclusively to Educational Data Systems, Inc.

Throughout this report, the five questions outlined in the Reading First evaluation study RFP are addressed. The five questions may be collapsed into two overarching issues for which California policymakers seek information: First, how well has the Reading First program been implemented in participating schools and districts? And second, what impact has the Reading First program had on participating schools?

Under program implementation, the questions specifically ask:

(a) How well did participating districts and schools implement their Reading First grants in accordance with California's Reading First plan? (See Chapter 3.)

(b) How effective are the resources, support and professional development activities that district-level administrative staff, school site administrators, and classroom teachers have received in implementing the Reading First grants? (See Chapters 3 and 5, and Appendices B, C, D, and E.)

Under program impact, the guiding questions ask:

(a) What is the impact of the Reading First program on K-3 students in participating districts and schools? (See Chapter 4.)

(b) What evidence is there that the Reading First program has improved the effectiveness of participating schools and districts? (See Chapters 4 and 5, and Appendix H.)

(c) Have any unintended consequences resulted from the implementation of the Reading First Program? (See Chapters 3 and 5, and Appendices B, C, and D.)

Evaluation Study Design

The EDS proposal for the Reading First evaluation study identified a wide range of variables available for addressing the guiding questions. These variables may be broadly grouped into three categories:

1. School Characteristics. These data include demographic characteristics for the students served by Reading First schools, including information about the socio-economic status of the school populations, ethnicity, and the degree of English language acquisition for each school. Data sources for student characteristics include statewide data collections, the California English Language Development Test (CELDT), and the Standardized Testing and Reporting (STAR) statewide testing program. The report also includes characteristics of school personnel including information about teacher qualifications and experience taken from the California Basic Educational Data System (CBEDS).

2. Implementation Measures. Program implementation was measured using comprehensive surveys given to teachers, principals and coaches of all Reading First schools. Secondary, qualitative sources of implementation data include open-ended questions on the surveys, individual interviews with district-level administrators, and a set of focus group interviews conducted with principals in selected districts in

2006. The implementation data include responses of teachers, reading coaches and principals regarding their own and their school's implementation of the Reading First program, including the use of scientifically based instructional materials, involvement in professional development, use of 6-8 week unit assessment data, and use of coaching services. For the 2003-04 school year (Year 2), EDS designed a set of three surveys (one each for Reading First classroom teachers, coaches, and school principals) and collected data from approximately 14,000 Reading First teachers, coaches, and school principals during the spring of 2004. For the 2004-2005 school year (Year 3), adjustments were made to the survey based on feedback from the EAG. Questions were added to capture information on the use of the Spanish versions of the curricular materials. The surveys collected data from approximately 20,200 teachers, coaches and principals in the spring of 2005. In 2006, minor adjustments were again made to the survey and approximately 20,000 participants responded.

3. Outcome Measures. Student academic progress in basic reading skills is of primary importance in evaluating the impact of Reading First on children in grades K-3. As shown in Table 1.1 below, this report includes student outcome data from three sources over the four years of implementation. The STAR testing program includes the California Standards Test (CST), designed specifically to measure the challenging English/Language Arts content standards adopted by the SBE in 1997, and a Norm-Referenced Test (NRT) for Grade 3, designed to measure a broader range of reading achievement. For each year of the evaluation study, California has used the California Achievement Test- 6th Edition Survey (CAT/6) for its NRT. An additional data source collected by C-TAC in their monitoring and assistance role includes the End-of-Year (EOY) assessment administered by classroom teachers, which measures oral fluency in Grades 1, 2 and 3 and phonological awareness and alphabetic knowledge in Kindergarten.

Table 1.1: Student Assessments Used in Reading First Evaluation Study

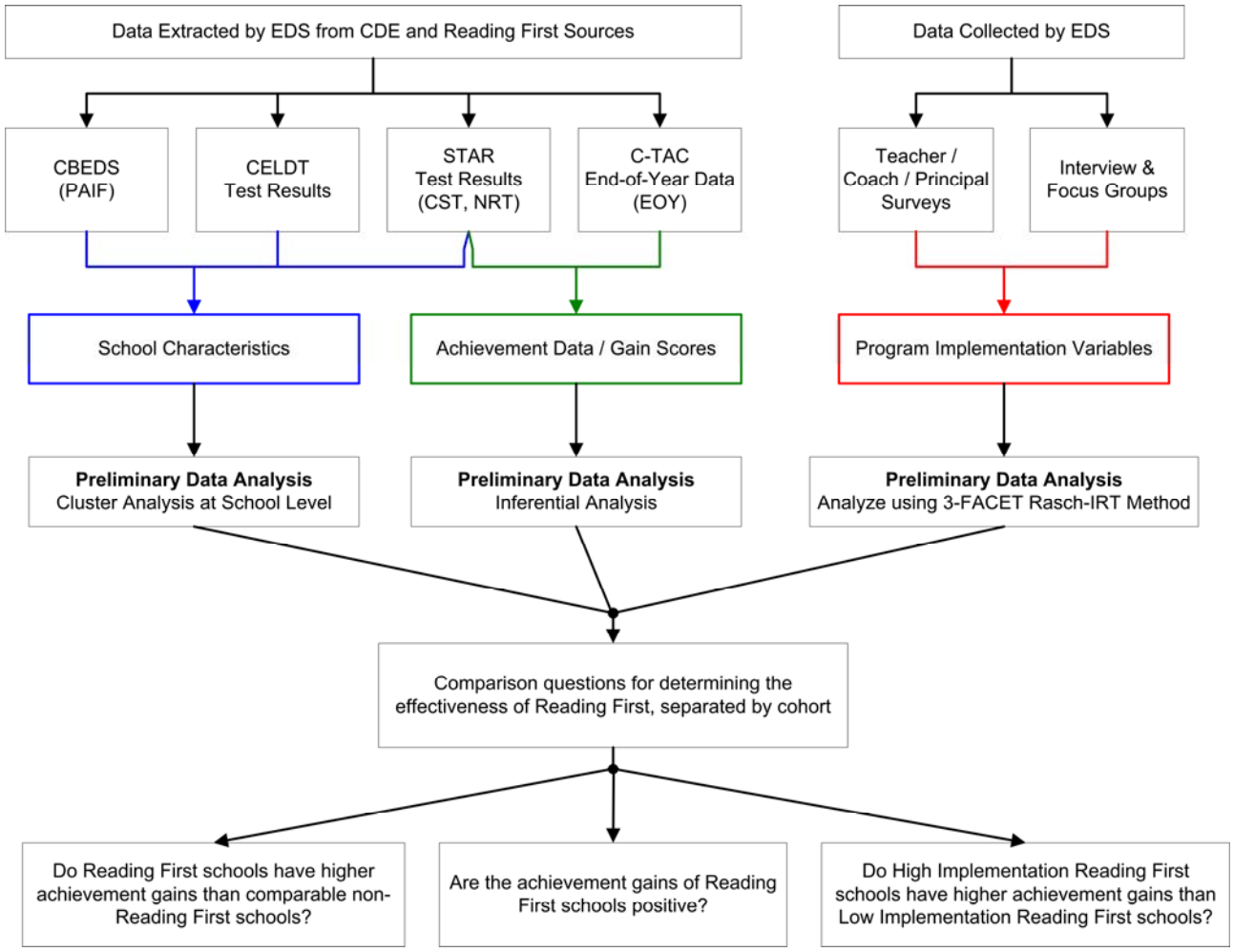
	EOY	CST	CAT/6
Kindergarten	X		
Grade 1	X		
Grade 2	X	X	
Grade 3	X	X	X

Note: The Kindergarten EOY test includes eight subtests: Naming Upper Case Letters, Naming Lower Case Letters, Matching Consonants and Sounds, Matching Short Vowels and Sounds, High Frequency Words, Decoding Consonant-Vowel-Consonant Words, Rhyming Words, and Phonemes in Words. The EOY tests for Grades 1, 2 and 3 consist of timed oral reading fluency passages. The CSTs include data from the English Language Arts component. The NRT used in California is CTB/McGraw-Hill's California Achievement Test, Version 6 (CAT/6) including the reading, language and spelling subtests.

The original conceptual framework presented in the EDS proposal for the Reading First evaluation study represented the overall plan for evaluating California's Reading First program. For the Reading First Year 1 Evaluation Report, several aspects of the original conceptual framework could not be addressed due to the lack of complete data. The Year 2 study allowed a more complete evaluation of all components, but the survey results were considered formative, as it was the first year of the instrument's use. The Year 2 survey results allowed for testing the instrument's psychometric properties and making adjustments and improvements for the Year 3 survey, used in spring 2005. The Year 4 survey maintained the psychometric properties. Surveys from Years 2 through 4 have been equated to allow cross-year comparisons.

The conceptual framework below illustrates the data sources for the Year 4 study and how they are grouped into demographic, implementation and student outcome variables. The framework then lists the types of statistical and qualitative analyses utilized to address each set of variables used to determine whether the Reading First program is effective. By relating school characteristics and program implementation to student outcomes, this study uses a quasi-experimental design and multiple approaches to addressing the research questions.

Figure 1.1: Conceptual Framework – Year 4



In this Year 4 Evaluation Report, Chapter 2 and Appendix A provide descriptive information for Reading First schools, a Statistical Comparison Group of schools, and all K-3 schools in California. It corresponds to the "School Characteristics" box in the conceptual framework. Chapter 3 derives program implementation variables, including the Reading First Implementation Index (RFII), from data collected from teacher, coach, and principal surveys (provided in Appendices B, C, and D). It corresponds to the "Program Implementation Variables" box in the conceptual framework. Information on the development of the survey and creation of the RFII, provided in previous years' reports, is included in Appendix E. Additional data sources include open-ended questions from the teacher survey, interviews with administrators and focus groups with Reading First school principals.

Chapter 4 provides information on student achievement measures, including CST, CAT-6, EOY, and RFAI gain scores for Reading First schools, a Statistical Comparison Group of schools, and all K-3 schools in California, and compares the achievement gains of High Implementation and Low Implementation Reading First schools. Chapter 4 also describes the construction of the Reading First

Achievement Index (RFAI) and its usefulness in determining significant progress.[9] Chapter 4 corresponds to the "Achievement Data/Gain Scores" box and to the three question boxes at the bottom of the Conceptual Framework. Additional analyses described in Chapter 4 address questions regarding the achievement of English learners in "waiver" classrooms, in which teachers provide instruction in Spanish and use materials written in Spanish, and non-waiver classrooms, in which teachers provide instruction in English and use materials written in English. Associated with Chapter 4, Appendix F provides results of a multivariate regression analysis of CST scores and reports disaggregated achievement gains. Appendix G describes how the RFAI is calculated. Appendix H provides a complete list of individual schools with selected achievement measures as required in Reading First reporting.

Chapter 5, corresponding to the "Program Implementation Variables" box of the conceptual framework, presents a qualitative, descriptive analysis of various narrative data sources. Administrators' views of implementation issues are examined using information from district administrators and focus group interviews conducted with principals of Reading First schools. Teachers' views of Reading First are also described, taken from the written comments of participant teachers in the teacher version of the survey conducted in 2005. This chapter examines what teachers view as positive and negative consequences of Reading First. This chapter also examines the implementation and efficacy of the coaching model used in Reading First using various sources of data. Also related to program implementation, this chapter examines the feasibility of sustaining the Reading First model beyond the grant funding.

[9] The use of the RFAI for determining whether schools have made significant progress and are eligible for continued funding has been approved by the State Board of Education.

Chapter 2: Demographic Characteristics of the Reading First Cohorts

Chapter 2 provides demographic profiles for the three Reading First LEA cohorts and for non-Reading First schools in California with the purpose of showing in what ways the various school populations are, and are not, comparable to each other. It explains the distinction between "LEA cohort" and "school years in program." It justifies the discontinuation of the "Comparison Schools" research strategy that is a feature of previous reports and introduces the idea of the "statistical control group" that we have adopted in its place.

Background and Overview

The Reading First program was first adopted during the 2002-2003 academic year by 13 school districts, known as LEAs. These 13 LEAs comprise what is called "Cohort 1", or "Round 1" since they received the first round of funding. In the 2003-2004 academic year, an additional 60 LEAs were funded to become Cohort 2. In the 2004-2005 academic year, 37 LEAs were funded to become Cohort 3. No new LEAs have been added to Reading First since the 2004-2005 academic year, so there is no Cohort 4. LEAs distribute their Reading First funds to schools within their districts that meet the eligibility requirements of Reading First.

LEA Cohorts versus School Years in Program (YIP)

In 2006, we clarify terminology to distinguish between LEAs, which are grouped into "LEA cohorts" according to the year they first received funding, and the schools within those LEAs which sometimes do not implement the program in the same year that their LEA first received funding. The schools are grouped according to "School Years in Program," called YIPs. Since we are focusing on LEA demographics in this chapter, LEAs and their member schools are organized by cohort. In Chapters 3 and 4 we focus on schools and their achievement, so we organize schools by YIP. In most cases schools are in the program the same number of years that their LEA is, in which case categorizing by YIP or cohort makes no difference. Schools in a Cohort 1 LEA tend to be YIP 4 schools. Schools in a Cohort 2 LEA tend to be YIP 3 schools. Schools in a Cohort 3 LEA tend to be YIP 2 schools. In some cases (e.g., Los Angeles Unified School District, which is a Cohort 1 LEA) there are schools that were added to the program as late as 2006 and thus have fewer years in the program.

No additional LEAs were added in 2006, so there are still only three LEA cohorts. However, within those cohorts some new schools have been added. There are 38 new Reading First schools in YIP 1.

Difficulties Establishing a Valid Comparison Group

In Years 1 and 2 of the evaluation, Reading First schools were compared to two groups of non-Reading First schools. One group consisted of Reading First Eligible schools that had not yet been accepted into

Reading First. The second group, consisting of demographically matched non-Reading First schools, was selected using cluster methodology and randomly selected from a list of schools in the state that most closely matched the Reading First schools for Socio-Economically Disadvantaged (SED) and English Learner (EL)[10] student population percentages.

In the Year 3 Report, it was found that the Reading First Eligible schools were not demographically representative of the schools that had already entered the program. In particular, the 2005 demographic data showed that on average 57% of the students in Reading First schools had limited English proficiency. Cohort 3 had the highest percentage of EL students (58.8%, compared to 57.6% for Cohort 1 and 56.2% for Cohort 2). The demographically matched schools were similar to the Reading First schools with regard to English Learners, with an average EL population of 57.5% in 2005, but the Reading First Eligible school group had much lower percentages of EL students, 47% in 2004 and 49.7% in 2005. This 8-10 percentage point difference would have had a significant effect on achievement comparisons. Proficiency in English is highly correlated with success in academics and with achievement scores (August & Hakuta, 1997; National Research Council, 2002). Given the large difference in the percentage of EL students between the Reading First and Reading First Eligible schools, there is little justification for treating the two groups as comparable for the purpose of this evaluation study. Thus, the Reading First Eligible schools were no longer used in comparative analyses.

Having rejected the Reading First Eligible Schools comparison group, the case was made in the Year 3 Report to use only the demographically matched group, renamed in the Year 3 report to "Comparison Group." All the gains tables in Chapter 4 of the Year 3 Report show scores for this demographically matched comparison group.

Difficulties with the Demographically Matched Comparison Group

In Year 4 several methodological problems became evident with the demographically matched comparison group. While it is more comparable to the Reading First schools in terms of EL and SED students, a review of the trend-lines showed that it sometimes had different, and higher, achievement starting points than the Reading First schools. Schools with higher achievement starting points can have different growth rates due to statistical artifacts resulting from the shape and position of the score distribution. Higher starting points can also signify secondary demographic characteristics that might effect growth, such as students having parents with a higher level of education. These effects are hard to control.

[10] Refer to Appendix A of this report or Year 1 and Year 2 Reading First Evaluation Reports for a detailed description of the selection of Comparison Group A and Comparison Group B schools.

More important, it became clear that the demographically matched schools, as would be true of almost any sample drawn from California elementary schools, is compromised by Reading First-like "treatments" to varying and unknown degrees. From the perspective of experimental design requirements, having both the control and experimental groups sharing even a portion of the same experimental treatment components seriously compromises the research design, making statistical comparisons between the groups unreliable as indicators of a true treatment effect.

For instance, an unknown number of non-Reading First schools have adopted the state-approved reading programs, the same Open Court and Houghton-Mifflin reading programs that Reading First uses. Non-Reading First schools that use these programs are automatically given access to AB 466 teacher professional development and AB 75 principal professional development. They also receive access to the 6-8 week skills assessments, which provide timely student diagnostic statistics to teachers. In addition, many non-Reading First schools have opted to hire their own reading coaches, though perhaps under a different coaching model. The main difference between Reading First schools and these non-Reading First schools is primarily the source of funding for their professional development activities and skills assessments, not their content.

Compromising the sample further, LEAs that adopt Reading First only apply those funds to that portion of the LEA's schools that are eligible under the Reading First legislation. Yet some district administrators reported in interviews a tendency to apply the same Reading First program components to the remaining non-eligible schools in the district in order to ensure uniformity and program coherence. Many schools that are classified as non-Reading First within a participating LEA have identical treatments as those that receive Reading First funds. This is in addition to the influence of state-approved materials on non-Reading First LEAs across the state. In the absence of implementation information from non-Reading First schools -- a level of data collection that falls outside the scope of the Reading First External Evaluator's contract -- the only certain difference between Reading First and non-Reading First schools is that Reading First schools receive funding for the program elements cited above, whereas non-Reading First schools (or their LEAs) have to pay for the educational program and the coach from another funding source if they opt to use the state-approved reading programs.

The fact that Reading First program elements are present to varying degrees in both the Reading First population and the non-Reading First population makes the use of an empirical sample of non-Reading First schools as a control group statistically invalid.

A Statistical Control Group

In the absence of a valid *sample* control group, it is in some cases possible to create a *statistical* control group using statistical techniques to hold constant the influence of school student demographics and, to

some extent, the use of common reading program components, in order to estimate the "true" impact of implementing the Reading First program. The procedure we use – a multiple regression model with level of implementation as one of the "predictor" variables – is described in detail in Chapter 4. The procedure allows us to estimate how a hypothetical set of schools that is assumed to have the same achievement starting point, demographic characteristics, and number of years in the program as a specified sample of actual Reading First schools, *would* have performed if they had *not* implemented the program. This hypothetical set of schools becomes our statistical control group. The performance "results" for this statistical control group are included in the tables and trend-lines presented in Chapter 4.

Note that this set of schools does not literally exist – it is a scenario run through a regression model. The validity of the statistical control group depends on the validity of the regression model and of its source data. The source data, in this case, consists of the set of Reading First schools (non-Reading First schools are excluded since they lack implementation information), the Reading First school student demographic characteristics, number of years in the Reading First program, and level of Reading First implementation. Level of implementation is measured using the Reading First Implementation Index (RFII), which is computed from surveys administered to teachers, coaches, and principals at each Reading First school. The computation of the RFII and its use in defining the statistical control group are described in detail in Chapter 3. Without a reliable RFII statistic, it would not have been possible to employ the statistical control group approach.

Demographic Information

The choice of a statistical control group rather than a sample control group makes some of the demographic information that follows obsolete since demographic effects are now controlled mathematically. Nonetheless, it provides a valuable profile of the similarities and differences across the three Reading First cohorts and all elementary schools in the state.

Most of the demographic data included in this and subsequent chapters are extracted from the Standardized Testing and Reporting (STAR) research file published on the California Department of Education (CDE) website[7]. In the STAR file, student-level data have been aggregated and presented at the school level. Therefore, the smallest unit of analysis in this study (other than waiver classrooms) is the school. Other sources of data include the California English Language Development Test (CELDT) research file, the Professional Assignment Information Form (PAIF) file, and the California Basic Educational Data System (CBEDS) file.

[7]The STAR research file used for the 2005-2006 data was the version obtained by EDS on October 4, 2006, referred to as "P2."

Table 2.1 lists the number of Reading First schools broken out by cohort and by year. It shows that 38 schools were added to Reading First in 2006, though the number of LEAs and cohorts is the same.

Table 2.1: Number of Reading First Schools (out of 5,834 elementary schools)

	Reading First Schools	
Cohort	2005 RF Schools	2006 RF Schools
Cohort 1	283	343
Cohort 2	391	372
Cohort 3	158	144
Cohorts 1, 2 and 3	821 (110 [1] LEAs)	859 (110 [1] LEAs)

[1] There were 108 LEAs that were school districts and two that were independent charter schools.

Socio-Economically Disadvantage (SED)

The second row in Table 2.2 compares Reading First school cohorts to non-Reading First schools with regard to socio-economic status (SED). We see that Cohort 1 Reading First schools started with the highest percentages (91.3% in 2003, 92.7% in 2004) of students with socio-economic disadvantage of all the school groups. Cohort 2 Reading First schools had lower percentages of students with SED (82.7% in 2004) than Cohort 1 when they began in the Reading First program, but the percentage increased to 83.5% in 2006. Cohort 3 is in the same range as Cohorts 1 and 2 with 85.8% of its students classified as SED in 2006. Thus, as of 2006, all three cohorts were similar in terms of SED.

Despite efforts to match them to Reading First schools, the demographically matched non-Reading First schools showed lower percentages of SED (about 82%) than the Reading First schools in 2004, 2005, and 2006. The Reading First schools were as high as 92.7% for Cohort 1. This raises a difficult matching problem. Even if a comparison group can be selected to match one of the Reading First cohorts, it cannot be selected to match all three at the same time due to cohort differences.

The RF Eligible schools are equally as dissimilar to the Reading First schools when compared on SED. Their average SED percent was 78.8% in 2004, 82.2% in 2005. When compared to the state average (53.4% in 2006), it is evident that the Reading First cohorts have a relatively high concentration of SED students, consistent with the federal guidelines that funds must go to schools with high levels of SED and academic underachievement.

These comparisons reveal some of the difficulties in matching a sample of non-Reading First schools to the Reading First schools, especially since each Reading First cohort has strong demographic differences. In combination with the difficulty posed by Reading First-like program elements compromising the non-Reading First schools, these demographic matching challenges further justify our decision not to include them as a control group in the Year 4 Report.

English Learners (EL)

In 2006 the demographic data show that on average 56% of the students in Reading First schools had limited English proficiency. Cohort 3 had the highest percentage of EL students (57.2%, compared to 56.5% for Cohort 1 and 54.9% for Cohort 2). The percent EL for All Elementary Schools was much lower than the Reading First cohorts, at only 29.3%. The demographically matched schools were very similar to the Reading First schools with regard to English Learners, with an average EL population of 56.1% in 2006. The RF Eligible school group, on the other hand, had much lower percentages of EL students, 47.4% in 2004, 49.7% in 2005, and 49.5% in 2006. This 8-10 percentage point difference may have a significant effect on achievement comparisons. Proficiency in English is highly correlated with success in academics and with achievement scores (August & Hakuta, 1997; National Research Council, 2002). Given the large difference in the percentage of EL students between the Reading First and RF Eligible schools, there is no justification for treating the two groups as comparable, which is why they were rejected in the Year 3 Report.

Table 2.2: Student Demographic Data, 2003 to 2006

	Reading First Schools									Demographically Matched Schools (used in Year 3			RF Eligible Schools (rejected in Year 3 Report)			All (inclu
	Cohort 1				Cohort 2			Cohort 3								
	2003	2004	2005	2006	2004	2005	2006	2005	2006	2004	2005	2006	2004	2005	2006	2003
Number of Schools	329	329	325	329	343	353	370	136	143	400	392	391	400	393	391	5823
SED (%)	91.3	92.7	89.1	91.4	82.7	86.7	83.5	85.1	85.8	82.9	82.1	82.3	78.8	82.2	82.0	51.0
EL (%)	58.5	58.6	58.8	56.5	53.0	55.5	54.9	57.5	57.2	57.0	57.5	56.1	47.4	49.7	49.5	27.2
Students with Disabilities (%)	7.5	8.4	8.5	8.4	8.0	7.7	7.9	7.1	7.7	9.4	8.5	9.3	7.9	7.8	8.4	9.8
African American (%)	17.2	16.6	15.4	14.1	8.8	8.2	8.0	6.6	6.4	5.2	5.1	5.0	6.7	6.4	6.0	7.8
American Indian (%)	0.3	0.3	0.3	0.3	1.0	0.9	0.8	0.8	0.6	0.8	0.9	0.9	1.4	1.6	1.8	1.3
Asian (%)	4.0	3.6	3.8	3.2	4.6	4.1	3.9	1.1	1.0	7.1	7.0	6.8	3.9	3.9	3.8	7.3
Filipino (%)	1.0	0.9	1.0	1.0	1.7	1.6	1.5	1.3	1.0	1.5	1.6	1.6	1.2	1.2	1.2	2.2
Hispanic (%)	71.5	73.3	74.4	76.5	72.0	74.1	75.2	77.1	78.6	71.5	71.6	72.5	67.9	69.4	70.8	40.2
Pacific Islander (%)	0.5	0.5	0.4	0.4	0.8	0.8	0.8	0.5	0.5	0.5	0.6	0.6	0.4	0.5	0.4	0.6
White (%)	3.7	3.4	3.2	2.8	9.6	8.8	8.0	11.2	10.3	12.1	11.6	11.1	17.2	15.8	14.7	36.5

[1]Cohort 2 demographics are provided beginning in 2004 because 2003-2004 was the first year of Reading First implementation in those schools.
[2]Cohort 3 demographics are provided beginning in 2005 because 2004-2005 was the first year of Reading First implementation in those schools.
[3]Demographics for the Comparison Group Schools and the RF Eligible Schools are provided even though they are not used in the Year 4 Report. Review of the EL (%) row reveals, for instance, that RF Eligible Schools have a substantially lower proportion of English Learners.
[4]The group "All Elementary Schools" *includes* Reading First schools in this chapter. Whereas in Chapter 4, "All Elementary Schools" *excludes* Reading First schools.

Urban-Rural Distribution

Table 2.3 presents the urban-rural locations of Reading First school districts in California. This information tells us that most of the school districts in Cohort 1 are located primarily in large or mid-size city areas (10 districts, 77%). Cohort 2 school districts are fairly evenly distributed across the large and mid-size cities and the fringe of large or mid-size cities (51 districts, 85%). Cohort 3 is most dissimilar to Cohort 1 as there are only 2 school districts (5.4%) in large city areas. The majority of school districts in Cohort 3 are in the urban fringe of cities and in rural areas (33 districts, 89%). A "typical" Reading First school district in California tends to be located in cities or suburbs. Across all three cohorts, only 13.6% of the districts are in rural areas.

Examining both the ethnic percentages in Table 2.2 and the locations of school districts in Table 2.3, it is interesting to note that Cohort 1 Reading First school districts are located in mostly urban areas and also have the largest percentage of African American students of the three cohorts. Cohort 2 has a more even distribution of schools across large and mid-size cities as well as urban and mid-size suburban areas than do Cohorts 1 and 3, and a more even distribution of students across the major ethnic groups. Cohort 3 has many school districts located outside cities and in rural areas and also has the largest percentage of Hispanic students.

Table 2.3: Urban-Rural Distribution for Reading First Districts, National Center for Education and Statistics 2005

District Location	Cohort 1		Cohort 2		Cohort 3		All Cohorts	
	N of Districts	Percent of Districts[1]	N of Districts	Percent of Districts	N of Districts	Percent of Districts	N of Districts	Percent of Districts
Large City	6	46.2	10	16.7	3	8.1	19	17.3
Mid-size City	4	30.8	11	18.3	7	18.9	22	20.0
Urban Fringe of Large City	1	7.7	17	28.3	10	27.0	28	25.5
Urban Fringe of Mid-size City	1	7.7	14	23.3	9	24.3	24	21.8
Small Town	0	0.0	1	1.7	1	2.7	2	1.8
Rural	1	7.7	7	11.7	7	18.9	15	13.6
Total	13	100.0	60	100.0	37	100.0	110 [2]	100.0

[1] The percent of the districts in that cohort in a particular type of location.
[2] There were 108 LEAs that were school districts and two that were independent charter schools.

Waiver Classrooms

California law (Proposition 227) mandates instruction in English for all students unless parents sign a waiver requesting bilingual instruction for their child. Classrooms with students whose parents have signed waivers for the students to receive primary language instruction in Spanish are referred to in this report as "waiver classrooms." In October 2003, the state legislature passed AB 1485, effective January 1, 2004, stating that students receiving instruction in their primary language (in waiver classrooms) would be held to the same standards as students receiving instruction in English, and that they would be required to participate in the STAR test in English. This led to the Reading First program's full support for such classrooms in Reading First schools. State-approved research-based reading materials in Spanish were available in January 2004.

Table 2.4 presents the number of waiver classrooms added to the Reading First program by cohort and year as of 2005. The 2006 statistics were not collected. As of 2005 there were 1734 waiver classrooms in Reading First schools across 51 districts. Cohort 1 has a lower percentage of waiver classrooms (23.8%) than Cohorts 2 and 3 (40.1% and 36.2%, respectively). Cohort 2 has the highest percentage of waiver classrooms of the three cohorts.

What is notable about this data is that a large number of Reading First school districts, 51 of 110 (46.4%), serve students receiving instruction in Spanish.

Table 2.4: Number of Waiver Classrooms Added to Reading First Districts by Cohort, 2003-2004 and 2004-2005 (2005-2006 was not collected)

	Cohort 1				Cohort 2				Cohort 3				Total		
		Waiver Classrooms				Waiver Classrooms				Waiver Classrooms				Waiver Classrooms	
	N of Districts	N	Percent by Cohort[1]	Percent by Year[2]	N of Districts	N	Percent by Cohort	Percent by Year	N of Districts	N	Percent by Cohort	Percent by Year	N of Districts	N	Percent by Year[3]
2003-2004	8	276	67.0	40.7	15	402	57.8	59.3	0	0	0.0	0.0	23	678	39.1
2004-2005	1	136	33.0	12.9	10	293	42.2	27.7	17	627	100.0	59.4	28	1056	60.9
Total	9	412	100.0	23.8	25	695	100.0	40.1	17	627	100.0	36.2	51	1734	100.0

[1] Percent by Cohort provides the percentage of Waiver Classrooms in the respective cohort added each year. For example, 67.0% of the total number of Waiver Classrooms in Cohort 1 (N=412) were added in 2003-2004. Similarly, 33.0% of the 412 Waiver Classrooms in Cohort 1 were classified as such in 2004-2005.

[2] Percent by Year provides the percentage of Waiver Classrooms added for the respective academic year. For example, 40.7% means that of the total Waiver Classrooms in 2003-2004 (N=678), 40.7% were in Cohort 1. Similarly, out of the total number of Waiver Classrooms in 2004-2005 (N=1056), 12.9% were in Cohort 1 (27.7% were in Cohort 2 and 59.4% were in Cohort 3).

[3] Under the Total column, the percentages across all cohorts are provided for the two academic years. 39.1% implies that of the total number of Waiver Classrooms (N=1734), 39.1% were added in 2003-2004 and 60.9% were added in 2004-2005.

Teacher Qualifications

Table 2.5 provides teacher credential and experience information as available in the CBEDS teacher Professional Assignment Information Form (PAIF) research files. This database shows the percent of teachers falling into each educational degree category by cohort and year. A comparison of the Reading First cohorts in terms of teacher qualifications shows that, in general, the teachers in Cohort 1 had lower educational degrees than teachers in Cohorts 2 and 3. Cohort 1 had fewer teachers with higher levels of education (Bachelors plus 30 or more semester units, Masters, Masters plus 30 or more semester units, and Ph.D.s) than the other cohorts. To more easily compare cohorts with each other and to the non-Reading First groups, a weighted index was computed based on CBEDS data sources relative to teacher qualifications. The weighted teacher qualification is an index ranging from a low teacher qualification of 1 to a high teacher qualification of 5. Refer to the note under Table 2.5 for a description of how this index was computed. This index shows that Cohort 1 Reading First schools had the lowest Weighted Teacher Qualification (2.0 to 2.2) of the other Reading First cohorts (ranging from 2.2 to 2.3) and of the non-Reading First schools. Of the Reading First cohorts, Cohort 2 had the highest percent of teachers with higher education.

It is interesting to note that in Table 2.3, the Urban-Rural Distribution of Reading First schools, Cohort 1 had the highest concentration of urban schools, and in Table 2.5 we see that Cohort 1 also had the lowest number of teachers with advanced degrees and full credentials. This is consistent with national reports indicating that urban schools tend to have a greater number of teachers with lower levels of qualifications, including more teachers with less than full credentials and fewer teachers with advanced degrees (Darling-Hammond, 1999; National Commission on Teaching and America's Future, 1996).

There are several indications of positive change in teacher qualifications in Reading First schools in Table 2.5. We see an increase over time in the percents of teachers holding a Bachelors degree plus 30 semester hours. In contrast, the Comparison Group schools and the RF Eligible schools show a slight decrease, and the All Elementary schools show a slight decrease across time. The number of fully credentialed teachers also increased over time in Cohorts 1 and 2, though increases also occurred in the Comparison Group schools, the RF Eligible schools and in the All Elementary schools groups.

Though the statewide increase in credentialed teachers may be a result of the pressure on districts to employ "highly qualified" teachers as a result of the *No Child Left Behind Act*, it is possible that participation in Reading First can explain the increases in the "Bachelors plus 30" category for Reading First schools. Reading First has provided opportunities for teachers to earn course credit for some aspects of advanced training. This is an important factor that will be monitored in future years.

Table 2.5: Elementary Teacher Credential and Experience 2003 to 2006

	Reading First Schools									Comparison Group Schools (used in Year 3 Report)[3]			RF Eligible Schools (rejected in Year 3 Report)[3]			All Elementary Schools (included in Year 4 report for context)[5]			
	Cohort 1				Cohort 2[1]			Cohort 3[2]											
	2003	2004	2005	2006	2004	2005	2006	2005	2006	2004	2005	2006	2004	2005	2006	2003	2004	2005	2006
Number of Schools	329	329	325	329	343	344	370	135	143	400	388	388	400	393	391	5647	5694	5720	5837
PhDs (%)	0.55	0.74	0.75	0.78	0.62	0.66	0.65	0.59	0.51	0.90	0.60	0.53	0.90	0.60	0.38	0.90	0.80	0.80	0.71
Masters plus 30 or more semester units (%)	9.4	11.7	12.6	14.2	13.9	13.6	13.4	16.3	14.3	11.7	11.2	11.3	12.3	11.8	11.8	14.0	14.5	14.3	14.1
Masters (%)	11.0	11.8	12.2	12.4	16.4	18.5	20.0	16.6	13.6	18.3	20.0	20.4	17.2	19.0	18.6	15.5	16.9	18.1	19.0
Bachelors plus 30 or more semester units (%)	41.2	44.2	45.3	45.7	49.9	49.6	49.0	47.1	53.0	52.8	51.3	50.9	53.1	51.8	52.7	51.3	51.7	50.6	49.8
Total Advanced Degrees (%)	**62.1**	**68.4**	**70.9**	**73.1**	**80.8**	**82.4**	**83.1**	**80.6**	**81.4**	**83.7**	**83.1**	**83.1**	**83.5**	**83.2**	**83.5**	**81.7**	**83.9**	**83.8**	**83.5**
Bachelors (%)	35.1	30.9	28.4	26.4	19.1	16.9	16.8	19.3	18.3	16.2	16.8	16.6	16.2	16.7	16.3	16.4	15.8	15.9	16.2
Less than Bachelors (%)	0.7	0.6	0.8	0.5	0.1	0.8	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.4	0.2
Total Bachelors or less (%)	**35.8**	**31.5**	**29.2**	**26.9**	**19.2**	**17.7**	**16.9**	**19.5**	**18.5**	**16.4**	**17.0**	**16.8**	**16.4**	**16.9**	**16.5**	**16.6**	**16.0**	**16.3**	**16.4**
Fully Credentialed Teachers (%)	77.8	82.1	91.3	94.9	93.7	96.0	97.2	92.1	93.8	93.4	95.1	95.8	93.6	95.4	96.3	90.9	93.7	95.8	96.5
Weighted Teacher Qualification[4]	2.0	2.0	2.1	2.2	2.3	2.3	2.3	2.3	2.2	2.3	2.3	2.3	2.3	2.3	2.3	2.2	2.3	2.3	2.3
Average years teaching	10.8	10.9	11.3	11.5	11.3	11.6	11.9	11.4	11.9	11.7	11.8	11.9	11.8	12.0	12.1	12.7	12.8	12.8	12.9

[1] Cohort 2 demographics are provided beginning in 2004 because 2003-2004 was the first year of Reading First Implementation in those schools.
[2] Cohort 3 demographics are provided beginning in 2005 because 2004-2005 was the first year of Reading First Implementation in those schools.
[3] Demographics for the Comparison Group Schools and the RF Eligible Schools are presented even though they are not used as "control groups" in the Year 4 Report. Because they were formed in the second year of the study, they start only in 2004.
[4]The Weighted Teacher Qualification is computed as follows: The percentage of teachers with PhDs is given a weight of 5, the percent of teachers with Masters plus 30 or more semester units is given a weight of 4, the percent of teachers with Masters is given a weight of 3, the percent of teachers with Bachelors plus 30 or more semester units is given a weight of 2 and teachers with Bachelors are given a weight of 1. The weighted degree percents are summed then divided by 100 to reach the Weighted Teacher Qualification. This index spans from 1 (lowest qualification) to 5 (highest qualification).
[5]The group "All Elementary Schools" *includes* Reading First schools in this chapter. Whereas in Chapter 4, "All Elementary Schools" *excludes* Reading First schools

Non-Reading First Schools

As discussed in this chapter, neither the Reading First Eligible schools nor demographically matched non-Reading First schools are now used as "control groups" in the analysis of Reading First schools, having been replaced by a statistical control group. The objective in examining demographic information in this chapter is not only to understand the nature of the Reading First schools, but also to justify the use of a specific set of schools for purposes of comparisons in outcome and impact analysis. In general, control groups are expected to fill a role similar to a placebo in clinical studies. They are intended to match the treatment group in all respects but the treatment itself. In educational research, while it is generally impossible to create a true control group, it is often possible to create a group that controls for the presence of "predictor" variables that have been found to impact outcomes. The traditional way to control for predictor variables is to identify a group of schools (or students) that most closely resembles the treatment group, the Reading First schools. Sometimes, however, it may be preferable to control for demographic variations using multiple regression, which is what we have done in 2006.

Nonetheless, in 2006 we continue to provide information regarding the two groups that were used as comparison groups in 2004 and 2005. That which was called the Comparison Group B in 2005—here referred to as the demographically matched non-Reading First schools—best filled the role of a control group in a demographic sense in 2005, although even demographically it had problems matching all three cohorts simultaneously. This group of schools was selected from the population of elementary schools using a two-step process that involved k-means Cluster Analysis and random selection. The two demographic variables controlled were percentages of SED and EL students per school. That which was called the Reading First Eligible group of schools in previous reports was randomly selected from the list of schools that were eligible for funding but resided in school districts that had not received any Reading First funding. Had Reading First funding been randomly assigned to the list of eligible schools, it is possible that the Reading First Eligible schools could have served as comparison schools for the Reading First cohorts. However, the demographic evidence presented in this chapter shows that funding in the first three years of the program went to schools with higher percentages of SED and EL students. As a result, the eligible schools that had not received Reading First funding as of 2005 appear to have been quite different from the funded schools in terms of demographic characteristics. Therefore, from the perspective of research design, it was not advisable to compare Reading First schools to the Reading First Eligible group of schools.

That said, neither the Reading First Eligible group or the demographically comparable non-Reading First group can be considered valid for purposes of this study due to their use of similar program elements as those required by Reading First. They are reported for informational purposes only.

Conclusions

This chapter yields the following findings:

- The term "Cohorts" refers to the year a Reading First LEA (district) accepted funding. The term "Years in Program" refers to the number of years a school within an LEA cohort has actually been implementing the program.
- For the Year 4 study, a "statistical control group" has been defined to serve as a "control group" in preference to the samples of non-Reading First schools called in the Year 3 Report the Comparison Group and the RF Eligible schools.
- The main reason for defining a "statistical control group" is that non-Reading First schools have to an unknown extent been incorporating the same program elements that are required in Reading First schools, compromising the control group.
- The Reading First cohorts have inherent differences in their demographic compositions. Cohort 1 has the highest percentages of SED and African American students and a high percentage of EL and Hispanic students. Cohort 3 currently has the highest percentages of EL and Hispanic students. The differences between cohorts make the selection of a single non-Reading First sampled comparison group very problematic. It is not possible for one control sample to be matched to all three cohorts at once.
- Cohort 1 schools largely represent urban city districts. The percentage of rural districts was relatively small in all Reading First cohorts, but when compared among cohorts, Cohort 3 had the highest concentration of rural districts.
- Waiver classrooms (Spanish language instruction) were fewest among Cohort 1 districts and more prevalent among Cohorts 2 and 3.
- Due to pronounced differences in demographics between Reading First cohorts, it is important to study outcome measures by cohort and not combine all three cohorts together.
- Reading First schools showed increases in teachers with advanced education, a trend that was not evident in non-Reading First groups. There was an increase statewide in the number of fully credentialed teachers and this trend was also evident in Reading First schools. Cohort 1 schools had the lowest percentages of well-qualified teachers of Reading First schools.

References

August, D., & Hakuta, K. (1997). *Improving Schooling for Language-Minority Children*. Washington, DC: National Academy Press.

Darling-Hammond, L. (1999). *Teacher Quality and Student Achievement: A Review of State Policy Evidence*. University of Washington Center for the Study of Teaching and Policy. Retrieved December 18, 2003 from http://depts.washington.edu/ctpmail/PDFs/LDH_1999.pdf

National Commission on Teaching and America's Future (NCTAF). (1996). *What Matters Most: Teaching for America's Future*. New York: Author.

National Research Council (2002). *Minority Students in Special and Gifted Education*. Committee on Minority Representation in Special Education, M. Suzanne Donovan and Christopher T. Cross, editors. Division of Behavioral and Social Sciences and Education. Washington, D.C., National Academy Press.

Chapter 3: Measurement of School Implementation

This chapter explains how survey data have been used to address the question, *How well has the Reading First program been implemented in each participating school and district?* It also presents measures of program implementation dimensions such as professional development, material and instructional resources, understanding of Reading First Assurances and curricular materials, and perceptions of the Reading First program.

To evaluate the implementation of Reading First in California, it is necessary to construct implementation measures to gather data from the field specifically for this purpose; no pre-existing data source is currently available. For this effort, Educational Data Systems (EDS) developed three surveys (paper-and-pencil in 2004, online and paper-and-pencil in 2005 and 2006) – one each for Reading First teachers, coaches, and principals – and administered them in the spring during 2004, 2005, and 2006. In 2004, approximately 14,000 surveys were returned to yield a response rate between 73% and 82%. The spring 2005 administration yielded approximately 20,200 surveys. The spring 2006 administration yielded 19,957 surveys. The response rates for 2005 and 2006 are not known precisely due to uncertainty regarding the total number of possible respondents at Reading First schools. However, we estimate it to be between 85% and 93%. This chapter discusses the organization, scoring, and analysis of the survey data to compute a Reading First Implementation Index (RFII) for each school. It presents the role of such information in assessing the efficacy of the Reading First program.

Further technical details are provided in Appendix E. Information on program implementation was also collected using focus groups of principals and district administrators from schools and LEAs selected according to their implementation and achievement levels, discussed in Chapter 5.

Rationale for Measuring Implementation

To fully evaluate the effectiveness of an educational program, it is not enough to look just at school-level achievement gains. Rather, two pieces of information are required: a) the effect of the program (e.g., achievement gains); and b) the *presence* or *absence* or *degree of implementation* of the program in question. If it is found that duration and intensity of program implementation are significant predictors of achievement, then we can say that evidence exists that the program is effective. If achievement gains bear no relation to the degree of program implementation, no evidence of program efficacy can be claimed (Schiller, 2001).

Unfortunately, there is no easy or obvious way to directly measure the presence, absence, or degree of implementation of Reading First in participating schools and districts (i.e., the degree to which Reading

First funding is being applied in the prescribed manner by school officials and teachers). There is no statewide database that would definitively reflect Reading First implementation.

Some of the analysis in Chapter 4 involves comparing the achievement gains of Reading First schools to non-Reading First schools. However, as Chapters 2 and 4 explain, such comparisons allow only limited conclusions because little is known about non-Reading First schools. Suppose comparable non-Reading First schools receive funding from similar programs, employ equivalent instructional materials, and demonstrate that they match or outperform Reading First schools. It would be unfair to claim that this constitutes evidence that the Reading First program is not working. The "control" group is in effect receiving the same treatment as the "treatment" group. Because all that differs is the funding source, no light is shed on the efficacy of the underlying pedagogy and educational philosophy.

Another approach is to ignore the non-Reading First schools and look only at absolute achievement gains. Suppose the gain is large. While such a gain would be consistent with the hypothesis that Reading First is effective, it does not rule out the possibility that the gain could have been caused by some other factor, a change in test forms for instance, or some other state-wide initiative that is not Reading First. We do not know how the schools *would* have done in the absence of the program.

These ambiguities justify the inclusion of a third approach, an attempt to quantify the actual degree of implementation occurring within each Reading First school. In this approach, a school that does not use the recommended materials, neglects professional development, or skimps on instructional time is not considered to be implementing the program, no matter how much Reading First funding it receives. When "implementation" is defined in this more tangible way, assuming it can be measured with reasonable accuracy, it becomes feasible to decide whether the program has the *potential* of working if it is well implemented, a determination that is critical in deciding whether the program should continue to be funded. Suppose we find that schools that rigorously and faithfully use the recommended program materials, and whose teachers, coaches, and principals receive the requisite training and understand the Reading First philosophy, do not show any higher achievement gains than schools that do not. If the achievement and implementation measures are sound, such a finding would constitute evidence that the program is unlikely to be effective no matter how much funding it receives.

By the same token, if we find that higher implementation corresponds to higher achievement gains, the central policy question becomes, "How can we ensure that Reading First is systematically implemented in all Reading First schools?"

An Implementation Measure Makes it Possible to Create a Statistical Control Group

An evaluation, whether of pharmaceuticals or educational programs, typically requires that a "treatment" be studied using two groups, one receiving the treatment, the other not, called a "control group." In order

for the comparison to rule out extraneous factors, the treatment group and the control group have to be the same in all respects except for the amount of "treatment" (program implementation) they are receiving.

In an experimental research design, schools would be randomly assigned to the treatment group and the control group. The "treatment" (Reading First program implementation) would be administered to the treatment group while the control group would receive reading curricula that are designed specifically to be different from those provided by the Reading First program. Randomization ensures that the groups are statistically equivalent. Control of the treatment ensures that any observed difference in outcomes between the two groups is due to the treatment, not to some other, unmeasured factor. Unfortunately, such a design is not possible or ethical in the real world of education.

The traditional alternative is to let schools assign themselves to the treatment group and compare them to non-treatment schools that are demographically similar. This approach was employed in the Year 1, Year 2, and Year 3 evaluation studies. Unfortunately, this approach requires information about non-treatment schools that is not available, and which funding and time constraints prevent us from collecting. We are indeed able to collect demographic information about non-Reading First schools, but we are unable to collect information about their reading curricula, amount of professional development, or extent of coaching support to ensure that they truly differ from Reading First schools in their program components. We have a treatment group but no valid control group.

A third, mathematical alternative becomes feasible as a result of our ability to construct a reliable implementation measure for each Reading First school. By quantifying how implementation relates to achievement while *controlling* for demographic variables using multiple regression, we can extrapolate that relationship to predict how a typical Reading First school would do *if they had not* implemented the program. Instead of a *sampled* control group, we are in effect creating a *statistical* control group of hypothetical schools that do not actually exist, but which are the same as the "typical" Reading First school except that they do not implement the program. While a traditional non-Reading First control group would be preferred if it could be controlled properly, the statistical control group is a close second best, and even offers some theoretical advantages.[12] However, this approach only becomes feasible when schools can be assigned a reliable school implementation measure that can be mathematically related to their achievement measures.

[12] Aside from the logistical advantage of not having to administer surveys to non-Reading First schools, statistical controls offer a mathematical means to create a theoretical control group that is the same as the treatment group in all respects except for the treatment. The methodology is valid to the degree the assumptions behind multiple regression are met and that the demographic predictor variables provide sufficient information about the schools to model them reliably. In research situations where it is not possible to assign schools randomly to Reading First and non-Reading First groups, a statistical control group can provide an alternative way to estimate the treatment effect.

That is why we attempt to construct a reliable implementation measure for each Reading First school. Without such information, it is virtually impossible to measure the impact of Reading First.

Rationale for Using a Survey

In 2006 there were over 850 Reading First schools in California. To measure implementation in each school, the external evaluator would ideally send trained auditors to observe each Reading First classroom over an extended period of time. While this would not be practical for the complete population of schools, it could in theory be done with a representative sample of schools (absent legal restrictions). The resulting data would address the question of program efficacy. However, the State specifically requested in its Request for Proposals an implementation measure for *each* Reading First school. The scope of this evaluation does not allow for classroom observation research designs for measuring implementation, though we have employed focus group interviews as described in Chapter 5.

The alternative is to administer a survey. This was done in the springs of 2004, 2005, and 2006. The advantage of using a survey is that it is easy to administer and analyze and the respondents (teachers, coaches, principals) are the most knowledgeable regarding what is happening inside their schools and classrooms throughout the school year. Nonetheless, there are unavoidable limitations and sources of bias:

1. The respondents are, to a certain extent, reporting on themselves. This could lead to a substantial upward bias in estimations of school implementation since respondents may feel a desire to respond "appropriately," or they may be unclear regarding what "full" implementation looks like.
2. Similarly, if school officials believe that survey results could be used to reduce or deny funding, there would be a strong incentive for some schools to encourage respondents to respond in a way that would raise the school's implementation score, also leading to an upward bias.
3. While an upward bias would probably apply to all schools to some degree, it might be more pronounced in some schools than others. This would introduce an extra source of error in the *relative* measures of schools.
4. In order for a survey to be specific enough to be useful, it needs to tailor its questions to particular types of respondents. For instance, there need to be questions tailored specifically to teachers, coaches, and principals, and to users of Open Court and Houghton Mifflin in the Spanish and English versions. This would seem to impair our ability to compare schools when they have different proportions of each respondent type.
5. To the degree the survey instrument is changed from year to year, results could lose their cross-year comparability.

6. Each question, taken on its own, inevitably carries ambiguities and imprecision. It is often difficult to be clear exactly what dimensional construct is being measured by a question, and whether it is indeed "implementation."

These are legitimate issues and we employed accepted survey analysis models to ameliorate the effects of each issue where possible. Issues 1 and 2 (upward reporting bias) are addressed by measuring schools relative to each other, not in absolute terms. Thus, to the degree all schools display the same upward bias, this bias has no effect when assessing program efficacy. Issue 3 (intentionally reporting higher scores), which causes differential bias, is the most difficult to solve. We address it in part by making it possible for teachers, who are presumably less influenced by funding concerns than principals, to take the survey anonymously. We are careful to include questions whose "correct" answers are not immediately obvious without having taken Reading First training. We also provide numerous opportunities for respondents to rate other respondent types. Coaches rate teachers. Teachers rate coaches. Both rate principals. This issue is most effectively addressed at the policy level by refraining from using the Reading First Implementation Index (RFII) as a criterion for determining "significant progress." The "significant progress" regulations approved in Fall 2006 are based entirely on achievement data.

Equating methods are used to address Issues 4 (tailored questions) and 5 (cross-year comparability) and to remove certain types of biases (e.g., principals being more "lenient" than teachers). Issue 6 (ambiguities in specific questions) is addressed by reporting measures composed of groups of questions that have been specifically written, then selected on statistical grounds, to "hang together" in coherent dimensions as defined by experts in the California Technical Assistance Center (C-TAC) and the Evaluation Advisory Group (EAG).

The statistical reliability (Cronbach-alpha) of the Reading First Implementation Index for 2006 is 0.92 (0.90 in 2005). A reliability of 0.85 is widely considered "sufficient" in the field of educational measurement. The cross-year correlation of the RFII across schools between 2004 and 2005 was 0.42. Between 2005 and 2006, it was 0.58. Taking into account changes in respondent samples per school and likely changes in implementation practices from year to year, these correlations support the claim that the RFII is measuring an objective school-level characteristic. Otherwise, we would expect the cross-year correlations to be closer to zero.

It is sometimes stated that survey instruments are "too subjective" to be used for measurement purposes despite their widespread use for measuring political beliefs, marketing preferences, attitudes and practices. Given the high content validity of the Reading First survey and its level of detail, the use of methodological tools that correct for common sources of bias, and the statistical and psychometric characteristics of the RFII, we feel that the RFII is approximately as valid and reliable as a typical large-

scale student assessment, bearing in mind that it is more vulnerable to intentional efforts to produce a high score.

Changes to the Survey in 2005 and 2006

In 2005, based on a change in the Reading First program to include Spanish curricular materials for waiver classrooms (instruction in Spanish), Section C of the teacher survey was expanded to include 74 additional questions involving the receipt and use of materials from *Foro abierto para la lectura* (the Spanish version of Open Court) and Houghton Mifflin's *Lectura: Herencia y futuro*. Additionally, the options in Question 6 of Section A were expanded to allow teachers to report whether they are teaching the Spanish version of their district's adopted reading program. (See Appendix B for the survey questions.)

The other major change was the removal of Section H from the teacher survey. The remaining changes to the survey were minor editorial corrections and clarifications of the instructions to respondents. Although the addition of the Spanish curricular questions made the survey appear larger, the actual number of questions that any single respondent faced decreased between 2004 and 2005 due to the removal of Section H.

In 2006, Section C went through further revisions to make the references to curricular materials more generic and less likely to result in missing data due to non-response resulting from an inexact match between the respondent's materials and those listed in the questionnaire. Some of the changes were substantive enough to be classified as new questions. Efforts were made to ensure that there were enough "old questions" to link the different survey administrations together. Individual questions throughout the survey underwent editorial modifications, often to clarify routing from section to section on the web survey. Anecdotal information received from teachers and coaches indicates that it took 20 to 30 minutes to complete the survey.

Calculating the Reading First Implementation Index (RFII)

Appendix E describes in detail the steps by which the RFII was constructed and how it is calculated. The procedure is as follows:

1. Using an Item Response Theory program called Facets, subsets of questions across the three surveys are used to generate measures on 17-19 dimensions.[13]
2. Three of these dimensions are used to calculate each school's RFII. They are: School Implementation Overall (SIO), Overall Reading First Understanding (OUND), and Teacher/Coach Professional Development (TCPD).
3. The measures on these dimensions are weighted and combined to calculate the school's RFII. The weights are:

   School Implementation Overall (SIO) = 70%

   Overall Reading First Understanding (OUND) = 20%

   Teacher/Coach Professional Development (TCPD) = 10%
4. The resulting RFII statistic is scaled to be between 0 and 100 and to have a distribution similar to that of the Reading First Achievement Index (RFAI).

Distribution and Interpretation of the RFII

Figure 3.2.1 shows how the RFII was distributed across all Reading First schools in 2006 (the 2005-2006 school year). The mean RFII was 39; the standard deviation around the mean was 6. As explained in Appendix E, this can be interpreted as follows: Reading First *teachers* on average found their schools to be "more than adequate" 39% of the time (i.e., on 39% of the relevant items). Note the emphasis on teachers; the RFII was intentionally calibrated relative to teacher perceptions of "adequate implementation." Teachers tended to give lower scores to their schools than did coaches and principals. While most of the dimension measures in the tables starting with Table 3.2.0 in the next section are calibrated relative to teachers, some of the dimensions are calibrated relative to coaches and principals as indicated in the footnotes to Tables 3.2.0 and 3.2.1 in the next section.

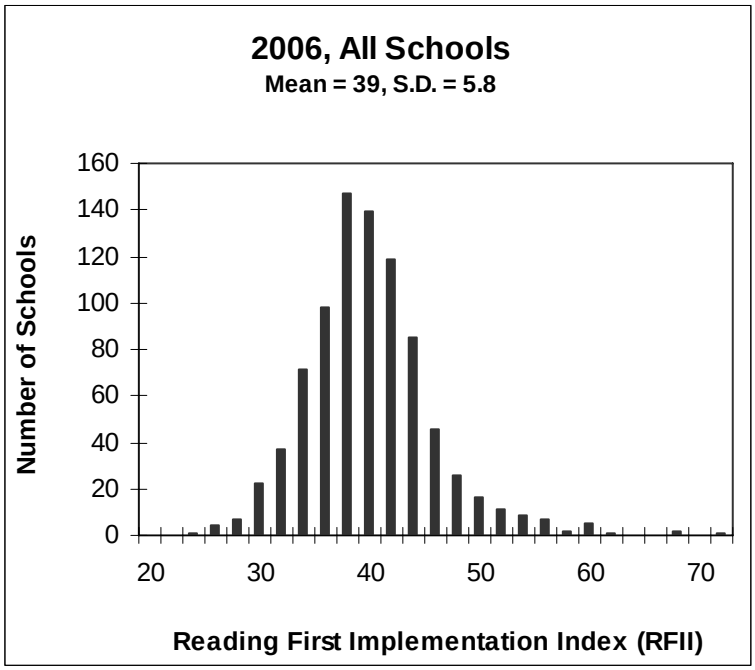
[13] There are a number of methods for analyzing survey data. The method used here, the Many-Facet Rasch Model or Facets, is well-suited to judging and equating designs in which there are large amounts of missing data and the data consist of "subjective judgments" (Linacre, 1994). Facets is a generalization of the Rasch Model, which is one of a number of psychometric models organized under the rubric of "Item Response Theory." These are the models behind many large-scale student assessments and licensure examinations, chosen especially for their ability to equate test forms so that students who are exposed to different test forms can nonetheless be measured accurately on a common scale.

Interpreting the RFII as a percentage of items is not strictly correct. The RFII is actually based on a probability that teachers in a school will rate their school "more than adequate" across the survey items. It is a theoretical statistical parameter used to explain the data, not a literal count of responses. Interpreting it as a percentage of items scored "more than adequate" makes it easier to understand, however.

Measures on the far right tail of the distributions of the figures below (above 55) should be viewed with caution; such schools tend to show unusually high inter-respondent agreement. This raises the possibility that respondents were coached to answer the questions in a specific way or that they completed the survey as a collaborative group.

Figures 3.2.2 through 3.2.4 are histograms that break out the school implementation distribution by Years in Program (YIP). All three groups have similar distributions. (YIP 1 schools are too few to be reported.) This is noteworthy because YIP 2, the schools that have been in the program for two years, had a notably bimodal distribution in 2005[14] when they had been in the program one year. An additional year of implementation has allowed the YIP 2 schools that formed the lower mode in the RFII continuum to catch up with the schools in the upper mode. This is accompanied by a 5-point average RFII gain in the YIP 2 schools, as shown in Table 3.5.0 in the next section, reinforcing the Year 3 finding that Reading First schools need at least one full year to become acclimated to the program.

Figure 3.2.1: All Schools - 2006 Reading First Implementation Index (RFII), distribution of schools



[14] The Year 3 Report can be accessed on the Educational Data Systems website at http://www.eddata.com/resources/publications/.

Figure 3.2.2: YIP = 4 – 2006 Reading First Implementation Index (RFII), distribution of schools

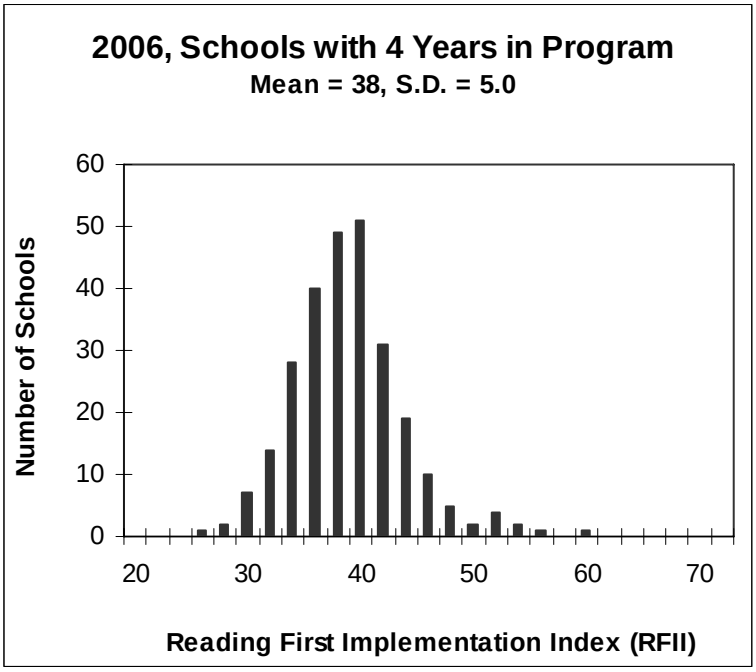


Figure 3.2.3: YIP = 3 – 2006 Reading First Implementation Index (RFII), distribution of schools

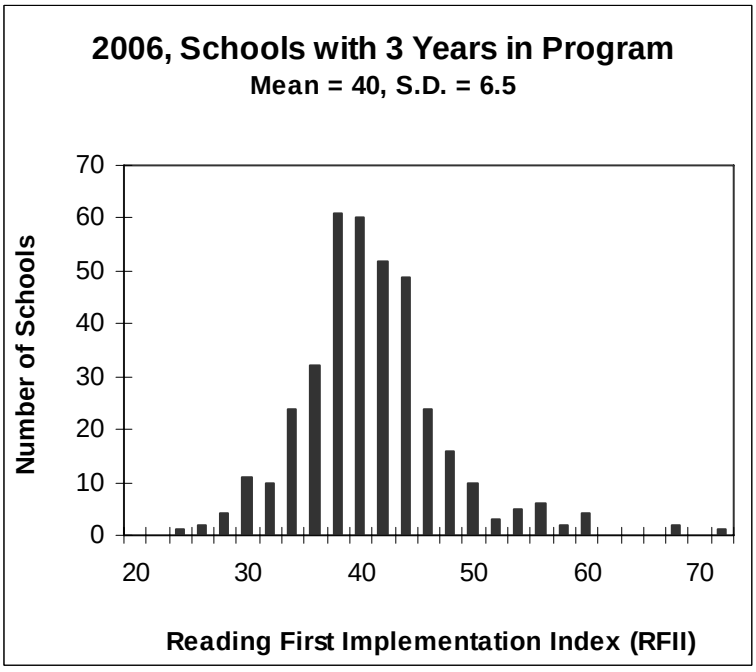
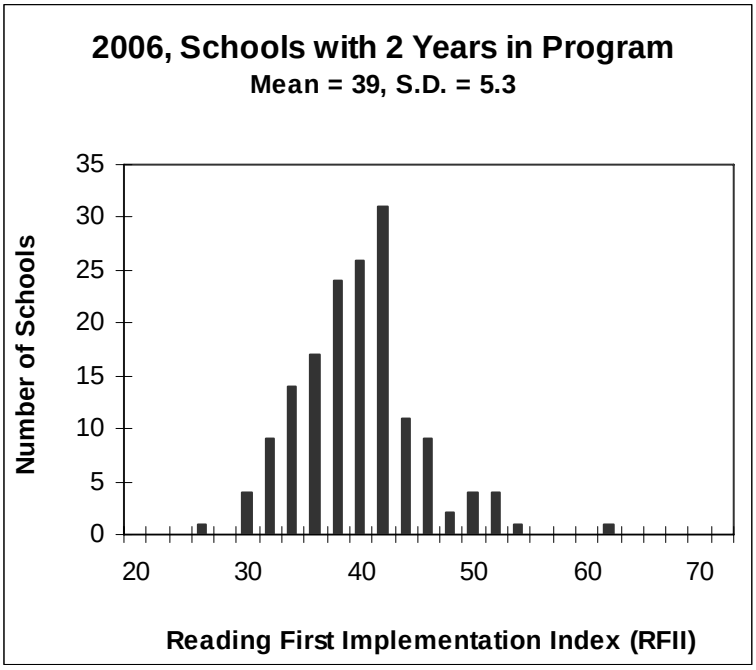


Figure 3.2.4: YIP = 2 – 2006 Reading First Implementation Index (RFII), distribution of schools



Tables of Measures on Each Dimension

The right half of the bottom row of Table 3.2.0 provides mean school RFII's for 2004, 2005, and 2006 for all schools in the Reading First population, with standard deviations. Means and standard deviations for 18 other dimensions are also reported. Three of these dimensions, set in bold, were used to calculate the RFII. The remaining dimensions were extracted from targeted subsets of items, some of whose items were used in calculating the RFII. They provide a more detailed profile of perceptions and implementation of the Reading First program. The statistics in the right six columns, excluding the standard deviations, may be interpreted as the percent of times (items) that teachers rated their school "more than adequate" on that dimension, averaged across schools. This is the same standard used for the RFII.

The six columns on the left of Table 3.2.0, excluding standard deviations, apply a more lenient standard of implementation to the population of Reading First schools. These statistics may be interpreted as the percent of times (items) that teachers rated their school "adequate" or better on that dimension, averaged across schools.

Included in the third column is the number of items per dimension. The more items, the smaller the standard deviation tends to be and the more reliable the measure is. Dimensions with only a few items have larger standard deviations and are much less reliable.

Table 3.2.1 uses the information in Table 3.2.0 to estimate the percentage of Reading First schools in California that meet the "adequate" or better criterion or the "more than adequate" criterion. Again, statistics are reported for each dimension. Note that the two percentages can sum to more than 100% since a school can be both "adequate" or better and "more than adequate."

The bottom row of Table 3.2.1 shows that almost all the schools met the "adequate" or better criterion, but that hardly any met the "more than adequate" criterion. For a school to be "more than adequate" it must receive an RFII greater than 50, and the average RFII from 2004 to 2006 increased from 36 to 39.[15]

Likewise, for a school to be scored as "adequate" or better in Table 3.2.1, it must receive a statistic greater than 50 in the left half of the bottom (RFII) row of Table 3.2.0. A discussion of why 50 is used to define "adequate" or better is provided in the next subsection. Here, the averages run from 58 to 62, and most schools are above the 50 threshold. Note that dimensions with smaller numbers of items have larger standard deviations (Table 3.2.0), which will artificially cause more schools in those dimensions to fall above and below the 50 threshold, as shown in Table 3.2.1. In the tables that report a percentage of schools that are "adequate" or "more than adequate," the small number of items for some dimensions (e.g., Principal Professional Development) can therefore lead to misleading percentage statistics. Interpret them cautiously by first referring to the "number of items" column. By the same token, dimensions with a large number of items will tend to show the schools clumped within a fairly small range on the scale – the effect of measurement error has been largely removed. When the average measure on a given dimension is well above 50, and the spread of the schools is small, this can result in all of the schools being above the 50 mark, which translates as 100% of the schools scoring "adequate" or better on that dimension. This is the case for the Teacher Implementation dimension, for example. It is also why almost all schools are rated "adequate" or better on the RFII.

Tables 3.3.0 to 3.5.1 break out the results in Tables 3.2.0 and 3.2.1 by Years in Program (YIP). Inspecting the bottom rows of these tables, we see that schools that have been in the program for four years (YIP = 4) have been implementing the program fairly steadily from 2004 to 2006, increasing from 36 to 38.[16] The YIP = 3 schools have increased their implementation from 36 to 37, with a jump to 40 in 2006. The YIP = 2 schools show a 5-point jump after their first year from 34 to 39, reflecting increases across the 18 dimensions, especially in implementation of the Assurances. A large proportion of YIP 2 schools had low implementation measures in 2005, sufficient to form a bimodal distribution (two peaks) on the RFII continuum. In 2006, the distribution has coalesced into a more conventional distribution with

[15] An RFII greater than 50 would mean that teachers rated that school "more than adequate" on more than 50% of the items on the survey. Few schools generate that score from their teachers.

[16] The Standard Error of Measure for the RFII is approximately 1.5 points. Changes or differences that are more than two standard errors (3 points) in either direction may be considered statistically significant at the 95% confidence level.

a single mode. For the whole population of Reading First school's, implementation has increased from 36 in 2004 to 39 in 2006.

The YIP 2 and YIP 3 schools show steeper increases than the YIP 4 schools. This may be in part because the YIP 4 schools only received an RFII after their second year of implementation, when the program was already in place. In addition, as noted in Chapter 2, the three LEA cohorts to which these schools belong are demographically quite distinct, the earlier cohorts being more urban and the later cohorts more rural. The degree to which urban characteristics might dampen or moderate rate of implementation is not known, but it may be a factor.

As regards the terms "adequate" and "more than adequate," we direct the reader to Appendix E. There it is explained that the terms "adequate" and "more than adequate" were not actually used in the surveys, but represent a linguistic interpretation imposed on the item rating scales by the EAG, analogous to writing a key for a student exam. In perusing the questions on the survey in Appendices B-D, the reader may detect the basis for the EAG's interpretations of "adequacy" in a large number of questions, especially those with 4 or 5 options.

Table 3.2.0: All Schools, N (2006) = 856, Mean and Standard Deviation of Each Dimension, 2004 - 2006[1,2, 3, 4]

			% of the time teachers rated their school "Adequate" or better averaged across schools						% of the time teachers rated their school "More than Adequate" averaged across schools					
			2004		2005		2006		2004		2005		2006	
	Dimension	# Items, 2006	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	Teacher Professional Development	9	63	*14*	62	*10*	61	*10*	38	*14*	36	*10*	35	*10*
2	Coach Professional Development	7	74	*19*	72	*19*	65	*23*	58	*22*	56	*21*	48	*22*
3	Principal Professional Development	3	64	*29*	61	*30*	71	*27*	48	*30*	46	*30*	57	*30*
4	**Teacher Coach Professional Development**	**11**	**64**	***15***	**63**	***11***	**61**	***11***	**40**	***16***	**37**	***11***	**35**	***10***
5	Evaluation of Professional Development	5	76	*12*	80	*9*	81	*9*	11	*6*	14	*7*	15	*9*
6	Implementation, Assurances	11	69	*15*	73	*13*	71	*14*	44	*18*	48	*16*	46	*16*
7	School Implementation, Materials	175	57	*11*	58	*9*	63	*10*	36	*10*	37	*9*	41	*11*
8	School Implementation, Instruction	28	58	*6*	59	*6*	64	*6*	34	*6*	36	*6*	40	*6*
9	**School Implementation Overall**	**210**	**58**	***7***	**58**	***6***	**62**	***7***	**39**	***7***	**40**	***6***	**43**	***7***
10	Coaching Implementation	32	70	*14*	73	*12*	75	*10*	46	*16*	48	*14*	50	*12*
11	Teacher Implementation	33	70	*4*	72	*4*	75	*4*	48	*5*	50	*5*	54	*5*
12	Teacher RF Understanding	17	56	*7*	58	*6*	59	*7*	27	*6*	29	*5*	30	*6*
13	Coach RF Understanding	17	69	*14*	72	*12*	73	*7*	36	*15*	39	*14*	38	*8*
14	Principal RF Understanding	17	59	*15*	64	*14*	67	*8*	17	*9*	19	*10*	20	*6*
15	**Overall RF Understanding**	**17**	**57**	***7***	**59**	***7***	**61**	***6***	**23**	***5***	**25**	***5***	**26**	***5***
16	Teacher RF Evaluation	4	59	*13*	61	*12*	63	*13*	14	*7*	14	*7*	16	*9*
17	Coach RF Evaluation	6	74	*18*	75	*18*	79	*17*	20	*18*	19	*18*	24	*21*
18	Principal RF Evaluation	6	83	*17*	84	*17*	87	*12*	23	*24*	24	*24*	23	*20*
19	**RF Implementation Index (RFII)**	**238**	**58**	***6***	**59**	***5***	**62**	***5***	**36**	***6***	**36**	***5***	**39**	***6***

[1] Dimensions 4, 9, and 15 are in bold because they are weighted contributors to Dimension 19, the RFII. The 2006 statistics are across 856 schools from the point of view of teachers for Dimensions 1, 4, 5, 7, 8, 9, 10, 11, 12, 15, 16, and 19. Dimensions 2, 10, 13, and 17 are from the point of view of coaches. Dimensions 3, 14, and 18 are from the point of view of principals. Dimension 6 is from the point of view of coaches and principals together. The 2004 statistics are across 628 schools, the 2005 statistics across 808 schools.

[2] The statistics in bold in the right six columns report the official RFII. Their counterparts in the left columns refer to what the RFII would have been had it been decided to use the "adequate" criterion to define implementation.

[3] The N count of schools in these tables does not exactly match those for all Reading First schools as reported in Chapters 2 and 4. The schools reported here are those whose teachers, coaches, and principals returned surveys.

[4] The rows in bold are those dimensions used in calculating the RFII.

Table 3.2.1: All Schools, N (2006) = 856, Percent of Schools Considered "Adequate" and "More than Adequate", 2004 – 2006[1,2]

			% of schools considered "Adequate" or better by teachers			% of schools considered "More than Adequate" by teachers		
	Dimension	# Items, 2006	2004	2005	2006	2004	2005	2006
1	Teacher Professional Development	9	87%	92%	89%	17%	9%	6%
2	Coach Professional Development	7	90%	90%	84%	68%	67%	51%
3	Principal Professional Development	3	67%	69%	78%	42%	41%	60%
4	**Teacher Coach Professional Development**	**11**	**89%**	**92%**	**89%**	**20%**	**12%**	**7%**
5	Evaluation of Professional Development	5	97%	99%	100%	0%	1%	1%
6	Implementation Assurances	11	88%	94%	92%	32%	41%	39%
7	School Implementation Material	175	76%	83%	94%	10%	7%	16%
8	School Implementation Instruction	28	90%	93%	99%	0%	1%	7%
9	**School Implementation Overall**	**210**	**88%**	**92%**	**97%**	**6%**	**5%**	**14%**
10	Coaching Implementation	32	92%	96%	98%	39%	46%	53%
11	Teacher Implementation	33	100%	100%	100%	35%	50%	80%
12	Teacher RF Understanding	17	79%	89%	95%	0%	0%	1%
13	Coach RF Understanding	17	87%	94%	100%	18%	20%	7%
14	Principal RF Understanding	17	73%	80%	97%	1%	0%	0%
15	**Overall RF Understanding**	**17**	**83%**	**92%**	**96%**	**0%**	**0%**	**0%**
16	Teacher RF Evaluation	4	77%	82%	83%	0%	0%	1%
17	Coach RF Evaluation	6	90%	91%	95%	5%	6%	10%
18	Principal RF Evaluation	6	93%	94%	98%	19%	20%	8%
19	**RF Implementation Index (RFII)**	**238**	**92%**	**96%**	**98%**	**1%**	**1%**	**4%**

[1] The 2006 statistics are across 856 schools from the point of view of teachers for Dimensions 1, 4, 5, 7, 8, 9, 10, 11, 12, 15, 16, and 19. Dimensions 2, 10, 13, and 17 are from the point of view of coaches. Dimensions 3, 14, and 18 are from the point of view of principals. Dimension 6 is from the point of view of coaches and principals together.

[2] The 2005 statistics are across 808 schools. The 2004 statistics are across 628 schools.

Table 3.3.0: Schools with 4 Years in Program, N (2006) = 267, Mean and Standard Deviation of Each Dimension, 2004 - 2006

			% of the time teachers rated their school "Adequate" or better averaged across schools						% of the time teachers rated their school "More than Adequate" averaged across schools					
			2004		2005		2006		2004		2005		2006	
	Dimension	# Items, 2006	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	Teacher Professional Development	9	62	*9*	63	*9*	57	*10*	36	*9*	37	*9*	32	*9*
2	Coach Professional Development	7	75	*20*	72	*21*	65	*25*	60	*23*	57	*23*	49	*24*
3	Principal Professional Development	3	61	*28*	55	*31*	69	*28*	44	*28*	40	*28*	53	*29*
4	**Teacher Coach Professional Development**	**11**	**64**	***9***	**64**	***9***	**58**	***11***	**37**	***10***	**37**	***10***	**32**	***10***
5	Evaluation of Professional Development	5	79	*10*	81	*8*	82	*7*	13	*6*	14	*7*	15	*7*
6	Implementation, Assurances	11	70	*15*	73	*12*	66	*12*	44	*17*	47	*14*	39	*13*
7	School Implementation, Materials	175	54	*8*	55	*8*	57	*7*	33	*8*	33	*8*	35	*7*
8	School Implementation, Instruction	28	59	*5*	60	*5*	64	*5*	36	*5*	36	*5*	40	*5*
9	**School Implementation Overall**	**210**	**58**	***6***	**58**	***6***	**60**	***6***	**39**	***6***	**39**	***6***	**42**	***6***
10	Coaching Implementation	32	72	*13*	74	*11*	75	*9*	47	*14*	50	*13*	50	*11*
11	Teacher Implementation	33	72	*4*	73	*4*	75	*4*	50	*5*	51	*5*	55	*5*
12	Teacher RF Understanding	17	59	*6*	60	*6*	62	*5*	29	*5*	31	*5*	32	*5*
13	Coach RF Understanding	17	73	*13*	74	*11*	74	*6*	40	*15*	41	*14*	39	*7*
14	Principal RF Understanding	17	65	*14*	67	*14*	67	*7*	20	*10*	22	*10*	20	*5*
15	**Overall RF Understanding**	**17**	**61**	***6***	**62**	***6***	**64**	***6***	**26**	***5***	**27**	***5***	**28**	***5***
16	Teacher RF Evaluation	4	61	*13*	60	*13*	61	*13*	15	*7*	14	*8*	15	*8*
17	Coach RF Evaluation	6	78	*17*	76	*18*	80	*16*	22	*20*	21	*19*	24	*21*
18	Principal RF Evaluation	6	85	*16*	83	*18*	86	*14*	26	*25*	24	*25*	23	*20*
19	**RF Implementation Index (RFII)**	**238**	**59**	***5***	**59**	***5***	**61**	***5***	**36**	***5***	**37**	***5***	**38**	***5***

Table 3.3.1: Schools with 4 Years in Program, N (2006) = 267, Percent of Schools Considered "Adequate" and "More than Adequate", 2004 - 2006

			% of schools considered "Adequate" or better by teachers			% of schools considered "More than Adequate" by teachers		
	Dimension	# Items, 2006	2004	2005	2006	2004	2005	2006
1	Teacher Professional Development	9	91%	94%	80%	7%	9%	2%
2	Coach Professional Development	7	90%	89%	80%	68%	68%	58%
3	Principal Professional Development	3	63%	60%	76%	35%	33%	57%
4	**Teacher Coach Professional Development**	**11**	**95%**	**96%**	**83%**	**10%**	**12%**	**4%**
5	Evaluation of Professional Development	5	99%	100%	100%	0%	0%	0%
6	Implementation Assurances	11	88%	97%	89%	36%	38%	19%
7	School Implementation Material	175	71%	71%	89%	3%	4%	4%
8	School Implementation Instruction	28	96%	96%	100%	0%	1%	5%
9	**School Implementation Overall**	**210**	**92%**	**92%**	**95%**	**3%**	**6%**	**7%**
10	Coaching Implementation	32	96%	97%	98%	41%	52%	51%
11	Teacher Implementation	33	100%	100%	100%	47%	60%	84%
12	Teacher RF Understanding	17	90%	94%	97%	0%	0%	0%
13	Coach RF Understanding	17	92%	97%	100%	24%	25%	6%
14	Principal RF Understanding	17	84%	88%	98%	2%	0%	0%
15	**Overall RF Understanding**	**17**	**92%**	**96%**	**97%**	**0%**	**0%**	**0%**
16	Teacher RF Evaluation	4	81%	80%	78%	0%	1%	0%
17	Coach RF Evaluation	6	95%	92%	97%	8%	7%	10%
18	Principal RF Evaluation	6	94%	92%	96%	24%	21%	9%
19	**RF Implementation Index (RFII)**	**238**	**97%**	**97%**	**99%**	**0%**	**1%**	**3%**

Table 3.4.0: Schools with 3 Years in Program, N (2006) = 379, Mean and Standard Deviation of Each Dimension, 2004 - 2006

			% of the time teachers rated their school "Adequate" or better averaged across schools						% of the time teachers rated their school "More than Adequate" averaged across schools					
			2004		2005		2006		2004		2005		2006	
	Dimension	# Items, 2006	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	Teacher Professional Development	9	63	*17*	64	*9*	63	*10*	39	*17*	38	*10*	37	*11*
2	Coach Professional Development	7	73	*18*	73	*17*	65	*22*	57	*21*	56	*19*	48	*21*
3	Principal Professional Development	3	65	*30*	67	*28*	74	*27*	50	*31*	52	*29*	60	*30*
4	**Teacher Coach Professional Development**	**11**	**65**	***18***	**65**	***9***	**64**	***11***	**41**	***20***	**38**	***10***	**38**	***11***
5	Evaluation of Professional Development	5	74	*13*	80	*8*	82	*10*	11	*6*	14	*7*	16	*11*
6	Implementation, Assurances	11	68	*15*	76	*13*	74	*15*	43	*18*	53	*17*	50	*17*
7	School Implementation, Materials	175	60	*12*	62	*8*	66	*10*	38	*12*	40	*9*	45	*12*
8	School Implementation, Instruction	28	57	*7*	59	*6*	64	*6*	34	*6*	35	*6*	40	*7*
9	**School Implementation Overall**	**210**	**57**	***8***	**60**	***6***	**63**	***7***	**39**	***8***	**41**	***7***	**45**	***8***
10	Coaching Implementation	32	69	*15*	73	*12*	75	*10*	45	*17*	48	*15*	51	*13*
11	Teacher Implementation	33	69	*5*	71	*4*	75	*4*	47	*5*	50	*5*	54	*5*
12	Teacher RF Understanding	17	53	*6*	56	*6*	58	*8*	25	*5*	27	*5*	29	*7*
13	Coach RF Understanding	17	66	*14*	71	*12*	73	*7*	33	*15*	38	*13*	38	*9*
14	Principal RF Understanding	17	57	*15*	61	*14*	67	*8*	15	*8*	18	*9*	20	*6*
15	**Overall RF Understanding**	**17**	**54**	***7***	**58**	***6***	**60**	***6***	**21**	***5***	**23**	***4***	**25**	***5***
16	Teacher RF Evaluation	4	58	*13*	62	*12*	65	*13*	13	*6*	15	*7*	17	*10*
17	Coach RF Evaluation	6	72	*19*	76	*16*	80	*16*	17	*16*	20	*19*	25	*21*
18	Principal RF Evaluation	6	82	*18*	85	*16*	88	*13*	22	*24*	24	*24*	24	*20*
19	**RF Implementation Index (RFII)**	**238**	**57**	***7***	**60**	***5***	**62**	***6***	**36**	***6***	**37**	***5***	**40**	***6***

Table 3.4.1: Schools with 3 Years in Program, N (2006) = 379, Percent of Schools Considered "Adequate" and "More than Adequate", 2004 - 2006

			% of schools considered "Adequate" or better by teachers			% of schools considered "More than Adequate" by teachers		
	Dimension	# Items, 2006	2004	2005	2006	2004	2005	2006
1	Teacher Professional Development	9	84%	94%	93%	24%	11%	10%
2	Coach Professional Development	7	89%	93%	86%	68%	70%	49%
3	Principal Professional Development	3	69%	78%	80%	46%	49%	64%
4	**Teacher Coach Professional Development**	**11**	**85%**	**94%**	**92%**	**28%**	**13%**	**12%**
5	Evaluation of Professional Development	5	95%	99%	99%	0%	1%	2%
6	Implementation Assurances	11	88%	96%	93%	29%	54%	52%
7	School Implementation Material	175	80%	94%	96%	15%	12%	25%
8	School Implementation Instruction	28	85%	93%	98%	1%	1%	9%
9	**School Implementation Overall**	**210**	**84%**	**93%**	**97%**	**9%**	**7%**	**20%**
10	Coaching Implementation	32	90%	95%	99%	37%	44%	53%
11	Teacher Implementation	33	99%	100%	100%	26%	47%	80%
12	Teacher RF Understanding	17	69%	87%	93%	0%	0%	1%
13	Coach RF Understanding	17	83%	94%	99%	12%	16%	10%
14	Principal RF Understanding	17	68%	75%	97%	0%	0%	0%
15	**Overall RF Understanding**	**17**	**75%**	**92%**	**94%**	**0%**	**0%**	**1%**
16	Teacher RF Evaluation	4	74%	85%	87%	0%	0%	2%
17	Coach RF Evaluation	6	86%	93%	94%	3%	7%	10%
18	Principal RF Evaluation	6	92%	96%	98%	17%	21%	7%
19	**RF Implementation Index (RFII)**	**238**	**87%**	**97%**	**98%**	**1%**	**2%**	**6%**

Table 3.5.0: Schools with 2 Years in Program, N (2006) = 158, Mean and Standard Deviation of Each Dimension, 2005 - 2006

			% of the time teachers rated their school "Adequate" or better averaged across schools						% of the time teachers rated their school "More than Adequate" averaged across schools					
			2004		2005		2006		2004		2005		2006	
	Dimension	# Items, 2006	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	Teacher Professional Development	9			56	*14*	61	*8*			31	*12*	35	*8*
2	Coach Professional Development	7			69	*18*	66	*21*			52	*21*	48	*19*
3	Principal Professional Development	3			63	*33*	72	*26*			49	*33*	57	*29*
4	**Teacher Coach Professional Development**	**11**			**56**	***14***	**61**	***8***			**31**	***12***	**35**	***8***
5	Evaluation of Professional Development	5			77	*10*	80	*10*			12	*9*	14	*7*
6	Implementation, Assurances	11			64	*15*	73	*12*			37	*15*	48	*15*
7	School Implementation, Materials	175			59	*7*	65	*7*			37	*7*	43	*8*
8	School Implementation, Instruction	28			57	*7*	64	*6*			33	*6*	40	*6*
9	**School Implementation Overall**	**210**			**56**	***5***	**62**	***6***			**38**	***5***	**44**	***6***
10	Coaching Implementation	32			69	*12*	76	*10*			44	*14*	51	*12*
11	Teacher Implementation	33			69	*5*	74	*4*			47	*5*	53	*5*
12	Teacher RF Understanding	17			55	*6*	58	*5*			26	*5*	28	*5*
13	Coach RF Understanding	17			67	*14*	72	*7*			35	*15*	37	*7*
14	Principal RF Understanding	17			60	*15*	66	*7*			17	*10*	19	*5*
15	**Overall RF Understanding**	**17**			**56**	***6***	**59**	***5***			**22**	***4***	**24**	***4***
16	Teacher RF Evaluation	4			59	*12*	62	*13*			13	*7*	15	*8*
17	Coach RF Evaluation	6			66	*19*	79	*18*			12	*12*	24	*22*
18	Principal RF Evaluation	6			84	*14*	88	*10*			20	*18*	22	*19*
19	**RF Implementation Index (RFII)**	**238**			**56**	***5***	**61**	***5***			**34**	***4***	**39**	***5***

Table 3.5.1: Schools with 2 Years in Program, N (2006) = 158, Percent of Schools Considered "Adequate" and "More than Adequate", 2005 - 2006

			% of schools considered "Adequate" or better by teachers			% of schools considered "More than Adequate" by teachers		
	Dimension	# Items, 2006	2004	2005	2006	2004	2005	2006
1	Teacher Professional Development	9		78%	92%		5%	4%
2	Coach Professional Development	7		82%	88%		55%	45%
3	Principal Professional Development	3		72%	77%		43%	60%
4	**Teacher Coach Professional Development**	**11**		**76%**	**91%**		**6%**	**3%**
5	Evaluation of Professional Development	5		98%	100%		1%	0%
6	Implementation Assurances	11		80%	96%		13%	45%
7	School Implementation Material	175		89%	97%		2%	16%
8	School Implementation Instruction	28		82%	99%		0%	5%
9	**School Implementation Overall**	**210**		**84%**	**97%**		**0%**	**14%**
10	Coaching Implementation	32		92%	99%		32%	58%
11	Teacher Implementation	33		100%	100%		31%	75%
12	Teacher RF Understanding	17		78%	94%		0%	0%
13	Coach RF Understanding	17		85%	99%		18%	2%
14	Principal RF Understanding	17		73%	97%		1%	0%
15	**Overall RF Understanding**	**17**		**78%**	**96%**		**0%**	**0%**
16	Teacher RF Evaluation	4		78%	85%		0%	0%
17	Coach RF Evaluation	6		80%	94%		1%	10%
18	Principal RF Evaluation	6		96%	100%		14%	9%
19	**RF Implementation Index (RFII)**	**238**		**90%**	**99%**		**0%**	**4%**

Are schools implementing "adequately" from the point of view of teachers?

There has been no official decision by the EAG to set an "adequate" or "more than adequate" cut-point on the RFII scale. By default, we employ the psychometric convention that a greater than 50% probability of scoring above the "adequate" threshold means that the school is "above" that threshold. Thus, if teachers rated their school "adequate" or above on more than 50% of the items, we say that the school is "adequate" or better. Table 3.2.0 tells us that in 2006 the average score in the "adequate" or better column across schools was 62, 12 points above the 50 threshold, and that 98% of California Reading First schools scored above this threshold. That is the basis for our claim that, on average, teachers feel that schools are implementing Reading First adequately or better.

Tables 3.2.0 and 3.2.1 also tell us that in 2006 only 4% of schools are implementing Reading First "more than adequately" according to teachers. Small as this percentage is, it represents an improvement over the 1% of schools reported to be implementing "more than adequately" in 2003 and 2004.

Profile of the "Comparable Non-Reading First" School

What is the most likely RFII of a "Comparable Non-Reading First" school?

Our report makes frequent use of the concept of a statistical control group of hypothetical "comparable non-Reading First" schools. For reasons described in Chapter 2 and above in Chapter 3, the population of actual "non-Reading First" schools is not suitable for making comparisons because many contain program elements that are required in Reading First.

As has been discussed, the statistical control group should have the same characteristics as the relevant Reading First population. This translates into the same average percent of English Learners, average percent of Socio-Economically Disadvantaged students, the same achievement starting point, and the same number of years in program. There remains the "treatment" variable. What should be the average RFII of a school that is not in the program? If we were to administer the Reading First implementation survey to a school that is in all respects the same as a typical Reading First school except that it has not adopted the program, what would its RFII be?

This question cannot be answered empirically because there are few, if any, schools in our Reading First population that have not been implementing Reading First to at least some degree. That means we need a theoretical basis for defining the "comparable non-Reading First" school. It might be tempting to assign it an RFII of zero, which is the smallest possible number on the RFII scale. This would correspond to a school whose teachers feel that the school is not "more than adequate" on *any* of the implementation items in the survey. There are two problems with this definition:

1. A perusal of the RFII distribution (Figures 3.2.1-3.2.4) shows that no schools have an RFII close to zero (the lowest is a 2003 school with an RFII of 17), and one would expect *some* schools in the total Reading First population, particularly those just entering the program, to approach the implementation level of a non-Reading First school.
2. As discussed in Chapter 2, many non-Reading First schools are implementing program elements that are required components of Reading First – the use of the SRA Open Court Reading or Houghton-Mifflin Reading curriculum, teacher participation in AB 466 and Advanced Series professional development, AB 75 principal training, use of the 6-8 week Skills Assessments, and the funding of reading coaches. Therefore, a typical non-Reading First school would be expected to get an RFII greater than zero.
3. Most of the items in the survey are forced multiple-choice and were keyed to have the equivalent of four scoring options, equivalent to a four-point rating scale ranging from "Far Below Adequate" to "Less than Adequate" to "Adequate" to "More than Adequate." Many of these "four-point" items were written so that it would not be immediately obvious which of the four item response options was the "correct" or "More than Adequate" option. Therefore, by random chance we would expect to see 25% of the items being scored "More than Adequate," making an RFII of zero unlikely.

In fact, it is this latter consideration that suggests a theoretical basis for determining the RFII of a "non-Reading First" school -- select the RFII that a school would receive if its respondents selected options at random. That leads us to conclude that the expected RFII of a non-Reading First school is 25.

Note that even though 25 is the score expected by chance, that does not mean the school is implementing *no* Reading First-like program elements. An RFII of 25 is conceptually equivalent to: using only one quarter of the program materials, attending one long day of professional development during the 40-hour AB 466 training, or understanding a quarter of the Reading First instructional practices. Even this degree of implementation – that of a non-Reading First school – might be expected to produce achievement gains relative to what one would expect without any of the program elements.

Empirical considerations also support specifying an RFII of 25 for the "comparable non-Reading First school." Examining the RFII distributions for each YIP (Figures 3.2.1 – 3.2.3), we see that the distribution of the lower tail tends to approach zero in the mid-twenties. An examination of the RFIIs in individual years shows that in 2003 there were 15 schools with an RFII less than 25, in 2004 there were 7, in 2005 there were 5, and when all three years are averaged together the lowest Mean RFII falls at exactly 25. Thus, the empirical distribution supports the theoretical expectation of 25 being the lower bound of the RFII.

Conclusions

This chapter finds:

- Measuring implementation is an essential element in assessing program effectiveness (i.e., the potential of a program to produce achievement gains given a sufficient level of implementation).
- School Reading First implementation measures can be computed from survey data given proper attention to methodological issues. In particular, care needs to be taken to tailor survey forms to specific types of respondents, to equate their responses across forms and years with the appropriate measurement technology, to construct and empirically verify that questions hang together in their intended dimensions, and to reduce incentives to overstate actual program implementation. Given that these conditions are met—and it appears that they have been—a Reading First Implementation Index (RFII) can be computed that is comparable in reliability to a standardized achievement test.
- The RFII can be interpreted as a (theoretical) percentage of times that teachers rate their schools "more than adequate" on relevant survey questions. Using the distribution of school RFII measures, it is possible to state how many schools in the state meet both the "adequate" or better standard and the "more than adequate" standard from the point of view of teachers on selected dimensions.
- The average RFII across all schools was 39 in 2006, compared to 36 in 2005 and 2004.
- 98% of schools in 2006 were rated "adequate" or better, up from 96% in 2005 and 92% in 2004.
- YIP 2, YIP 3, and YIP 4 schools have all shown steady growth in implementation, especially YIPs 2 and 3. It appears that most of this growth is achieved by the end of the second year of program funding.
- YIP 2 schools (labeled Cohort 3 schools in the 2005 Report) have shown a surge in implementation in their second year, reflected particularly in the dimension called "Implementation of Assurances." Last year (2005), we noted that a large proportion of these schools indicated low implementation, forming a bimodal distribution on the RFII continuum. With the recent upsurge, the YIP 2 schools now form a distribution with one mode and they are implementing on a par with other schools in the program.
- The definition of a "non-implementing" Reading First school is one that receives an RFII of 25. This number was arrived at as a result of both theoretical and empirical considerations.

References

Linacre, J. M. (1994), Many-Facet Rasch Measurement, Chicago: MESA Press. Also visit www.rasch.org/rmt/rmt162h.htm.

Schiller, E. (2001). Policy Brief I: Using Implementation Data to Study State, District, and School Impacts, Bethesda, MD: Abt Associates Inc.

Chapter 4: Achievement Results

This chapter addresses the questions: *What is the impact of the Reading First program on K-3 students in participating districts and schools? What evidence is there that the Reading First program has improved the effectiveness of participating schools and districts?* Achievement results for Reading First schools are presented in terms of the Standardized Testing and Reporting (STAR) Program – the California Standards Test (CST) and the California Achievement Test (CAT/6) – and the End-of-Year (EOY) curriculum-embedded assessments. Achievement is compared in four ways: a) between years (gain scores); b) between Reading First and non-Reading First schools; c) between Reading First schools and a statistical control group; d) between High Implementation and Low Implementation Reading First schools.

The Argument

The objective of this evaluation is to determine whether or not, and to what degree, the Reading First program is effective. What is meant by "effective"? According to the federal guidelines for Reading First, the program is effective to the degree it ensures "that all children in America learn to read well by the end of third grade" (U.S. Department of Education, 2002). There are four ways to examine the effect of Reading First on reading in California given the limitations of a non-experimental research design.

1. Measure the size of the achievement gains of Reading First schools
2. Compare Reading First schools to comparable non-Reading First schools
3. Compare Reading First schools to a statistical control group
4. Compare High Implementation Reading First schools to Low Implementation Reading First schools

The first approach looks at the absolute size of the achievement gains of Reading First schools from when they started (and implementation was low) to the present (when the program has been in place and is presumably well-implemented). A significant positive gain would suggest that Reading First is working. However, it is difficult to rule out the possibility that such gains are the effect of other causal factors that came into play over the same time period, especially factors that may cause all schools to show an increase or decrease in scores. That is why it is useful to compare Reading First schools to comparable schools across the same time span, as a way to control for causal factors that may universally lead to increases or decreases in scores.

The second approach, comparing Reading First schools to comparable non-Reading First schools, is discussed in Chapter 2. Although such comparisons were employed in previous reports, we opted not to take that approach in Year 4 due largely to the realization that many non-Reading First schools are

implementing educational components that are similar to those required by Reading First, making them inappropriate as a control group.

The third approach, comparing Reading First schools to a statistical control group, is discussed in Chapter 2 and Chapter 3. It entails using multiple regression to predict the achievement gains of a hypothetical sample of schools that are similar to Reading First schools except that they have a minimal level of implementation, as defined using the Reading First Implementation Index (RFII). This approach has been employed in this report.

The fourth approach is a variation of the third, except that we are comparing a sample of actual low-implementing Reading First schools with a sample of actual high-implementing Reading First schools. As with the third approach, it depends on having a way to measure the degree of implementation within a school and a clear definition of what that implementation entails. It is important that implementation not be an accidental proxy of some other causal factor, such as school resources or demographic advantage. Assuming a valid implementation measure, a significant positive difference between High and Low Implementing Reading First schools establishes that Reading First at least has the *potential* of working if properly implemented. If implementation of Reading First bears little or no relation to achievement gains (assuming both measures are sound), then the program may not be effective.

Therefore, Reading First will be said to show evidence of being effective to the degree that:

1. Achievement gains in Reading First schools are positive;
2. Reading First schools show higher achievement gains than non-Reading First schools;
3. Reading First schools show higher achievement gains than a statistical control group;
4. High Implementing Reading First schools show higher achievement gains than Low Implementing Reading First schools.

The Achievement and Gain Metrics

In attempting comparisons, it is important to be familiar with the various achievement metrics used to measure growth in reading ability. The school achievement metrics used in our evaluation are: (1) the California Standards Test or CST, (2) the norm-referenced California Achievement Test Mean Percentile Ranking (CAT/6 MeanPR), (3) the EOY test administered only by funded Reading First districts, and (4) the Reading First Achievement Index (RFAI), a weighted combination of the CST, CAT/6 and EOY data, also available only for Reading First schools. In keeping with our research design, change in achievement is measured over time in year increments, the unit of analysis being not the student but the school, thus the percentage of students within a school meeting some specified performance level or benchmark. The statistics in the charts and tables that follow are generally in this percentage metric and report the 2006

percentage of students meeting benchmark minus the percentage for the year immediately preceding the program, which yields a gain score.[17] The CAT/6, EOY, and RFAI metrics and gain scores are handled a little differently, as described below. Thus, the reported gain scores can in all cases except the CAT/6 be interpreted as the subtraction of the percentage of students meeting some specified benchmark in one year from the corresponding percentage in a later year.

YIPs and Measures of School Progress

What are Reading First YIPs and why use YIPs instead of cohorts?

YIP stands for "Years in Program" and refers to how many years a school has been in Reading First as of 2006. The maximum YIP in 2006 is four years. The minimum is one year, though the YIP 1 schools are so few and unusual (including a large number of pre-Kindergarten and Kindergarten schools) that they are not discussed in detail in this Report.

The Year 3 Report uses the term "cohort" instead of YIP, which refers to the year in which a school started in the program. Thus what is called a YIP 4 school in the Year 4 Report is called a Cohort 1 school in the Year 3 Report in which we used the terms interchangeably. The terminology was changed because it was noted that the term "cohort" (synonymous with "Round of Funding") should be reserved for the year that an *LEA* or *district* first received Reading First funding, which is not necessarily the same as when an individual school *within* that district might have begun implementing the program. According to Reading First guidelines, awards were made to LEAs, not individual schools. However, as some schools dropped from the program (due to school closures, mergers or other circumstances), LEAs may have replaced some schools. Thus, a Cohort 1 district, for example, would have most schools with four years in the program, but may have one or two with fewer than four years. Therefore, we refer to "LEA Cohorts" but "School Years in Program." For example Los Angeles Unified School District, which entered Reading First in Cohort 1, has 28 schools that began implementing the program in 2005.

Because we have chosen the school as our unit of analysis, our achievement tables are in terms of YIPs, not cohorts.

Why distinguish between YIPs (or cohorts) at all? It is often found in educational research that intervention program effects vary over time and across cohorts. For example, a change in state achievement tests can cause an anomaly in gain scores that affects the perceived growth curves of different cohorts, or YIPs, quite differently. It may also turn out that they have different demographic

[17] Such percentage metrics are not the preferred way to measure gains since they discard large amounts of information and are prone to measurement errors, such as floor and ceiling effects. Mean scale scores are the preferred metric. However, the percentage of students achieving benchmark is relatively easy to interpret and is useful for comparing results from different assessments. Also, in some cases it is all that is available. The CAT/6 is reported in a Mean Percentile Ranking metric (MeanPR) which has advantages over the percent of students metric.

profiles and thus respond to educational programs differently. This proved to be particularly true of Reading First, whose first cohort of LEAs was much more urban than subsequent cohorts. Because such cohort or YIP effects appear to be an important aspect of Reading First, we have elected to analyze them separately.

As explained in Chapters 2 and 3, there are four Reading First YIPs, though we are focusing on YIPs 4, 3, and 2. YIP 4 refers to schools that received the first round of funding and began implementation in the 2002-2003 academic year, though funding was not received until the middle of the school year. YIP 3 refers to schools that received the second round of funding and began implementing the program just prior to the 2003-2004 academic year. YIP 2 refers to schools that received the third round of funding and began implementation just prior to the 2004-2005 academic year. YIP 1 refers to schools that received the fourth round of funding and began implementation just prior to the 2005-2006 academic year. For each YIP, cross-year gains from their starting year (or the year prior in the case of STAR) to 2006 are reported in each achievement metric for Reading First and non-Reading First schools, with results broken out for High and Low Implementers of Reading First, as well as the relevant statistical control group.

Measures of School Progress

School progress or growth, also known as achievement gains, is measured using the CSTs, the CAT/6 MeanPR, the End-of-Year (EOY) test, and the Reading First Achievement Index (RFAI), which is a composite of the others. Each metric has unique characteristics.

1. The California Standards Test (CSTs). The CSTs are administered to students in Grades 2 and above toward the end of the school year. For purposes of this study, we use the English Language Arts (ELA) component of the CSTs for Grades 2 and 3. Within ELA, we study the percentage of students per school that fall within each of the three following performance categories, which are a simplification of the five CST performance categories (Advanced, Proficient, Basic, Below Basic, Far Below Basic).
    a. "Proficient and Above" means the percentage of students in a school that are in the Proficient and Advanced performance categories;
    b. "Basic" means the percentage of students in a school that are in the Basic performance category;
    c. "Below Basic and Far Below Basic" means the percentage of students in a school that are in the bottom two performance categories.

    The CST gain score for each school is the 2006 percentage of students in a specified category

minus the corresponding percentage in the year immediately preceding the first year of Reading First funding. These gain scores are averaged across schools within a cohort to yield the "Average School Gain Score" statistics reported in the tables below.

By reporting CST gain scores for each of our three defined performance categories, we track not only student movement into the Proficient and Above category but also movement out of the Below Basic and Far Below Basic categories into higher categories. It is just as important to study movement out of the bottom categories as movement into the upper categories since reading programs often have differential impacts on diverse student groups. Also, studying program impact on the bottom categories sheds light on the long-term sustainability of growth trends associated with the program, since movement into the upper categories relies on movement out of the bottom categories.

2. The CAT/6 MeanPR. As of the Spring 2005 administration of the California STAR assessment, the CAT/6 component of STAR was discontinued in all elementary grades except for Grade 3, so only Grade 3 CAT/6 Reading, Language Arts, and Spelling data are used in this study. The "MeanPR" of a school is the average of the National Percentile Rank (NPR) scores of each of its students after they have been suitably rescaled for purposes of aggregation. The MeanPR gain score for each school is its MeanPR in 2006 minus its MeanPR in the year immediately preceding its first year of Reading First implementation. The CAT/6 gain scores reported in the tables below are an average of these MeanPR gain scores across schools in that cohort. Note that they are interpreted as a change in national percentile ranking, not as a change in the percent of students meeting some benchmark or performance standard.

3. End-of-Year Test (EOY). As the name suggests, the EOY is a curriculum-based test administered by all Reading First schools to students in Grades K-3 at the end of the academic year. The Kindergarten EOY test consists of eight subtests: Consonants, Lower Case Letters, Phonics, Rhyming, Syllables, Upper Case Letters, Vowels, and Consonant-Vowel-Consonant. The EOY tests for Grades 1, 2 and 3 consist of timed oral reading fluency passages in which success is measured in terms of words correct per minute. The EOY is unique and valuable for this study because it is the only test that can be used to measure achievement in Grades K and 1. (The STAR metrics are available only for Grades 2 and above.) It is also the only test that is also administered in Spanish to students in "waiver" Reading First classrooms (in which instruction is conducted in Spanish). One limitation of the EOY is that it is not administered to non-Reading First schools, which prevents comparisons in this metric. Another is the relative lack of information about its psychometric properties – its reliability, internal consistency, and

comparability across testing populations – though the overall content validity of similar tests is established (Hasbrouck & Tindal, 2005, Yovanoff & Tindal, 2003). The correlation between EOY scores and CST scores has also been noted (Parker, 2003). The EOY score for each grade within a school consists of the percentage of students that meet the benchmark established for that grade based on national norms recommended by Hasbrouck & Tindal. The gain score for that grade is its 2006 EOY minus its EOY at the end of its *first* year of Reading First funding, which for YIPs 4 and 3 is 2004. YIP 2 has only a 1-year EOY gain score since it was not in Reading First in 2004. Note that, unlike the CSTs and CAT/6, the base year is not the year *preceding* the first year of Reading First implementation. Also note that the gain scores are averaged within a given grade across schools in the YIP to produce an average gain score for that grade.

4. Reading First Achievement Index (RFAI). The RFAI is a weighted combination of school-level percentages of students meeting various performance levels and benchmarks drawn from the CSTs, the CAT/6 MeanPR, and the EOY, with the heaviest weights placed on the CSTs. Refer to Appendix G for a detailed explanation of how the RFAI is computed. The RFAI was first computed in 2004. As of this study there are three RFAI indices for YIPs 4 and 3 (for 2004, 2005, and 2006), and two for YIP 2 (2005 and 2006). Because it contains an EOY component, the RFAI can be computed only for Reading First schools. Like the CST, each school RFAI can be interpreted as a percentage of students meeting a set of combined benchmarks and performance levels. The gain score for that school is its 2006 RFAI minus its RFAI at the end of its *first* year of Reading First implementation, which for YIPs 4 and 3 is 2004. For YIP 2 it is 2005.

The Relationship between Reading First YIPs and Achievement Gains

Achievement Gains on the STAR CSTs and CAT/6 MeanPR

The first Reading First LEA cohort (Cohort 1) was funded in the 2002-2003 academic year. As of this report, schools in that Cohort (YIP = 4) tested for STAR in 2003, 2004, 2005, and 2006 since the Reading First program began. However, we use STAR 2002 as the baseline year or pre-program year and compute student gains against that year. Therefore YIP 4 schools have a maximum of four years of student gains data, from 2002-03, 2003-04, 2004-05, and 2005-06. YIP 3 schools first implemented the program in the 2003-2004 academic year and have tested in 2004 and 2005 since program implementation. STAR 2003 is their baseline year so this YIP has a maximum of three years of student gains data. YIP 2, entering Reading First in 2004-05, has two years of gains data, 2004-05, and 2005-06. YIP 1 has only one year of gains data (which we do not report). The point is that for each YIP there are multiple ways to compute 1-year gains, 2-years gains, and so forth. However, because we are most interested in measuring total growth, we restrict our examination to gains relative to just the baseline year.

Achievement Gains on the End-of-Year Test

The EOY was first administered in 2003, the end of the first academic year in Reading First for YIP 4 schools. The 2003 EOY database proved too incomplete to use as a basis for computing 2-year gains. Therefore, for YIP 4 we report only 2-year gains, 2004 to 2006. For YIP 3 we report the same 2-year gain, 2004 to 2006. There is a 1-year EOY gain for YIP 2. One issue in the use of the EOY test is that in 2005 schools with "waiver classrooms" (as defined by AB 1485) had the option of testing students in those classrooms on a Spanish version of the EOY test. Testing in Spanish could only be done in grades K-2, since schools are required by law to test all students in English at the end of Grade 3. Therefore, in 2005, EOY data existed for Spanish-instruction and English-instruction students in Grades K-2. Since this was new as of 2005, there is no comparable Spanish-instruction group for 2004. Therefore, EOY gain scores were computed and presented only for non-waiver or English-instruction students for Grades K-3. In 2006, as part of the regulations on measuring "significant progress," it was decided to include Grade 3 Spanish-scores for students in Grade 3 Spanish-instruction classrooms in the RFAI calculation, and so we also report them separately.

For both Spanish-instruction and English-instruction students for Grades K-2 we report the percentage of students at benchmark. Note, however, that the two language groups are tested using different oral fluency prompts that are not psychometrically equated, which complicates efforts to make comparisons.

Achievement Gains on the RFAI

As of this Year 4 Report the published RFAIs are for 2004, 2005, and 2006 for YIPs 4 and 3, and 2005 and 2006 for YIP 2. When reporting RFAI gains, we report only 2-year gains for YIPs 4 and 3 and 1-year gains for YIP 2.

Comparison of Reading First to non-Reading First Schools

Chapter 2 provides a detailed description of how we decided to compare Reading First to non-Reading First schools. Reading First schools were initially compared to two comparison groups of non-Reading First schools—labeled Comparison Group schools and Reading First Eligible schools—and to All Elementary schools. In the Year 3 report, the Reading First Eligible schools were dropped as a comparison group because of their demographic dissimilarity with the YIP 4 schools. In this Year 4 report we are also dropping the so-called Comparison Group which, although comparable demographically to Reading First schools, is contaminated to an unknown degree by programs and educational practices that are similar to Reading First. We retain the All Elementary schools group (minus Reading First schools) in order to provide an overview of the rest of the state and show how it has been trending since 2002. However, it is emphasized that the All Elementary schools group is so demographically dissimilar to the Reading First group that no valid conclusions about program effects can be drawn from comparing them.

In the trend-line charts presented later in this chapter (e.g., Figure 4.1.1), the All Elementary schools group (which has a much higher starting point than the Reading First schools) is adjusted to have the same starting point as the Reading First schools so that their trend-lines can more conveniently be compared.

What are High Implementation and Low Implementation Reading First schools?

One of the features of this evaluation is that Reading First is studied not only in terms of student achievement but also in terms of program implementation at the school level. Chapter 3 describes the teacher, coach, and principal surveys that were administered to all Reading First schools and used to compute a Reading First Implementation Index (RFII) statistic for each school that responded. The RFII is intended to measure the degree to which the teachers, coaches, and principals are implementing the Reading First program in their school. RFII measures have been computed for 2004, 2005, and 2006, based on survey administrations in the spring of each year. Therefore, there are three years of RFII scores for YIPs 4 and 3, and two for YIP 2.

The RFII was used to divide Reading First schools into two groups labeled High Implementation schools and Low Implementation schools. The method for categorizing schools into these two groups was as follows: For every Reading First school in YIPs 4 and 3 the 2004, 2005, and 2006 RFII scores are averaged. For a YIP 2 school, 2005 and 2006 RFII scores are averaged. The mean RFII for all schools

(across all three YIPs) was computed in 2005 and found equal to 36. This is our cut-point. If a school's RFII falls at or above 36 points, it is classified as a High Implementation school, otherwise as a Low Implementation school. It is expected that this cut-point will remain the same in future years and across YIPs to maintain longitudinal comparability and to ensure that "high implementation" has a uniform meaning in all cases. In the tables below, the High Implementation achievement gains are those for schools with a mean RFII higher than 36. The Low Implementation gains are for schools with a mean RFII less than 36.

What is the Statistical Control Group?

As discussed in Chapters 2 and 3, the statistical control group is defined using regression models (such as those presented in Tables 4.0.1 and 4.0.2 below) to create a hypothetical sample of schools that has the same demographic characteristics as the Reading First YIP under consideration (or of the whole Reading First population, depending on the comparison) but that does not participate in Reading First. A school is said to be a "comparable non-Reading First" school if it receives an average yearly RFII of 25 or less for reasons outlined in Chapter 3. Therefore, 25 is entered into the regression equation to create an expected 2006 achievement score and gain score for the statistical control group.

The 2006 achievement score for the statistical control group is calculated as follows:

1. *Define a dependent variable*. The dependent variable is a school's 2006 score on the achievement metric under consideration. An example might be: "Grade 3 CST, Percent Below Basic and Far Below Basic." Another might be: "Grade 3 CAT/6, Spelling MeanPR."
2. *Compute a regression equation*. A regression equation is computed to model that dependent variable given a small set of predictor variables. These predictor variables are: starting point (a school's achievement score on that metric in the baseline year), number of years in the program (YIP), the school's average yearly RFII to date, its percent of Socio-Economically Disadvantaged (SED) students, and its percent of English Learners (ELs). Due to what is known as collinearity (excessive correlation) between EL status and SED status, the regression will allow only one variable or the other into the final equation, whichever has the strongest effect on that dependent variable. The output is a linear equation that predicts the value of the dependent variable (the 2006 achievement score) given a set of inputs for each of the predictor variables.
3. *Calculate inputs*. We now define our Reading First sample, for example all YIP 4 schools. Looking only at the YIP 4 schools, we calculate their average starting point and their average percent of SED students (or EL students, whichever is called for in that regression equation). We already know their number of years in the program (YIP = 4). We also know the theoretical

average yearly RFII of a non-Reading First school (Mean RFII = 25). These become inputs for the regression equation.

4. *Calculate a 2006 prediction*. Plugging the inputs into the regression equation, we get a predicted 2006 achievement value on the specified metric, representing the value a school would be expected to receive if it had the same characteristics as the Reading First schools in that YIP but had an average yearly implementation of 25. This predicted value, as well as the starting point, is graphed on the trend-line charts presented later in this chapter (e.g., Figure 4.1.1). Note that there are values only for the baseline year and 2006.
5. *Calculate a gain score*. We subtract from the predicted 2006 achievement value the mean starting point for the schools in the YIP to obtain the gain score for the "Statistical Control Group." This gain score may be interpreted as the difference in scores between the starting point of the hypothetical non-Reading First school, which is defined to be equal to the average starting score of the rest of the schools in the YIP to make it maximally comparable, and its predicted end-point. This difference is pasted into the gains table.
6. *Repeat for each achievement metric*. A new regression equation is computed for each achievement metric and new inputs are calculated for each YIP to generate the necessary gains predictions that are listed under Statistical Control Group in each table.

Achievement Results

Regression Effect Sizes

As mentioned, we perform a separate regression analysis for every achievement variable in order to calculate gains for the statistical control group. To summarize the effect sizes of the variables that predict Reading First achievement, we present regression tables for the Grade 2 CSTs (percent Proficient and above) and the RFAI, which is a composite of all the Reading First achievement metrics. In each case, the dependent variable to be predicted is the 2006 score for the relevant achievement variable. The predictor variables are: the school's baseline starting point[18], its number of years in the program, its average yearly RFII, its percent of Socio-Economically Disadvantaged (SED) students, and its percent of English Learners (ELs). In a separate regression we compute an effect size for a composite variable which is the school's average yearly RFII multiplied by its number of years in the program, equal to the sum of its RFII statistics over time. This effect is presented with the others.

Tables 4.0.1 – 4.0.2 summarize our findings in terms of the Grade 2 CST Proficient and Above and the RFAI regression models. Our regression model yields predictor variables ($p < 0.05$) that account for 33% of the total variance in 2006 Grade 2 CST scores (percent "Proficient" and above) and 63% of the total variance in the 2006 RFAI scores, a composite of all grades and metrics. Similar effect sizes are found across the achievement metrics except for the Grade 3 CSTs, for which the Number of Years in Program verges on being a statistically non-significant predictor. This is a result of the statewide negative Grade 3 CST trends between 2003 and 2004, generally regarded as an artifact of the test and not of an actual decline in the student population's performance.

Effect of Measurement Error

It is worth noting that the magnitudes of the standardized beta coefficients reported in the tables below (which quantify the "effect" of that variable in predicting achievement as described below) are to some unknown degree inaccurate due to measurement error. This is true both of the achievement tests and especially of the self-report implementation survey instrument. In this study, if the survey fails for whatever reason to measure true Reading First implementation in the classroom and school, the Average Yearly RFII effect will approach zero, even if the school is in fact implementing the program vigorously. There exist statistical corrections that can be applied to regression coefficients to correct for measurement error, but they are complex and have not been attempted here (Osborne, 2003).

[18] Starting point is defined differently for the CSTs and the RFAI. The CST starting point is the score immediately previous to the first year of implementation. The RFAI starting point is the score at the end of the first year of implementation.

Table 4.0.1: Effect Size of Variables Predicting Percent of Students Proficient and above on Grade 2 CST's in 2006 ($R^2 = 0.33$)[1]

Predictor Variable	*Standardized BetaCoefficient Effect (standard deviation units)*	*t-test (t > 1.96 implies significance with 95% confidence)*	*Probability the Effect is by Chance*
Starting CST score in year preceding first year in program	0.46	14.7	0.0000
Number of Years in Program	0.15	5.3	0.0000
Average Yearly RFII	**0.14**	4.6	0.0000
Yearly RFII * Years in Program[2]	**0.21**	7.1	0.0000
Percent of ELs in School	-0.13	-4.0	0.0000

1. The "Standardized Beta Coefficient" column shows how much the CST percent Proficient and above increases for every unit increase of that predictor variable. The "t-test" column shows how many times bigger the effect is than what would be predicted by chance. The "Probability" column gives the probability that the effect occurred by chance.

2. The "Yearly RFII * Years in Program" predictor variable is the product of a school's Average Yearly RFII and Number of Years in Program (equal to the sum of its RFII statistics over time). To avoid collinearity, its effect size was computed in a separate regression run in which Number of Years in Program and School's Average Yearly RFII were removed.

Table 4.0.2: Effect Size of Variables Predicting School RFAI in 2006 ($R^2 = 0.63$)[1]

Predictor Variable	*Standardized Beta Coefficient Effect (standard deviation units)*	*t-test (t > 1.96 implies significance with 95% confidence)*	*Probability the Effect is by Chance*
Starting RFAI in first year in program	0.72	32.0	0.0000
Number of Years in Program	0.15	7.0	0.0000
Average Yearly RFII	**0.09**	4.0	0.0000
Yearly RFII * Years in Program[2]	**0.17**	7.9	0.0000
Percent of SEDs in School	-0.09	-4.2	0.0001

The Non-Implementation Predictor Variables

As predictors in a multiple regression, each standardized beta coefficient gives the effect size of the variable that *would* occur if all the other variables were held constant. Thus, we see that the implementation variables are significant, taking into account the EL and SED status of the school.

For both tables, the strongest predictor of 2006 achievement is the school's starting point, its achievement level in the relevant baseline year. This is to be expected and has no special significance other than to convert the 2006 achievement score into the equivalent of a gain score. The remaining predictor variables are thus conceptually equivalent to predictors of gain scores.

These models support the conclusion that the effect of program implementation is to a large degree statistically independent of school demographic and language makeup. This is another way of saying that

the Reading First program effect is not biased against English Learners. The effect, as measured by the standardized beta coefficient, is approximately the same regardless of the number of English Learners that receive instruction in that school. This is discussed in greater detail towards the end of the chapter.

The Implementation Predictor Variables (Average Yearly RFII and Years in Program)

There are two predictor variables in these tables that relate to Reading First implementation. The first, Average Yearly RFII, relates to the average level of implementation that occurs in a school per year. Based on the implementation survey results, it is strictly a measure of Reading First implementation and nothing else.

The second predictor variable relating to Reading First implementation is Number of Years in Program. The longer a school is in Reading First, the more it has implemented the program. The difficulty with this variable is that it is potentially sensitive to non-Reading First factors that may influence achievement during its time in the program. These factors could include:

- Non-Reading First reading program elements in the schools (though these have been largely excluded in high implementing Reading First schools), as well as non-reading specific programs such as participation in HP or SAIT.
- Artifacts in the achievement tests, such as may have occurred with the Grade 3 CSTs
- Changing demographics in the state
- Some as yet unknown factor causing California students to become better at taking tests

The influence of these factors is unknown, a common and well-know affliction of trend analysis. Therefore, assuming minimal influence by non-Reading First trend effects, the *total* Reading First implementation effect is properly defined as its Average Yearly RFII multiplied by its Number of Years in Program. That means the Reading First total implementation effect is in the neighborhood of 0.17 to 0.21 standard deviations based on the tables above. That is equivalent to saying that for every standard deviation unit increase in total Reading First implementation, Grade 2 CST achievement increases by as much as 0.21 standard deviation units, or roughly one-fifth.

In real terms, this 0.21 effect means that a Reading First school with average or above average program implementation can, after four years, expect to see an increase of 3-7 percentage points in the percent of students scoring "Proficient" and above as compared to a comparable non-implementing school (i.e., a school with an average yearly RFII of 25) over the same period. This effect manifests in Figure 4.1.1 (on p. 77) as the increasing gap between the diverging trend-lines. Adding the effect of years in the program, an average or above average Reading First school that has been in the program for 4 years can expect to increase its percent of "Proficient" students by 8 to 11 percentage points above the score of a comparable

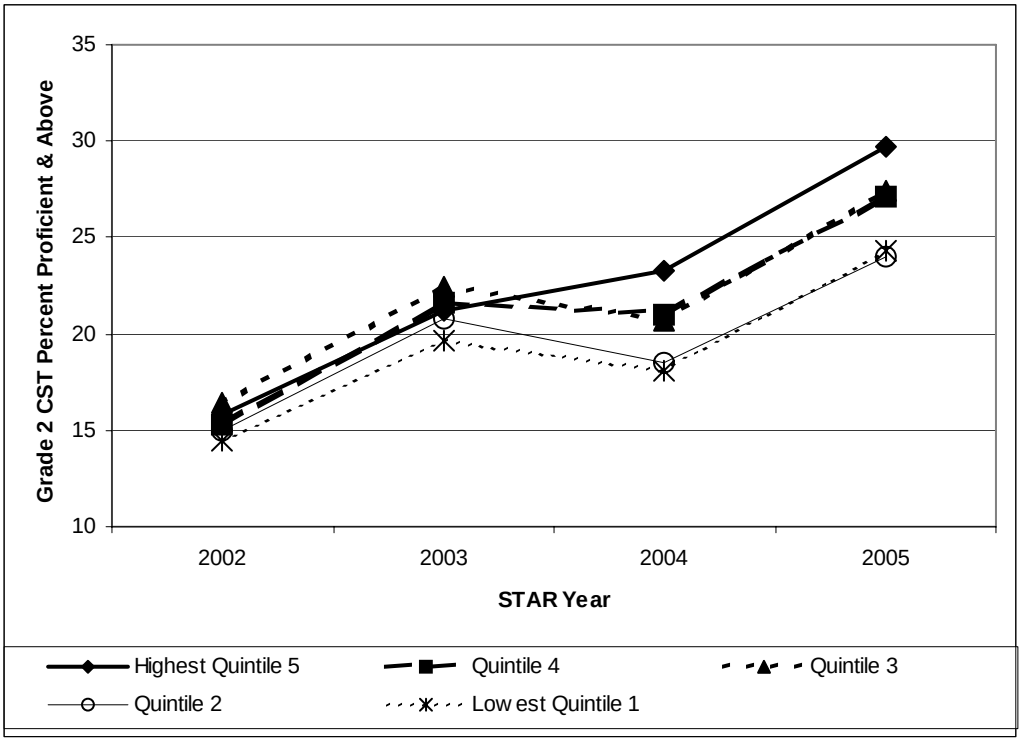
non-implementing school (RFII = 25). Thus, a 3 percentage point effect (relative to the statistical control group) is the minimum that can be expected if we assume that the effect of years in the program is solely an artifact of non-Reading First factors influencing the trend, and discount it accordingly. If we assume that the effect of years in the program is a consequence of Reading First, then the average Reading First effect is around 8 percentage points. If the school is a high-implementer, it can expect an effect of 11 percentage points.

Figure 4.0.1 is a trend-line chart sorted by implementation quintiles. Schools are sorted by RFII, divided into quintiles, and the average achievement of each quintile is graphed. The chart displays graphically both the years in program effect and the average implementation effect. The higher the implementation level of the school, the steeper its trend-line. Thus, the various implementation quintiles "fan" out over time in terms of their achievement levels. This "fan" pattern is an effect solely of average yearly implementation and in itself shows the efficacy of Reading First, independent of trend effects. It would occur even if there were no overall positive trend. We see that after 4 years the difference between the lowest implementation quintiles and the middle quintiles is around 3 percentage points; between the lowest quintiles and the highest quintiles it is around 6 percentage points.

In addition to the fan pattern, we see an overall positive slope of the trend-lines. The lowest quintiles grow around 9 percentage points, the highest quintiles by around 15 percentage points. This overall positive effect is the consequence of years in the program. If we can assume that there are no extraneous non-Reading First factors influencing the years in program effect, then around 11 percentage points of growth can be attributed to the combination of the implementation effect and the years in program effect, which is around 75% of the total 15 percentage point gain experienced by the schools in the highest quintiles.

Given these trend-lines and their statistical underpinnings, we have strong evidence that the Reading First program, when implemented with fidelity, has a positive impact on student achievement.

Figure 4.0.1: YIP 4 Grade 2 CSTs, 2002 to 2006, Percent Proficient and Above, by Implementation Quintiles



Should "Years in Program" be included in the implementation effect if all schools are trending upward?

When Average Yearly RFII is multiplied by Number of Years in Program, we see that the total implementation effect is 0.21 (consistent with the effects reported in the Year 3 Report for the Grade 2 CSTs), roughly double the effect of the Average Yearly RFII alone. We argue that the 0.21 effect size is the more valid estimate of the total Reading First effect, but it requires us to assume that the Years in Program effect is sensitive mainly to the school's implementation of Reading First and that it is independent of non-Reading First effects on achievement. This is a strong assumption, but not unreasonable. That is because Reading First schools agree not to implement competing programs or initiatives that are not aligned with Reading First. Because the achievement gains hold up, and are in fact higher, in schools that have high RFIIs and thus are implementing Reading First more exclusively, we conclude that the total Reading First gain scores since the start of the program are *not* an artifact of a statewide non-Reading First trend effect but are primarily a Reading First implementation effect.

These considerations lead to an interesting question. If the achievement gains experienced by Reading First schools over a four-year period are primarily a Reading First effect, as we suggest, then why do non-Reading First schools also show substantial gains over the same period of time? Assuming that the gains are real, we see two possibilities:

1. Non-Reading First schools have, over the same period, begun implementing non-Reading First educational strategies that happen to be effective; or
2. Non-Reading First schools have been implementing some or all of the same program elements that make Reading First effective.

A review of state educational initiatives supports the second possibility. The state, in January of 2002, adopted two "approved" reading curricula for K-8 schools to use. These are the same Houghton-Mifflin and Open Court reading programs used in Reading First. Schools that adopt these programs, have access to AB 466 teacher professional development, AB 75 principal professional development, and the 6-8 week skills assessments. In addition, many LEAs and schools have opted to hire reading coaches at their own expense. Such non-Reading First schools become virtually indistinguishable from Reading First schools in terms of educational practices in the classroom. The main difference is that the non-Reading First schools must use other funding sources to hire reading coaches and provide professional development.

Therefore, we believe that the statewide trend is fundamentally an effect of the same educational practices and program elements that are required by Reading First. To prove this hypothesis, a separate non-Reading First implementation study is necessary. This would mean administering a version of the Reading First implementation survey to a representative sample of non-Reading First schools.

Is the implementation effect as strong in the Year 4 Report as it was in the Year 3 Report?

The answer is no. A review of the Year 4 Report trend-lines and regression coefficients, including the quintile chart above (Figure 4.0.1), shows that the implementation effect is not as pronounced in Year 4 (Educational Data Systems, November 2003). The gap between high- and low-implementing schools, while substantial and significant in the Year 4 Report, is somewhat narrower than in Year 3. The explanation for this lies in part with the implementation measures. The implementation level of a school is defined to be the average of its RFII's across survey administrations. To the degree that statistical noise, or error, enters into these average RFII measures, the trend-lines become more entangled. In the case of the Year 4 Report, the increase in precision gained by averaging a school's RFII statistics over time is somewhat lost by increased variation in the RFII's over time. The effect may be impacted by schools that increased their implementation levels significantly since entering the program. It is also possible that the surveys, after being re-administered three years in a row, are losing some of their discriminating power. These issues need to be tracked carefully in future reports.

YIP 4 Tabular Results

Tables 4.1.1 through 4.1.3 provide achievement gains for YIP 4 Reading First Schools on STAR (CSTs and CAT/6), the EOY test, and the RFAI.

STAR Results

The CSTs and CAT/6 metrics show:

- In Reading First schools for Grade 2, the percentage of students scoring Proficient and above increased by 16 percentage points over four years. Grade 3 shows much smaller gains of 4-5 percentage points due to an anomalous counter-trend from 2002-2005. The Grade 3 CST metric is an example where the years in program effect is relatively weak but the average yearly implementation effect remains strong.
- The percentage of students scoring Below Basic and Far Below Basic decreased by the same amount in Grade 2. Grade 3 also shows large movement out of the bottom categories. Reading First is equally effective across the performance continuum.
- CAT/6 shows that for Reading and Language, Reading First schools rose approximately 4 percentiles relative to a nationally-normed population. For Spelling they rose 11 percentiles.
- The CST's and CAT/6 agree in terms of direction of the overall trend, but the magnitudes are different. There are many psychometric factors that could explain the differences in magnitudes, but they are subtle and beyond the scope of this study. We are forced to study each metric's trends relative only to itself.
- Reading First schools show dramatic gains in the Spelling section of the CAT/6.
- The Reading First schools show larger gains than the statistical control group and the average of All Elementary Schools (non-Reading First). Note that the All Elementary Schools trend-line has been adjusted downward to have the same starting point as the Reading First schools to allow visual comparison.
- High Implementing Reading First schools show larger gains than Low Implementing Reading First schools.

RFAI Results

Table 4.1.2 compares RFAI gains for High and Low Implementation YIP 4 schools. The pattern is consistent with the CST and CAT/6 Mean PR gains in Table 4.1.1. This is not surprising since the RFAI is a weighted combination of the CSTs and the CAT/6 MeanPR. However, it includes an EOY component which, given some anomalies in EOY Table 4.1.3 that are noted below, may narrow the difference between the High and Low Implementers.

- The 3-year gains are comparable in magnitude to the STAR gains, taking into account the more recent baseline year.

- Reading First schools show higher gains than the statistical control group.
- High implementers show higher gains than low implementers.

EOY Results

Table 4.1.3 presents 2-year and 1-year EOY gains for YIP 4 schools from 2004 to 2006. Even though three years of EOY gain data are available for YIP 4, we ignore the 2003 results as being less reliable due to incomplete data sets. The 2002-2003 academic year was the first year of Reading First implementation and the first year that the EOY test was administered. As a new test, its implementation was not mandatory. There were also typical start-up problems with data submission and data collection. School participation in the EOY assessment and data collection efforts improved in 2004, providing more reliable EOY numbers.

Table 4.1.3 shows that YIP 4 Reading First schools produced gains ranging from 6 to 9 percentage points across grades on the EOY test between 2004 and 2006. These are substantial gains for only two years, comparable in magnitude to the STAR results over four years. There is also a pattern of gains from 2005 to 2006, with the exception of Grade 3 Spanish. The Grade 3 Spanish mean gain scores are low as a result of one extreme negative gain for one school. Omitting that school results in a positive gain of 1.6. Another complicating factor is that students in Grade 3 waiver classrooms are tested in English rather than Spanish, unlike waiver students in earlier grades.

Aside from total gains, the pattern seen in the STAR scores of higher growth for High Implementation Reading First schools is not reproduced here; only Grade 2 shows High Implementation schools with higher gains. English-instruction students generally show smaller gains than waiver (Spanish-instruction) classrooms, which suggests that Reading First Spanish-instruction is effective for English Learners. In some cases the Spanish gains are very large, perhaps anomalous. While they may be the result of Reading First, the relatively small number of schools with waiver classrooms makes it hard to tell. The small N for waiver classrooms in Grade 3 (9 schools) is driven in part by the fact that most schools try to transition their EL students out of waiver classrooms into English-only classrooms before Grade 3.

What are we to conclude? On the one hand, these results fail to confirm that High Implementation schools grow faster than Low Implementation schools. On the other hand, if we accept the premise that growth across years is driven by Reading First, then Reading First has produced strong and consistent gains in the EOY measures from 2004 to 2006, especially in the waiver Spanish-instruction classrooms. Of the two effects, the High/Low Implementation effect is in general more subtle than the Years in Program effect, not surprising given the difficulty of measuring school implementation with a self-report survey. In this case it appears that the statistical fluctuations in the EOY data are too large to capture a self-reported implementation effect but not too large to obscure an overall positive growth effect.

It is important to bear in mind the limitations of the EOY tests as psychometric instruments. They do not have established reliability statistics and have not been scaled for difficulty, but appear to have content validity (Hasbrouck & Tindal, 2005). The Grades 1 – 3 Oral Fluency results are based on performance on a small number of reading passages that are not the same for everybody and have not been equated to control for difficulty. Assessments are conducted and scored by classroom teachers. While such testing practices are common for formative assessments such as the 6-8 week skill assessments and EOY tests, they lack the standardization of summative assessments such as the CST's and CAT/6.

Table 4.1.1: 4-Year Gain, 2002 to 2006 for YIP 4 Reading First and non-Reading First Schools

Years in Program: 4	Average School Gain Score				
	Reading First Schools[1, 2]				Non-Reading First Schools
	All Reading First Schools	High Implementation Schools[3,6]	Low Implementation School[3,6]	Statistical Control Group[4]	All Elementary Schools[5]
CST					
Grade 2	(N=267)	(N=162)	(N=105)	(N=N/A)	(N=4096)
% Proficient & Above	15.7[abc]	16.4 [ab]	14.6 [a]	11.8	13.6
% Basic	0.0[b]	-0.5[b]	0.7[b]	0.4	-5.9
% Below Basic & Far Below Basic	-15.7[abc]	-16.0[ab]	-15.3[ab]	-12.6	-7.7
Grade 3	(N=266)	(N=161)	(N=105)	(N=N/A)	(N=4105)
% Proficient & Above	4.3[abc]	5.2[ab]	2.9[b]	2.2	1.0
% Basic	6.8[bc]	7.4[ab]	6.0[b]	5.6	2.9
% Below Basic & Far Below Basic	-11.1[abc]	-12.5[ab]	-9.0[b]	-8.0	-3.9
CAT/6					
Grade 3 (Mean Percentile Rank)	(N=266)	(N=161)	(N=105)	(N=N/A)	(N=4105)
Reading	3.7[abc]	4.1[ab]	3.0[b]	2.5	0.7
Language	4.3[abc]	5.0[ab]	3.1	2.8	2.4
Spelling	10.9[abc]	12.4[ab]	8.6[ab]	6.6	4.3

[a] Significantly different ($p < 0.05$) relative to the Statistical Control Group.
[b] Significantly different ($p < 0.05$) relative to All Elementary Schools.
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., Gain = 0. Applied to the "All Reading First Schools" column only.

[1] The N of schools may be different between Grades 2 and 3 within the same cohort because not all schools have both grades. However, the N of schools within the same cohort for the Grade 3 CSTs and CAT/6 MeanPR match.

[2] N is always in reference to schools and not students.

[3] The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.

[4] Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First school.

[5] Elementary Schools' refers to all non-Reading First elementary schools in California.

[6] High Implementation schools are those with an RFII greater than the mean RFII for 2004, which was 36. Low Implementation schools have an RFII less than 36.

Table 4.1.2: Two Year RFAI Gain 2004 to 2006 YIP 4 Reading First Schools

Years in Program: 4	Reading First Schools			
	All Reading First Schools	High Implementation Schools[1]	Low Implementation Schools	Statistical Control Group[2]
RFAI Gain	(N=267) 8.1	(N=162) 8.2	(N=105) 7.9	(N=N/A) 7.2

[1] The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.

[2] Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First Schools.

Table 4.1.3: End-of-Year Fluency Test: Gains (2004-2006, 2005-2006) and Percent Meeting Benchmark 2006, YIP 4 Reading First Schools

	All Reading First Schools						High Implementation Schools						Low Implementation Schools					
	2004-2006 Gain		2005-2006 Gain		2006 % at Benchmark		2004-2006 Gain		2005-2006 Gain		2006 % at Benchmark		2004-2006 Gain		2005-2006 Gain		2006 % at Benchmark	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Kindergarten																		
English	249	9.0	251	6.2	255	85.6	152	8.7	153	5.7	157	86.0	97	9.6	98	7.1	98	85.0
Spanish	-	-	46	6.2	56	87.2	-	-	23	6.6	29	88.6	-	-	23	5.8	27	85.8
Grade 1																		
English	263	6.0	265	2.2	265	45.4	160	6.1	161	3.3	161	46.0	103	5.8	104	0.5	104	44.4
Spanish	-	-	50	6.4	59	54.9	-	-	26	2.4	32	54.1	-	-	24	10.6	27	56.0
Grade 2																		
English	264	7.3	267	1.4	267	45.7	160	6.9	162	0.2	162	45.7	104	7.8	105	3.2	105	45.8
Spanish	-	-	45	7.7	54	45.4	-	-	22	11.3	28	47.2	-	-	23	4.3	26	43.4
Grade 3																		
English	263	6.1	266	1.7	266	46.1	160	6.0	161	1.6	161	45.8	103	6.1	105	1.8	105	46.6
Spanish			9	-0.6	36	32.0	-	-	1	-20.0	15	32.6	-	-	8	1.9	21	31.7

YIP 4 Trend-lines

One disadvantage of studying only total gain scores is that they say nothing about the shape of the trend-lines across years. Neither do they reveal the absolute percentage of students in a given category at a given time. To provide a more complete picture of the historical performance of these schools, we present trend-line graphs for each YIP. Figure 4.1.1 displays the percentage of Proficient and above students for the Grade 2 CSTs from 2002 to 2006. Trend-lines are given for All Reading First schools, High Implementation schools, Low Implementation Reading First schools, the Statistical Control Group, and All Elementary schools. Figure 4.1.2 displays the same trend-lines for Grade 3. Figures 4.1.3, 4.1.4, and 4.1.5 track the school MeanPR for Reading, Language Arts, and Spelling for Grade 3 from 2002 through 2006.

The Grade 3 CST trends for Reading First schools in Figure 4.1.2 mimic other statewide CST trends (see http://star.cde.ca.gov/ for statewide STAR reports) and reflect no significant gain in percent Proficient and Above in Grade 3 through 2005. Even in a negatively trending environment, however, the High Implementation schools out-perform the Low Implementation schools, the Statistical Control Group, and All Elementary schools. The trend-lines reverse course and start heading upwards in 2004, and by 2006 the average direction of the trend relative to the 2002 starting point is positive. The High Implementation schools maintain and increase their lead relative to the other school trend-lines throughout this time-frame.

The CAT/6 MeanPR Reading, Language Arts and Spelling trend-lines, while not as dramatic as the Grade 2 CST trend-lines, tell the same story. Instead of a decline, the trend-lines for all groups are positive. High Implementation leads to higher gains than Low Implementation. In Spelling, the Reading First schools not only start out ahead but also increase their lead dramatically, revealing that Spelling is a notable strength of the approved Reading First curricula relative to national norms.

Figure 4.1.1: YIP 4 CSTs, 2002 to 2006, Percent Proficient and Above Trend-lines, Grade 2

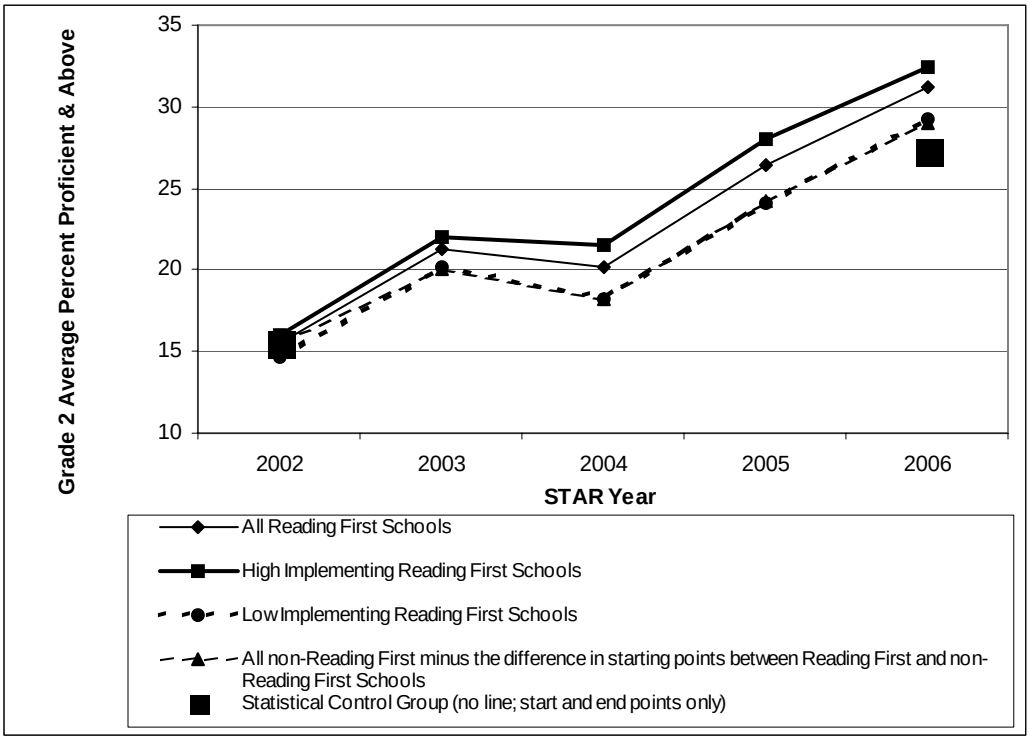


Figure 4.1.2: YIP 4 CSTs, 2002 to 2006, Percent Proficient and Above Trend-lines, Grade 3

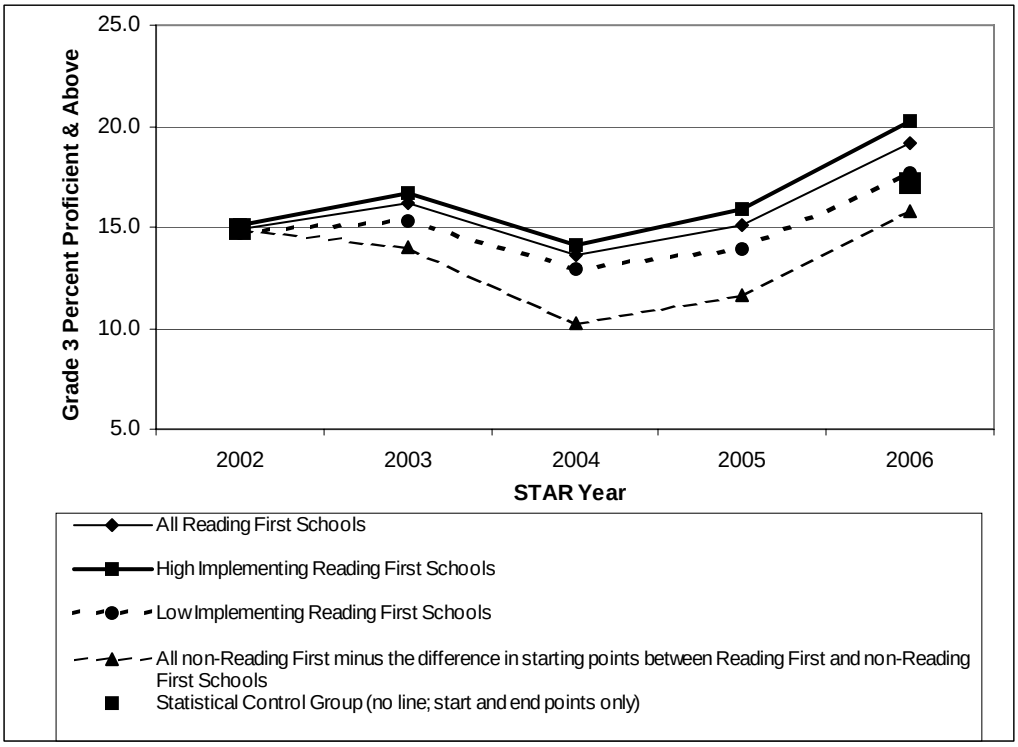


Figure 4.1.3: YIP 4 Mean Percentile Rank, CAT/6 Reading, 2002 to 2006, Grade 3

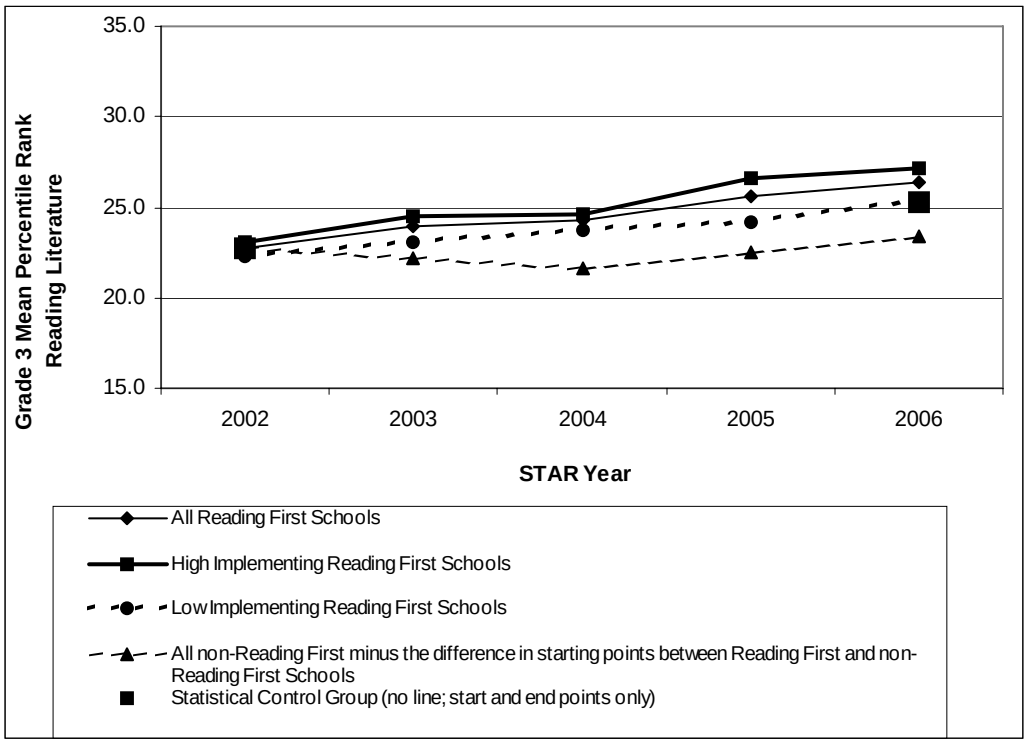


Figure 4.1.4: YIP 4 Mean Percentile Rank, CAT/6 Language Arts, 2002 to 2006, Grade 3

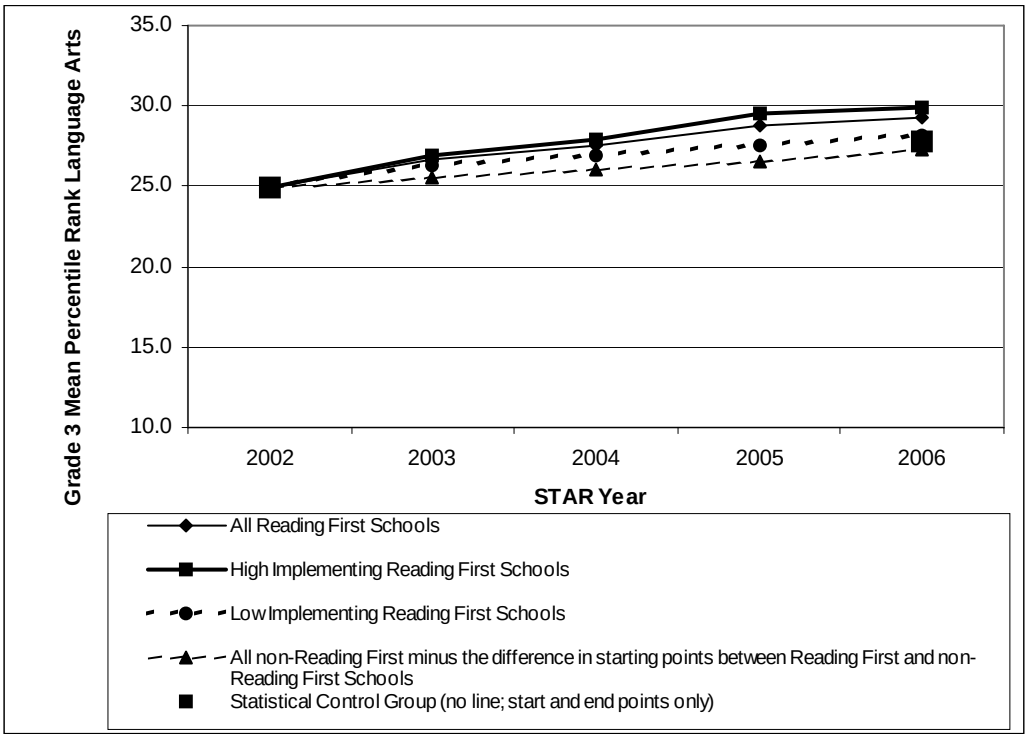
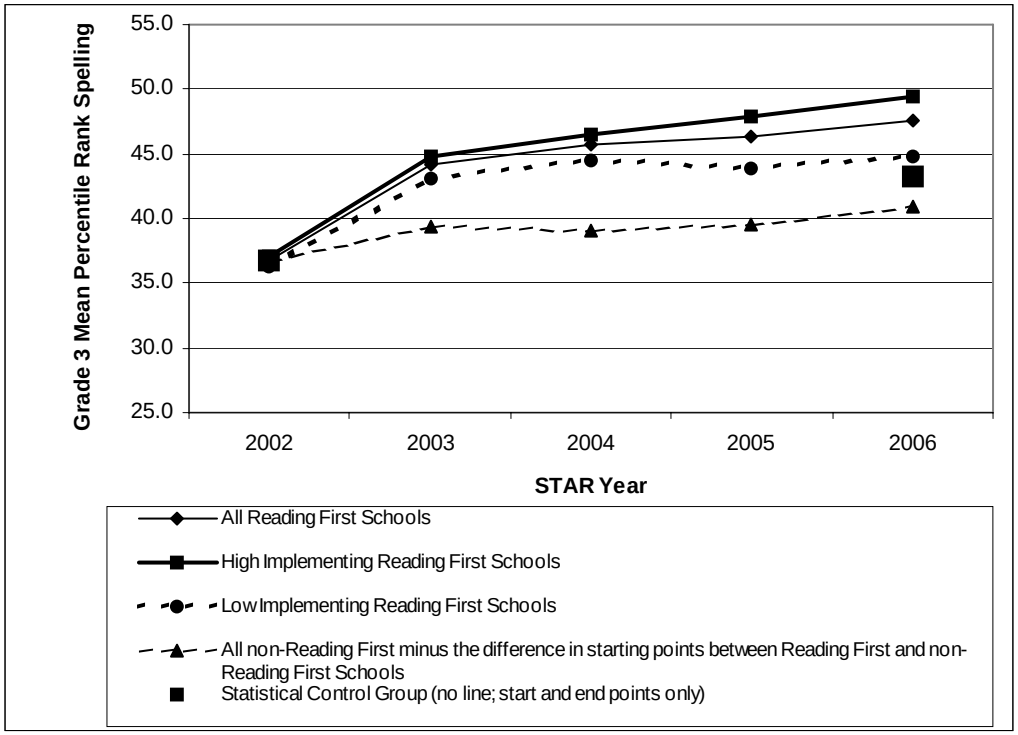


Figure 4.1.5: YIP 4 Mean Percentile Rank, CAT/6 Spelling, 2002 to 2006, Grade 3



YIP 3 Tabular Results and Trend-lines

Table 4.2.1 shows the STAR gains for YIP 3 Reading First schools. The gains tables are similar to those for YIP 4 and lead to the same conclusions, but lack the benefit of the extra year. The CAT/6 Spelling gains are not as remarkable as those for the YIP 4 schools.

Table 4.2.1: YIP 3 Schools 3-Year Gains, 2003 to 2006

Years in Program: 3	Average School Gain Score				
	Reading First Schools[1,2]				Non-Reading First Schools
	All Reading First Schools	High Implementation Schools[3,6]	Low Implementation Schools[3,6]	Statistical Control Group[4]	All Elementary Schools[5]
CST					
Grade 2	(N=372)	(N=224)	(N=148)	(N=N/A)	(N=4189)
% Proficient & Above	10.5[abc]	11.0[ab]	9.9[a]	7.7	8.7
% Basic	-2.0[bc]	-2.2[b]	-1.6[b]	-2.2	-5.9
% Below Basic & Far Below Basic	-8.6[abc]	-8.8[b]	-8.3[b]	-6.6	-2.8
Grade 3	(N=372)	(N=225)	(N=147)	(N=N/A)	(N=4195)
% Proficient & Above	3.7[bc]	4.1[ab]	3.2[b]	2.4	1.5
% Basic	5.1[abc]	5.0[b]	5.2[b]	4.0	1.2
% Below Basic & Far Below Basic	-8.8[abc]	-9.1[ab]	-8.3[b]	-6.7	-2.7
CAT/6					
Grade 3 (Mean Percentile Rank)	(N=372)	(N=225)	(N=147)	(N=N/A)	(N=4194)
Reading	3.7[abc]	4.0[ab]	3.1[b]	2.1	0.9
Language	4.5[abc]	4.9[ab]	3.8[ab]	2.5	1.5
Spelling	7.0[abc]	7.6[ab]	6.2[ab]	4.0	1.2

[a] Significantly different ($p < 0.05$) relative to the Statistical Control Group.
[b] Significantly different ($p < 0.05$) relative to All Elementary Schools.
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., Gain = 0. Applied to the "All Reading First Schools" column only.

[1] The N of schools may be different between Grades 2 and 3 within the same cohort because not all schools have both grades. However, the N of schools within the same cohort for the Grade 3 CSTs and CAT/6 MeanPR match.

[2] N is always in reference to schools and not students.

[3] The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.

[4] Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First school.

[5] Elementary Schools' refers to all non-Reading First elementary schools in California.

[6] High Implementation schools are those with an RFII greater than the mean RFII for 2004, which was 36. Low Implementation schools have an RFII less than 36.

Table 4.2.2: Two Year RFAI Gain 2004 to 2006 YIP 3 Reading First Schools

Years in Program: 3	Reading First Schools			
	All Reading First Schools	High Implementation Schools[1]	Low Implementation Schools	Statistical Control Group[2]
RFAI Gain	(N=377) 8.9	(N=228) 9.2	(N=149) 8.5	(N=N/A) 6.4

1. The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.

2. Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First Schools.

Figure 4.2.1: YIP 3 CSTs, 2003 to 2006, Percent Proficient and Above Trend-lines, Grade 2

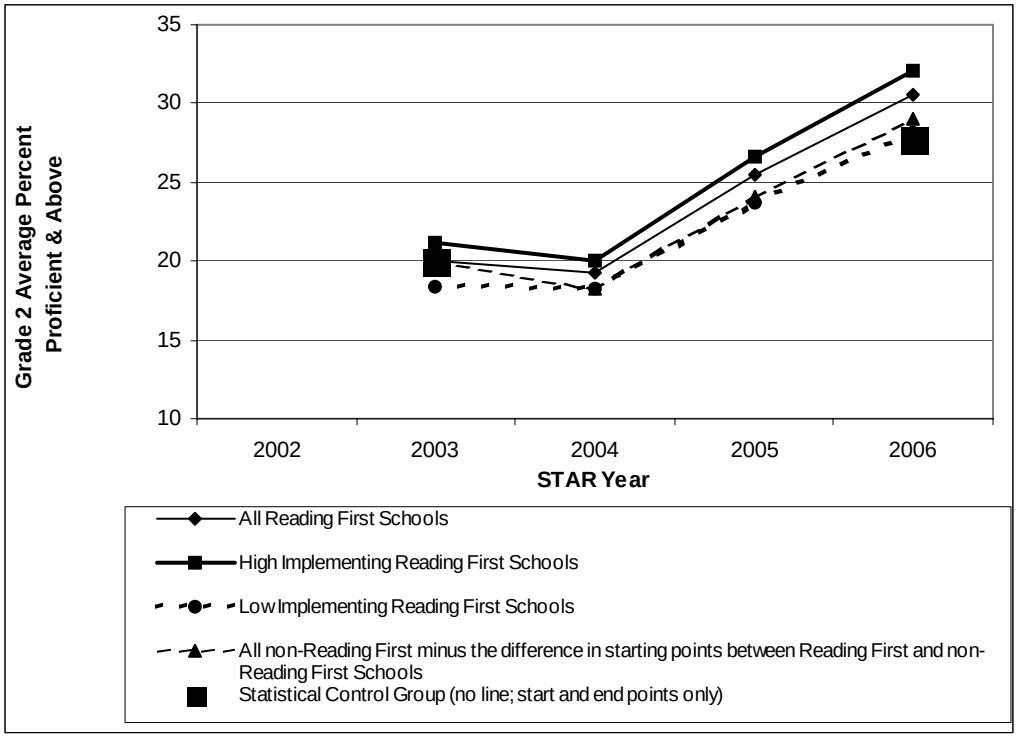


Figure 4.2.2: YIP 3 CSTs, 2003 to 2006, Percent Proficient and Above Trend-lines, Grade 3

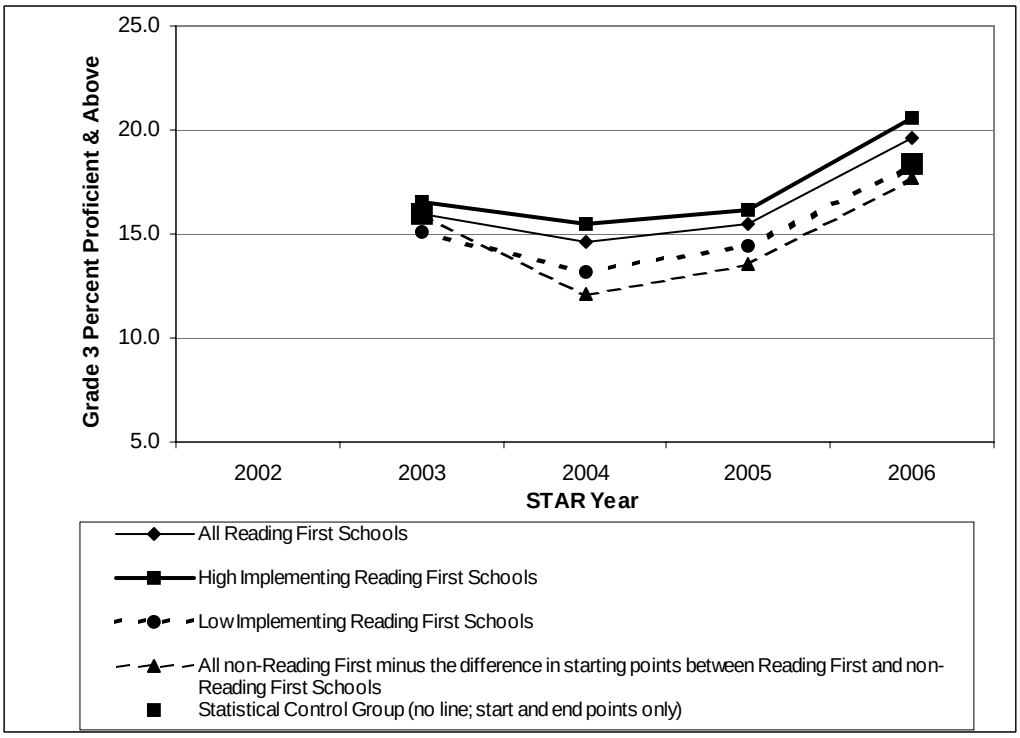


Figure 4.2.3: YIP 3 Mean Percentile Rank, CAT/6 Reading, 2003 to 2006, Grade 3

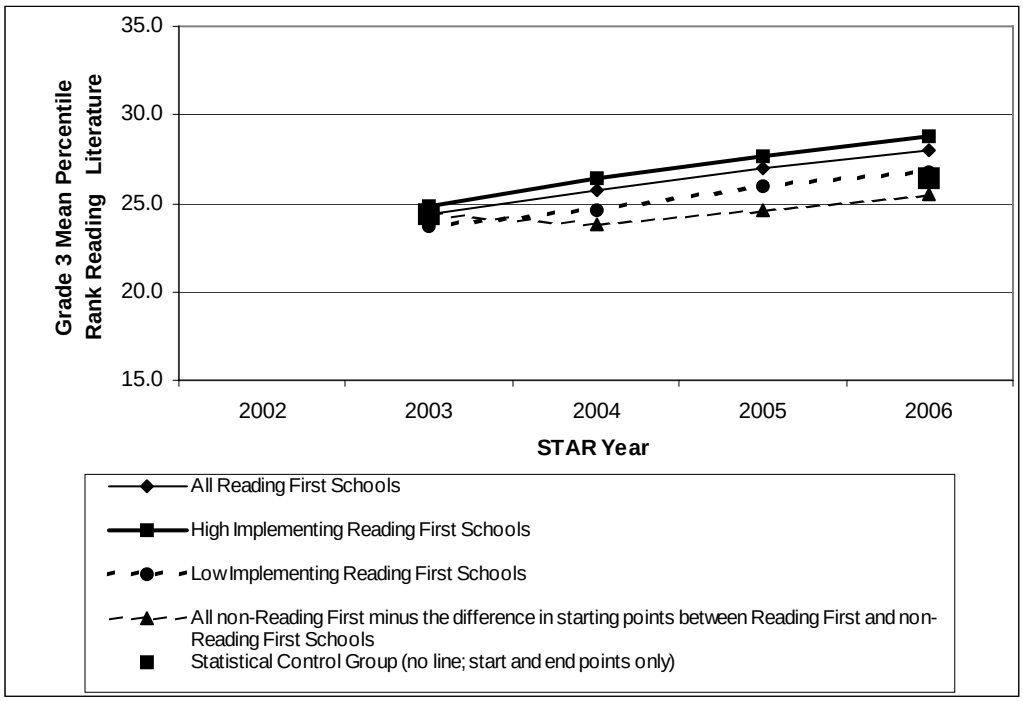


Figure 4.2.4: YIP 3 Mean Percentile Rank, CAT/6 Language Arts, 2003 to 2006, Grade 3

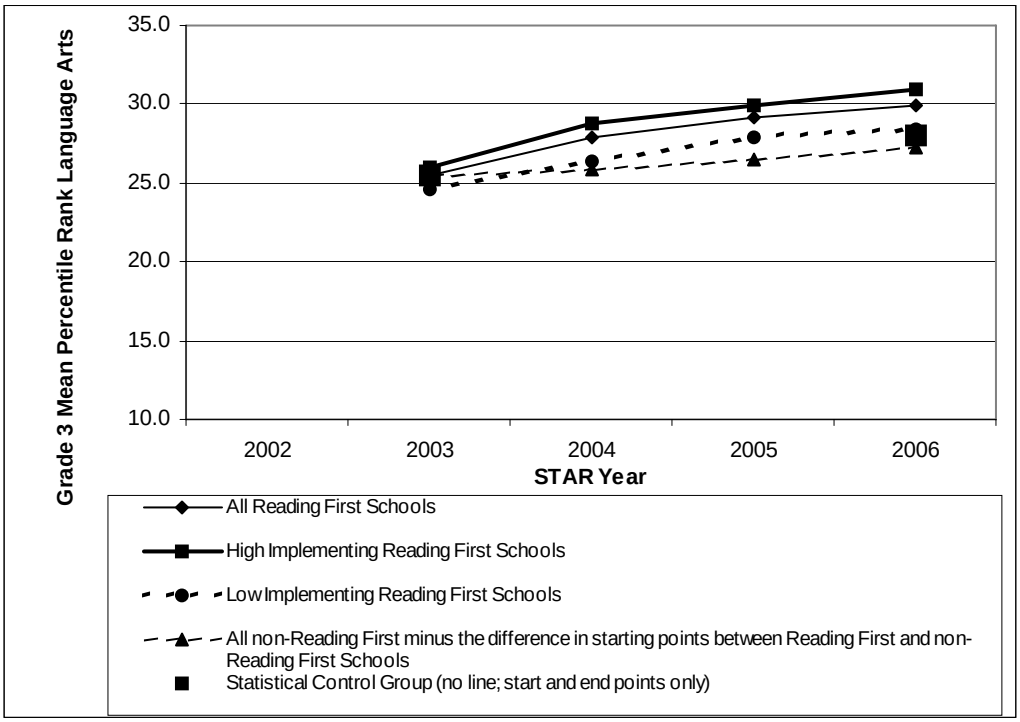
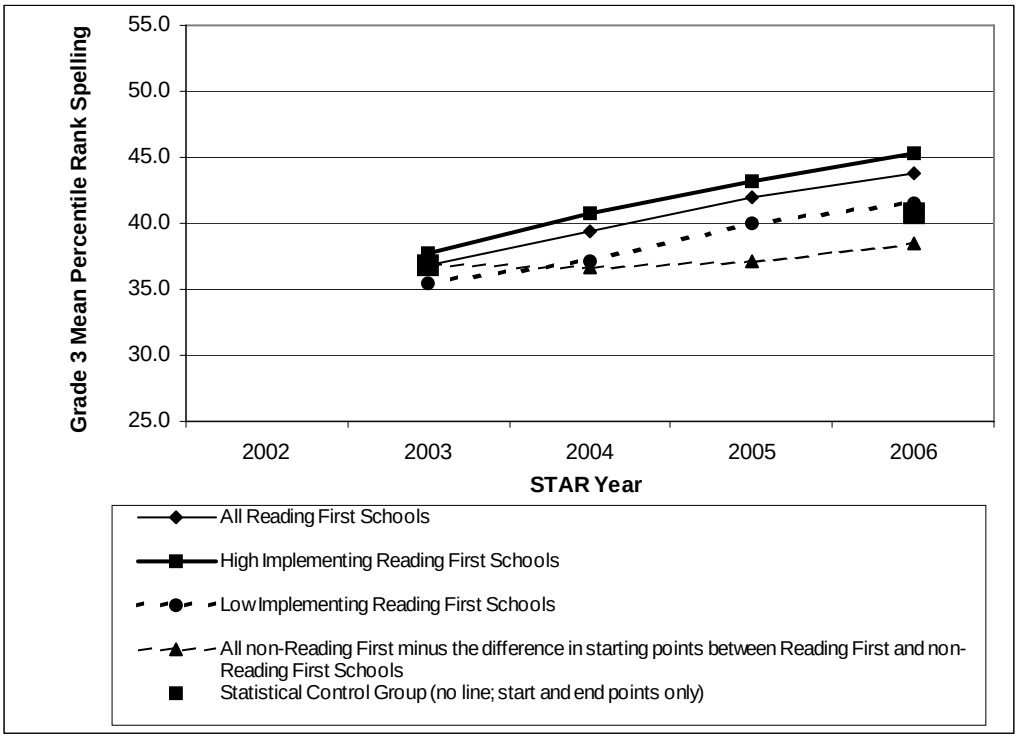


Figure 4.2.5: YIP 3 Mean Percentile Rank, CAT/6 Spelling, 2003 to 2006, Grade 3



YIP 2 Tabular Results and Trend-lines

The YIP 2 results and trend-lines are similar to those for YIPs 3 and 4 with some interesting exceptions:

- The Grade 2 CST gains are almost as large as those for YIP 4, despite taking place over a much shorter time period. YIP 2 schools appear to be growing much more quickly than YIP 4 and YIP 3 schools, a possible acceleration that should be studied closely.
- High Implementing schools have larger gains than Low Implementing schools for the Grade 2 and Grade 3 CST's, and the Grade 3 CAT/6 for Spelling.
- On the other hand, Low Implementing schools are growing faster than High Implementing schools on the Grade 3 CAT/6 Reading and Language metrics. They are also growing faster on the RFAI metric.

We do not have a complete explanation for why the Low Implementation schools post higher gains than High Implementation schools for two of the CAT/6 metrics and for the RFAI. This does represent a departure from the pattern shown for the Grade 2 metrics and for the other YIPs. However, while the CAT/6 trend-lines (Figures 4.3.1 to 4.3.3) reflect these different rates of growth, they also show that High Implementation schools consistently have higher percents of students that are Proficient and above. We also note that these are schools that have only been in the program for two years, and that the RFAI gain is only across one year, reducing the amount of time available for program effects to reveal themselves.

Table 4.3.1: YIP 2 Schools 2-Year Gains, 2004 to 2006

Years in Program: 2	Average School Gain Score				
	Reading First Schools[1,2]				Non-Reading First Schools
	All Reading First Schools	High Implementation Schools[3,6]	Low Implementation Schools[3,6]	Statistical Control Group[4]	All Elementary Schools[5]
CST					
Grade 2	(N=144)	(N=75)	(N=69)	(N=N/A)	(N=4253)
% Proficient & Above	12.0[ac]	13.0[ab]	11.0[a]	7.2	10.3
% Basic	0.4[b]	0.1[b]	0.6[b]	0.5	-4.6
% Below Basic & Far Below Basic	-12.3[abc]	-13.0[ab]	-11.5[ab]	-8.2	-5.7
Grade 3	(N=145)	(N=74)	(N=71)	(N=N/A)	(N=4262)
% Proficient & Above	6.7[abc]	6.9[a]	6.5	4.3	5.1
% Basic	5.1[abc]	5.4[ab]	4.7[b]	3.5	-0.3
% Below Basic & Far Below Basic	-11.6[abc]	-12.3[ab]	-11.0[ab]	-7.9	-4.8
CAT/6					
Grade 3 (Mean Percentile Rank)	(N=145)	(N=74)	(N=71)	(N=N/A)	(N=4257)
Reading	2.8[bc]	2.7[a]	2.9[b]	1.7	1.3
Language	3.3[abc]	3.3[ab]	3.4[ab]	1.7	0.9
Spelling	6.1[abc]	7.3[ab]	4.8[ab]	2.0	1.2

[a] Significantly different ($p < 0.05$) relative to the Statistical Control Group.
[b] Significantly different ($p < 0.05$) relative to All Elementary Schools.
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., Gain = 0. Applied to the "All Reading First Schools" column only.

[1] The N of schools may be different between Grades 2 and 3 within the same cohort because not all schools have both grades. However, the N of schools within the same cohort for the Grade 3 CSTs and CAT/6 MeanPR match.

[2] N is always in reference to schools and not students.

[3] The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.

[4] Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First school.

[5] Elementary Schools' refers to all non-Reading First elementary schools in California.

[6] High Implementation schools are those with an RFII greater than the mean RFII for 2004, which was 36. Low Implementation schools have an RFII less than 36.

Table 4.3.2: One Year RFAI Gain 2005 to 2006 YIP 2 Reading First Schools

Years in Program: 2	Reading First Schools			
	All Reading First Schools	High Implementation Schools[1]	Low Implementation Schools[1]	Statistical Control Group[2]
RFAI Gain	(N=151) 6.6	(N=78) 6.3	(N=73) 6.9	(N=N/A) 4.7

1. The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.

2. Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First Schools.

Figure 4.3.1: YIP 2 CSTs, 2004 to 2006, Percent Proficient and Above Trend-lines, Grade 2

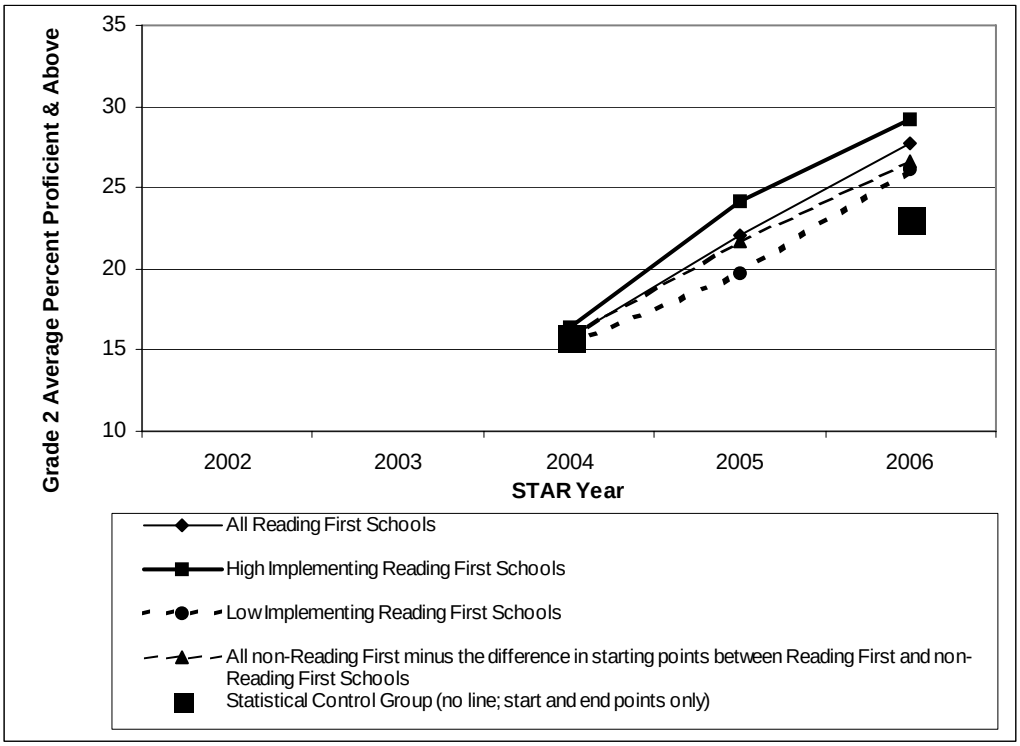


Figure 4.3.2: YIP 2 CSTs, 2004 to 2006, Percent Proficient and Above Trend-lines, Grade 3

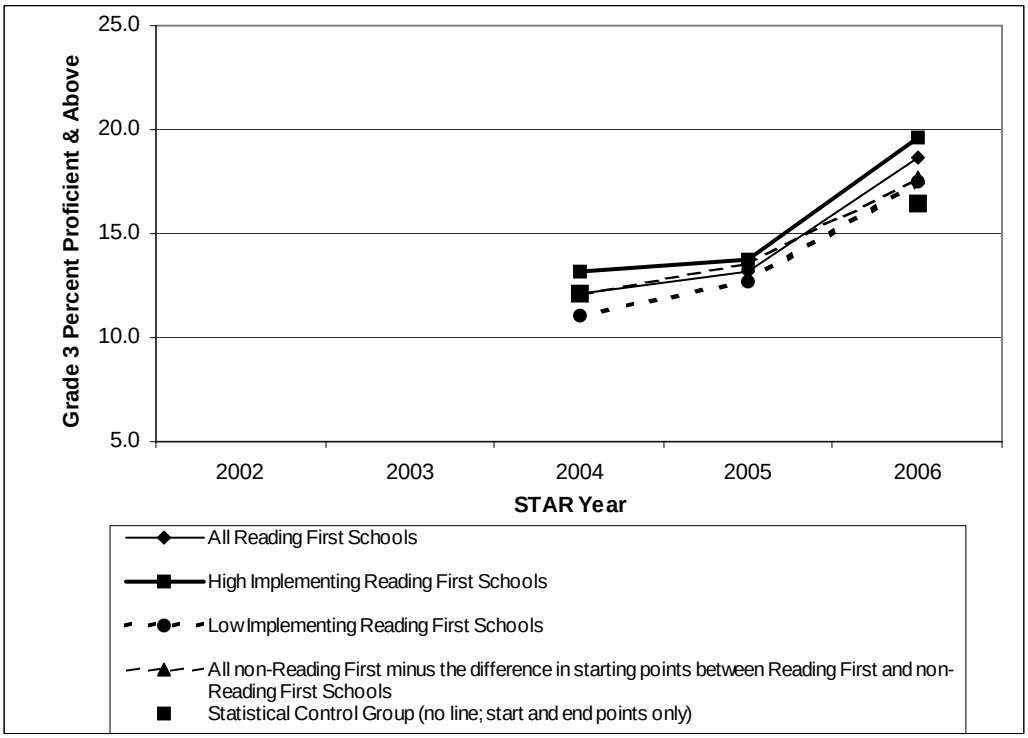


Figure 4.3.3: YIP 2 Mean Percentile Rank, CAT/6 Reading 2004 to 2006, Grade 3

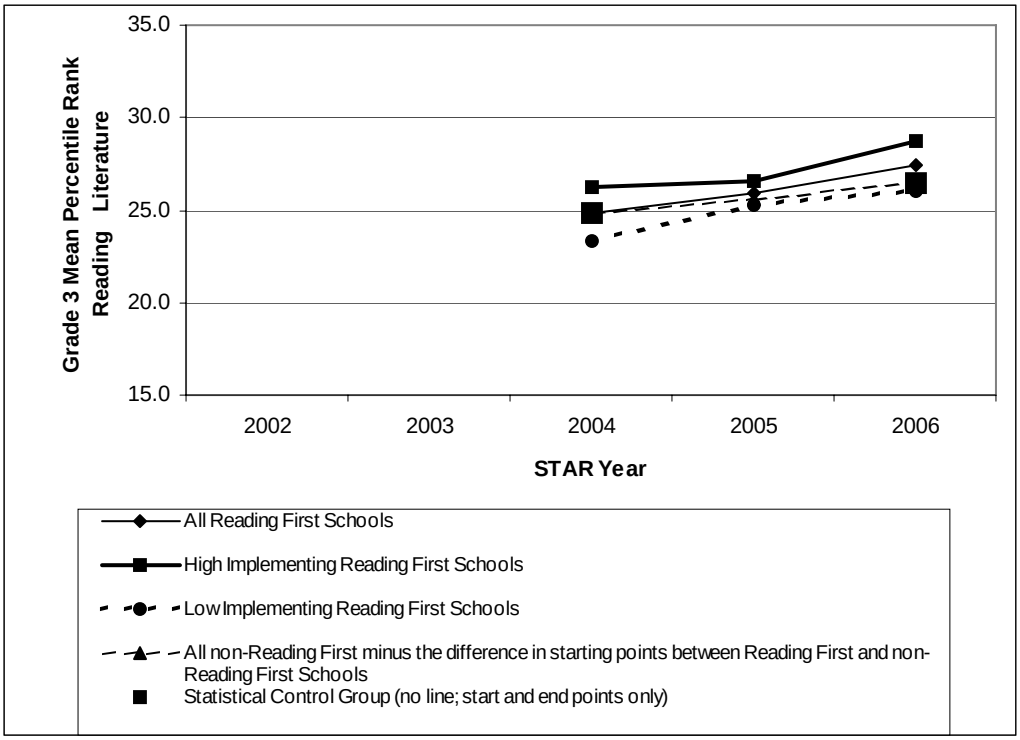


Figure 4.3.4: YIP 2 Mean Percentile Rank, CAT/6 Language Arts 2004 to 2006, Grade 3

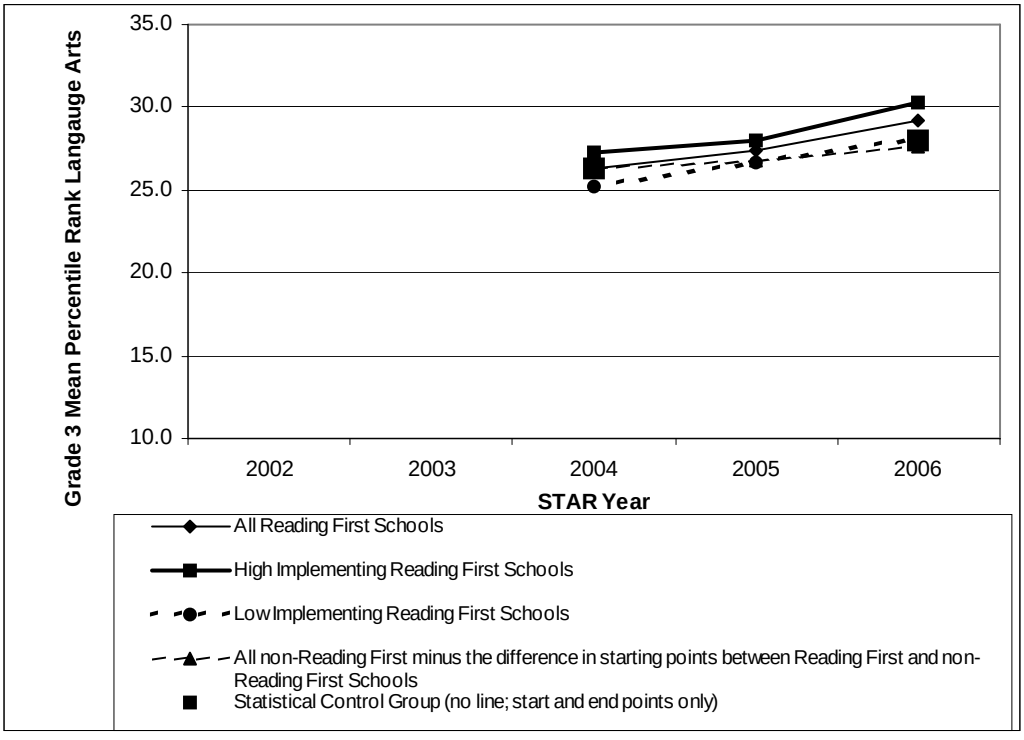
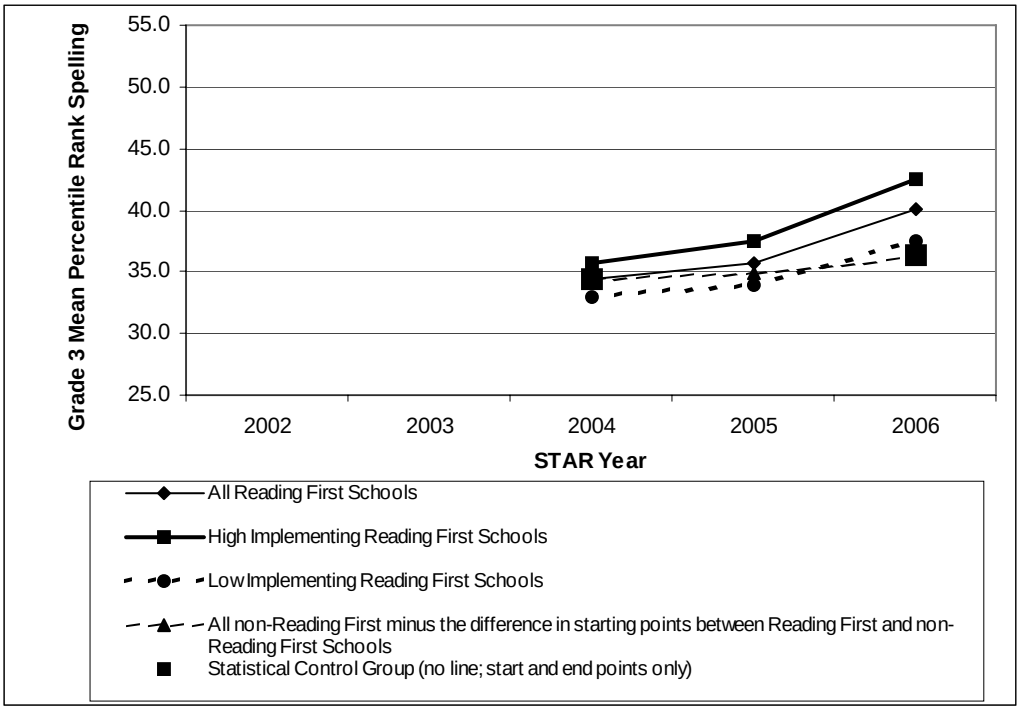


Figure 4.3.5: YIP 2 Mean Percentile Rank, CAT/6 Spelling 2004 to 2006, Grade 3



Does Reading First work with English Learners?

In an earlier section of this chapter, "Regression Effect Sizes," we noted that regression Table 4.0.1 provides statistical evidence that Reading First is not biased against schools with a high number of English Learners, that the English Learner effect on achievement is independent of the Reading First implementation effect. In this section, we study the effect of Reading First on schools with more than 50% English Learners in greater detail.

First, it is important to understand that our finding that Reading First is not biased against schools with a majority of English Learners is not meant to imply that they have the same overall level of achievement as schools with fewer English Learners. Schools with more English Learners typically have more challenges in teaching English reading skills than do schools with students for whom English is their primary language. This results in lower overall achievement scores in schools with high proportions of English Learners, which is why regression Table 4.0.1. shows that a school's percentage of English Learners is a negative predictor of 2006 Grade 2 CST achievement.

Our finding is based not on absolute achievement scores but on gain scores – the *rate* of growth. As shown in Table 4.4 and Figures 4.4.1 and 4.4.2 (which look at YIP 4 schools only), the gains and trend-lines for schools with a majority of English Learners are similar to those for the entire YIP 4 population as given in Table 4.1.1, and are even somewhat larger. The implementation effect is also present. High Implementation schools with a majority of English Learners outperform the Low Implementation schools. Thus, level of implementation of the program is just as important for schools with a majority of English Learners as for the overall Reading First population.

It is not yet known whether Reading First "waiver" classrooms (in which instruction is carried out in Spanish) are more effective than Reading First English-only classrooms (in which instruction is carried out in English) in moving students into "Proficient" and above. However, out of a total of 267 Reading First schools with grade 2 classrooms, 222 schools do not have active "waiver" classrooms, roughly 83%. It is likely that a similarly high proportion of schools within the 186 schools with a majority of English Learners also do not have active "waiver" classrooms, though this proportion has yet to be determined. That means that the Reading First effect presented in Table 4.4 and Figures 4.4.1 and 4.4.2 for schools with a majority of English Learners can be attributed largely to English-only classrooms. Subsequent research may find that "waiver" classrooms out-perform English-only classrooms (a hypothesis raised by EOY table 4.1.3) in teaching how to read English. However, based on the data gathered so far, it appears that the English-only Reading First classrooms are at least as adequate for schools with a majority of English Learners as they are for the Reading First population as a whole.

Table 4.4: YIP 4 Schools with >50% English Learners, CSTs, Grade 2, 2002 – 2006

Schools with more than 50% English Learners	Majority EL Reading First Schools (in the program 4 years)				
	All Reading First Schools	High Implementation Schools[1]	Low Implementation Schools[1]	Statistical Control Group[3]	All (>50% EL) Non-Reading First Elementary Schools[4]
Number of Schools (2006)[2]	186	104	82	N/A	683
Number of Students (2006)	24,370	12,775	11,595	N/A	70,376
% Proficient & Above					
2002	14.2	15.0	13.2	14.2	18.3
2006	30.6	32.0	28.7	27.0	33.2
Gain	16.4[6]	17.0	15.5	12.8	14.9
% Below Basic & Far Below Basic					
2002	55.8	54.4	57.6	55.8	51.1
2006	39.2	37.4	41.5	42.0	36.6
Gain	-16.6	-17.0	-16.1	-13.8	-14.5
Mean Performance Level Per Student[5]					
2002	2.35	2.39	2.30	2.35	2.48
2006	2.83	2.88	2.78	2.75	2.92
Gain	0.49	0.49	0.48	0.40	0.44

[1] High Implementation schools are those with an RFII greater than the mean RFII for 2004, which was 36. Low Implementation schools have an RFII less than 36.

[2] The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools. Schools may be missing an RFII because they did not submit surveys.

[3] Calculated using multiple regression, these statistics are the expected gain of schools with the same starting point, number of years in program, and percent of English Learners and/or percent SED students as the Reading First schools profiled in this table, with the exception that the RFII is specified to equal 25, the expected RFII of a non-implementing Reading First school.

[4] Elementary Schools' refers to all non-Reading First elementary schools in California.

[5] Mean Performance Level is on a scale from 1 to 5, where: 1 = Far Below Basic, 2 = Below Basic, 3 = Basic, 4 = Proficient, 5 = Advanced.

[6] The gain statistics do not exactly match the difference between the 2002 and 2006 means because they are calculated differently, as the mean of individual school gains rather than as a difference between the means of the 2002 and 2006 school populations.

Figure 4.4.1.: YIP 4 Schools with Majority of English Learners, CSTs, Grade 2, 2002 – 2006, Percent of Students Proficient & Advanced

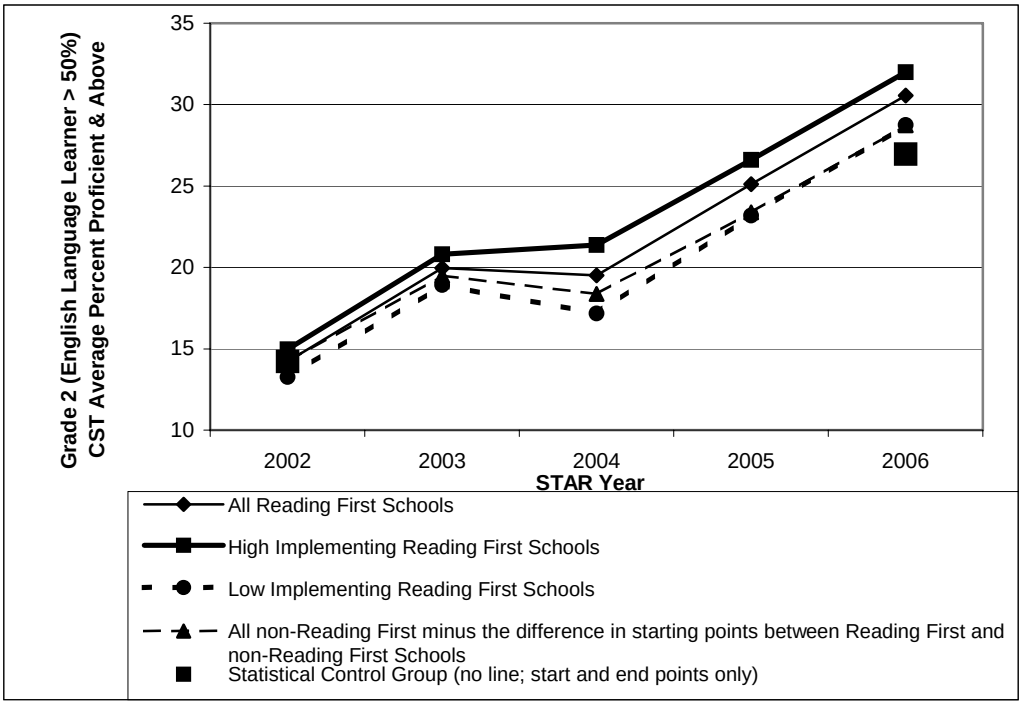
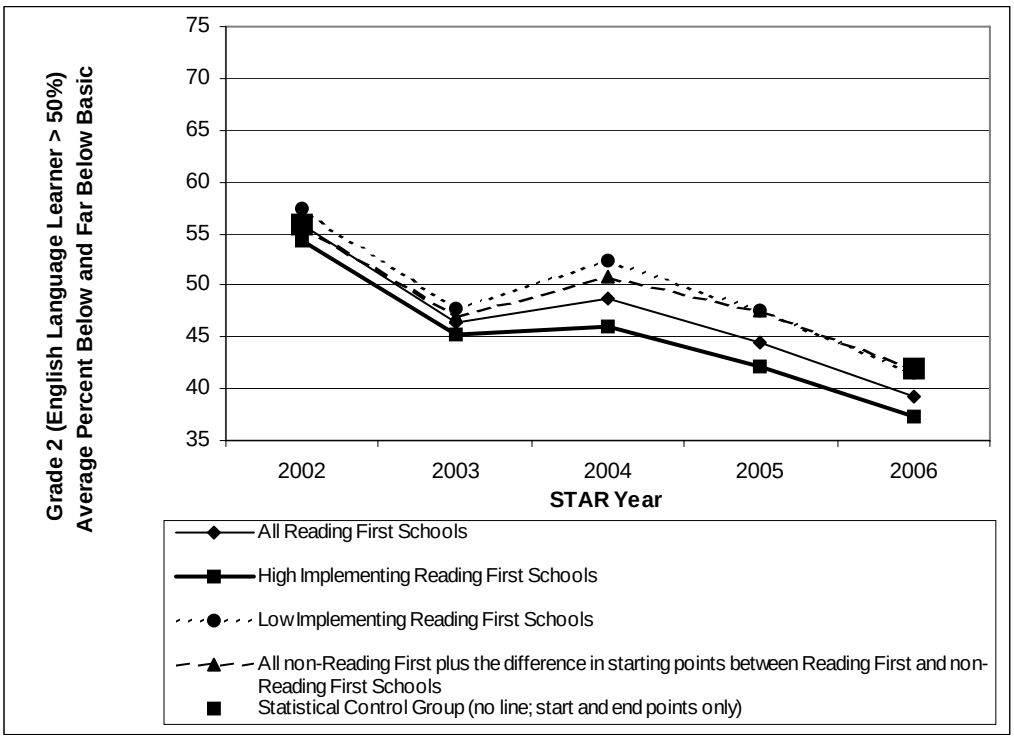


Figure 4.4.2: YIP 4 Schools with Majority of English Learners, CSTs, Grade 2, 2002 – 2006, Percent of Students Basic and Far Below Basic



Does low implementation of Reading First hurt schools?

It is important to know whether low implementation of the Reading First program hurts Reading First schools in terms of student achievement, relative to how they would have done without Reading First. Based on the data presented in this chapter, we would say no. Tables 4.1.1, 4.1.2 and 4.1.3 compare the gain scores of the average Reading First school with the gain scores of High Implementation and Low Implementation Reading First schools, the Statistical Control Group, and All Elementary schools. In all three YIPs we see that for most metrics of analysis, Low Implementation schools show gains that are similar to, or higher than, the Statistical Control Group and All Elementary non-Reading First schools. If low implementation of the Reading First program were having a negative impact, we would expect to see Low Implementation schools perform worse than these other groups. It is notable that Low Implementation schools generally move *more* students out of the bottom categories of the CSTs than do non-Reading First schools. We therefore conclude that low implementation of the Reading First program, though not desirable, is still at least as effective in helping the students in the bottom CST categories as non-Reading First schools, and that it is not causing a decline in students achieving standards.

Conclusions

Reading First shows evidence of being effective if:

1. Achievement gains in Reading First schools are positive;
2. Reading First schools show higher achievement gains than non-Reading First schools (though such differences are not statistically valid if the samples are not matched);
3. Reading First schools show higher achievement gains than a statistical control group;
4. High Implementing Reading First schools show higher achievement gains than Low Implementing Reading First schools.

Criterion 1: Are the achievement gains of Reading First schools positive?

For YIPs 4, 3, and 2, the answer is "yes" on all metrics including the EOY metric, and the gains are statistically significant.

Criterion 2: Do Reading First schools show higher achievement gains than non-Reading First schools?

The answer is "yes" on all the STAR metrics for YIPs 4, 3, and 2, and the difference is often statistically significant. The answer is not known on the EOY metric.

Criterion 3: Do Reading First schools show higher achievement gains than a statistical control group?

For YIPs 4, 3, and 2 on all the STAR metrics, the answer is "yes", and the difference is significant across achievement metrics. The answer is not known on the EOY metric.

Criterion 4: Do High Implementing Reading First schools show higher achievement gains than Low Implementing Reading First schools?

For YIPs 4 and 3 on the STAR and RFAI metrics, the answer is "yes." For YIP 2, the answer is "no" on the RFAI and two of the CAT/6 Grade 3 metrics. On the EOY metrics, the results are ambiguous.

We conclude that overall the Reading First program is having a positive impact on student Reading/Language Arts achievement.

* * *

An important idea in this chapter is the hypothesis that the cross-year positive trend observed in high-implementing Reading First schools is primarily a result of Reading First implementation, since Reading First precludes the use of non-aligned program elements. If this is true, it raises the possibility that the high growth rates seen in non-Reading First schools over the past five years arise from their adoption of the same program elements that are also driving the Reading First gains. If subsequent research should bear out this hypothesis, it would validate the efforts so far to make such program elements available to all California schools, not just those in Reading First, and it would justify expanding them further.

References

Cohen, J. (1992). A power primer. Psychological Bulletin, 112 (1), 155-159.

Haager, D., Dhar, R., Moulton, M., Varma, S. (2005). *The California Reading First Year 3 Evaluation Report, November 2005*. Educational Data Systems: http://www.eddata.com/resources/publications

Hasbrouck, J., & Tindal, G. (2005). *Oral Reading Fluency: 90 Years of Measurement (Tech. Rep. No. 33)*. Eugene, Oregon: University of Oregon, College of Education, Behavioral Research and Teaching.

Osborne, J. W. (2003). Effect sizes and the disattenuation of correlation and regression coefficients: lessons from educational psychology. Practical Assessment, Research & Evaluation, 8(11).

Parker, J. (2003). *Do Interim Assessments for Reading First Predict Success on the English/Language Arts California Standards Test?* Paramount Unified School District. Paper presented to the California Educational Research Association (CERA), November 14, 2003.

Yovanoff, P., & Tindal, G. (2003). *Scaling Early Reading Performance Tasks (Tech. Rep. No. 21)*. Eugene, Oregon: University of Oregon, College of Education, Behavioral Research and Teaching.

U.S. Department of Education. (2002). *Guidance for the Reading First Program*. It can be found on the U.S. Department of Education website at www.ed.gov/programs/readingfirst/legislation.html.

Chapter 5: Participants' Views of Reading First

The purpose of this chapter is to provide further insight into the views of personnel most affected by the Reading First program. Several data sources were used to examine the perceptions of teachers, reading coaches, principals and district administrators who have participated in the Reading First program. These data sources differ from those reported in other chapters because they are narrative in nature, derived from interviews, focus groups and text responses to open-ended questions on surveys. Such narrative sources offer rich, descriptive characterizations of participants' perceptions. However, it is important to note that qualitative research is not intended to provide an objective, unbiased perspective such as that obtained from statistical analyses. When used in conjunction with quantitative statistical analysis, qualitative research can provide a deeper explanation of statistical results.

The primary purpose of this chapter is to examine the perceptions of the different personnel groups regarding Reading First in general as well as issues regarding implementation, coaching and sustainability of practices. Given that both the Year 3 report and the quantitative analyses of this Year 4 report indicated that the level of implementation significantly affected student outcomes, one important purpose of this chapter is to compare the perceptions of personnel in high- and low-achievement schools regarding specific aspects of implementation. To that end, the district administrator interviews and principal focus groups were structured to draw information from districts that had reported a relatively high level of implementation, but in fact, had both high-achieving and low-achieving schools. These findings are reported in a section later in this chapter, "Comparison of High and Low Achieving Schools." In this evaluation study, the only source of data from teachers and coaches is the surveys given yearly. Teachers and coaches represent the "front line" of Reading First and potentially offer a perspective on how the program plays out at the classroom level. In this chapter, their responses to the open-ended survey questions from the Year 3 survey were analyzed to gain insight into classroom-level implementation issues.

Interviews with District-Level Administrators

To better understand the impact and implementation of the Reading First program at a district level, assistant superintendents were selected to participate in one-hour individual interviews held at their district office. Districts were selected for participation in the interview process if they had schools that fell into two categories based on Year 3 implementation (RFII) and achievement (RFAI) scores. One group of schools received high RFII and high RFAI scores (High-High) and the other group had high RFII with low RFAI scores (High-Low). From this pool of 12 districts that fit the criteria, eight were selected that fit geographically into four locations (two in the north and two in the south) and represented a range of smaller and larger districts. For the purpose of this evaluation, two Los Angeles Unified School district

Local Districts were included as separate districts and the administrators involved were at the local district level. In all other cases, the district administrator was an assistant or associate superintendent who had responsibility for direct oversight of the Reading First program. For each district and district-level administrator selected to participate in the interview process, focus groups of principals were also conducted (See "Focus Groups with Principals, below).

Interviews were conducted by independent researchers who had expert interviewing skills and a working knowledge of the Reading First program. Interviewers had no direct involvement with the districts or the Reading First program. The interviewers used a semi-structured interview format in which they asked predetermined questions with an opportunity to probe for further information or to explore further a topic of interest. Interviews were recorded for the purpose of transcribing and no identifying information was included in the audio recording. The interview questions are provided below.

Administrator Interview Protocol

4. (Background information) How long have you been involved in Reading First in your district? What is your role?
5. Describe the impact of the Reading First program in your district.
6. What do you see as the critical elements of Reading First for improving reading achievement in your district?
7. Do you have "waiver" classrooms in Reading First schools in your district? (Note: Waiver classrooms conduct reading instruction in Spanish, using designated reading curriculum in Spanish, and Spanish language testing. At the end of third grade, students must be transitioned into English and take the CST testing in English.) If the answer is no, skip to the next question. If the answer is yes, ask the following: "How has Reading First impacted the waiver classrooms with regard to reading instruction? Do you see any differences in how the waiver classrooms and English-instruction classrooms implement reading instruction?"
8. How will your district sustain the initiative after the end of Reading First funding? (Probe: What aspects of Reading First will your district maintain? What steps will you take to ensure sustained implementation?)
9. What has been the impact of the reading coaches in your district? How does the coaching model factor into your long-term plans for your district?
10. In this evaluation study, we look at reading achievement as measured by the Reading First assessments and the STAR testing. In your district you have Reading First schools that have made better gains than others in terms of reading scores. Generally speaking, can you explain why some schools have done better than others?

11. We are trying to better understand why some Reading First schools in California had high achievement gains while others did not. We have identified schools in your district that have a high reported implementation. Some of these schools also had high achievement, while others had less dramatic reading gains. The next few questions have to do with these specific schools. Again, I want to remind you that the information you provide is confidential. No information about your district, specific schools or individuals at schools will be shared with the California Department of Education, Reading First personnel, or the public. [Hand interviewee the list of high-gain and low-gain schools.] The first list of schools had high reported implementation and high achievement gains. The second list had high reported implementation and lower achievement gains. In your opinion, how do these two groups of schools differ with regard to school characteristics or contexts?
12. In your opinion, how do these two groups of schools differ with regard to implementation of the Reading First program? (Probe regarding time devoted to instruction, adherence to program.)
13. In your opinion, how do these two groups of schools differ with regard to school leadership?
14. In your opinion, how do these two groups of schools differ with regard to professional development?
15. In your opinion, how do these two groups of schools differ in terms of peer coaching in reading?
16. In your opinion, how do these two groups of schools differ in terms of reading assessment?

Focus Groups with Principals

Focus groups were formed that included principals from the same districts as the administrators described above and were held in four geographic locations. Each focus group included principals from one to three districts. At each location, two 1.5-hour focus groups were conducted – one for principals from high implementing/high achieving (High-High) schools and one for high implementing/low achieving (High-Low) schools. Six to eight principals per group were selected by their district's Reading First coordinator to participate. Thus, there were eight focus groups, with four representing High-High schools and four representing High-Low schools, across four geographic locations.

Focus groups were conducted by the same independent interviewers who conducted the administrator interviews. Focus groups were recorded for the purpose of transcribing and no identifying information was included in the audio recording. Principals were assured that no identifying information would be reported. Principals and interviewers were not aware of the High-High or High-Low status of the group being interviewed nor were the district coordinators informed of this structure at the time of arrangements. The interviewers were instructed to facilitate the discussion so that all participants had both an opportunity to respond to the question posed and to respond to prior comments made by others in a conversational format. Facilitators were instructed to monitor the time allotted per question as well as the

participation of individuals, and to ensure that no individual was allowed to monopolize the conversation and that diverse opinions were allowed. The concept of a focus group allows not only for individual responses to questions, but also allows for participants to add to others' responses or pose contrasting views. The focus group questions follow.

Principal Focus Group Protocol

- Describe your school and how Reading First has impacted your students.
- Each principal in this group is at a school that has been involved with Reading First for two or three years. Given a scale of one to ten, with ten being the highest possible, how would you rate your school's fidelity of implementation, and why?
- In your opinion, how important is coaching to your school? Describe how you work with the reading coach at your school.
- Think about how you will operate as a school when Reading First ends. In your opinion, what aspects of Reading First will your school continue to implement, and why?
- How does the coaching model factor into your long-term plans for continued implementation?
- To what extent do you believe the district will support sustained implementation of Reading First after the end of the grant? Why?

Open-Ended Questions from Year 3 Teacher, Coach and Principal Surveys

To better understand the views of school-level participants in the Reading First program, text responses to open-ended survey questions were analyzed by participant group. (Note that this Year 3 survey is similar to the Year 4 survey discussed in Chapter 3 of this report.) The open-ended questions were provided on the survey to give respondents an opportunity to provide more information than was asked in the multiple-choice, quantitative questions. Participants had the option of completing the surveys online or on paper. For this qualitative analysis, only online surveys were used because the text was available in an electronic database and would not require time-intensive transcribing.

It is important to note the limited scope of the open-ended responses. A relatively small percentage of respondents offered comments. Of the 18,492 teacher surveys collected, 7,146 (38.6% of surveys collected) wrote narrative responses online. Of the 903 coach surveys collected, there were 199 (22% of surveys collected) narrative comments submitted online. Of the 811 principal surveys collected online, there were 356 (43.9% of surveys collected) comments submitted online. It is also important to note that the wording of the questions may have invited negative responses. The questions were worded as follows. Teacher Survey: "Describe any unintended consequences of your district's adopted Reading/Language Arts program."

Coach Survey: “List any programs, school initiatives or activities that are having an adverse effect on your school’s implementation of your district’s adopted reading/language arts program. Briefly describe each adverse impact.”

Principal Survey: “List any programs, school initiatives or activities that are having an adverse effect on your school’s implementation of your district’s adopted reading/language arts program. Briefly describe each adverse impact.”

A Note about Qualitative Research

The advantage of qualitative research is to get an “insider’s view” of a phenomenon and “give voice” to participants in that phenomenon (Brantlinger, Jimenez, Klingner, Pugach & Richardson, 2005). Qualitative research provides rich, contextual information about phenomena of interest. Unlike quantitative research that uses statistical analysis to provide answers to specific questions, qualitative analysis provides an opportunity to explore issues that may emerge from narrative data, such as interviews, field notes or open-ended questions. Having specific research questions in qualitative research may limit the analysis and force the investigator to ignore potentially important and rich information. Instead, the goal of qualitative analysis is to systematically examine data for converging evidence of themes or categories of information. For this analysis, guiding questions were formed regarding each data source, giving the analysis direction, but not limiting the scope of investigation. Individual comments that did not fit within the parameters of the guiding questions were not ignored if they were validated by recurring instances and provided insight into the overall purpose of the Reading First evaluation study. Despite the potential for surprising or unexpected findings, this did not occur in this study. All responses that were coded fit well into the guiding questions.

Limitations of qualitative research, and this study in particular, should be noted. Though qualitative research may provide in-depth insight into phenomena and why they occur, results are viewed as inconclusive. Generalizability of findings beyond the respondents is somewhat limited. Being able to generalize findings requires knowing specific information about the sample and having some assurance that the sample is representative of a particular group. Furthermore, it is difficult to interpret the weight or meaningfulness of findings without the ability to quantify them. Reoccurrences of findings from multiple data sources or respondent groups certainly gives some insight into their importance, but they are not weighted or counted as in quantitative methods. The key advantage of using qualitative methodology is the ability to derive findings from multiple data sources and interpret them within a specific context. In this evaluation study, the qualitative data sources were examined for converging evidence of phenomena from multiple data sources (surveys, focus groups and interviews) and multiple perspectives (teachers, reading coaches, principals, district administrators) and interpreted through the lens of the expectations and assurances of the Reading First program.

The extent to which the sample is representative of a particular group in this study varies by data source. For principal focus groups and district administrator interviews, participants were randomly selected according to preset criteria and there is some assurance that the sample is representative of groups of interest. However, the survey responses from teachers, coaches and principals do not involve random selection. Those who chose to respond to an optional question may, in fact, differ significantly from the total pool of Reading First participants on key demographic characteristics such as length of time they have participated in the program and prior experience. These respondents represent program participants who *volunteered* to offer an opinion. Thus, a validity issue arises. Do these findings represent the views of participant teachers, coaches and principals in general, or is there some characteristic of those who chose to respond that differentiates them from those who did not?

An additional issue regarding the survey responses is the nature of the question. The wording of the open-ended questions led to specific types of responses that were slanted toward negative opinions (See survey analysis section below). Teachers were prompted to list unintended consequences of the program, whereas coaches and principals were asked to describe any competing programs or mandates that would interfere with Reading First. Survey respondents who held a negative bias or had a particular negative experience may have viewed the open-ended question as an opportunity to air complaints while those who were relatively satisfied with the program may not have felt compelled to offer an opinion at all. One way to address this would be to cross-validate individual comments with individual responses to the quantitative items on the survey. Did respondents who provided positive comments generally give positive ratings on the survey items and vice versa? This relationship between quantitative and qualitative responses is impossible to determine because, in this study, the narrative responses were pooled together into one text file without any identifying information regarding respondent, school or district. Thus, the reader is cautioned against comparing the frequency or weight of positive with negative response categories. Though there were more negative comments than positive, many written responses, in fact, contained both positive and negative statements that were treated as separate statements. Furthermore, it is possible that a lack of response from an individual may actually represent a positive opinion of the program.

Data Analysis Methodology

For this evaluation study, each data source consisted of a text file that was subjected to analysis using a qualitative software package. Data reduction involved coding segments, or “chunks” of data that contained meaning related to the study purpose. According to Miles and Huberman (1994), data reduction “is a form of analysis that sharpens, sorts, focuses, discards, and organizes data in such as way that ‘final’ conclusions can be drawn and verified (p. 11).” For this study, one researcher read through several pages of comments to establish an initial set of codes, or categories of meaning represented in the data, that was

consistent with the guiding questions. Two researchers then met to discuss code descriptions and coded one-fourth of the dataset together. To do this, the coders assigned one or more codes to each segment of text that constituted a unit of meaning. This might be a phrase, sentence or multiple sentences. When there was no obvious existing code for a segment, the two coders discussed whether the segment was relative to the study, and if so, created a new code with descriptors. Once the two coders established consensus regarding the code definitions and assignment of codes to segments in the subset of data, each coder then worked independently until the entire data set was coded. The coders met frequently to discuss new codes that emerged or alter code definitions.

Using a grounded theory approach and a recursive coding and analysis process, all segments were coded and categorized using a constant comparison method (Strauss & Corbin, 1990; 1994). Reliability was addressed in this study by the use of frequent conferencing among coders and researchers. Validity was addressed through an audit trail of the analysis process, the examination of confirming and disconfirming evidence, and triangulation of data across data sources (e.g., comparing teacher, administrator and coach views on a similar topic). Once all segments were coded, the software package facilitated refinement of codes and categories to find recurrent patterns and discern themes and their relative strength (Brent, Slusarz & Thompson, 2002). Segments of data that did not "hang" with other comments or fit into any category were considered lacking in weight and were discarded. The software includes tools for searching, categorizing, and sorting data as well as hypothesis testing and theory building.

Guiding Questions for Data Analysis

The research team developed guiding questions that provided direction for the analysis. Questions were designed to address the overall research questions for the larger evaluation study and also fit the nature of the data sources. Findings that did not fit within the framework of the guiding questions were not ignored if they occurred with a relatively high frequency within a data source. Guiding questions were as follows:

(d) What are the positive aspects of Reading First?

(e) What factors affect implementation?

(f) How do participants view the coaching model?

(g) What aspects of Reading First will be sustained?

(h) How do High Implementation/ High Achievement schools differ from High Implementation/ Low Achievement schools?

District and School Administrators' Positive Perceptions of Reading First

The positive impact of the Reading First program was a theme that ran through the administrators and principals Year 4 data sources (interviews and focus groups). Findings are summarized in this section by

respondent group and in the order of the weight of their occurrence. Table 5.1 provides a summary of the perceptions of district administrators and principals, with selected representative quotes[19].

District Administrators' Perceptions from Individual Interviews

There were four main findings related to the positive impact of Reading First from the perspective of district administrators that occurred with significant strength and frequency in the individual interviews conducted in Year 4. The frequency of the code is included in parentheses below, with frequencies from this data source ranging from 2 to 89. They are summarized below and quotes are provided in Table 5.1:

1. *Coaches (34).* Administrators viewed coaches as critically important to the district's adherence to and implementation of Reading First. Coaches were viewed as instructional leaders who are knowledgeable and resourceful. Though this category of comments fell under the heading of positive perceptions, it contained several sub-categories that are discussed in more detail below under a separate theme heading of coaches.
2. *Coherence and focus (27).* Administrators described the program as providing an impetus for district administrators to articulate clear instructional goals and follow through with them, so that the entire district, often including non-Reading First schools, could operate with a clear instructional agenda and message. Reading First as a "package" of components and guidelines provided districts with a positive model for reading instruction. Administrators referred to fidelity of implementation, or adherence to the model, as a key ingredient for success, and described how components of the program enabled them to maintain fidelity. Administrators also commented on the application of elements of the Reading First model, such as assessment and coaching, to other areas of curriculum, such as mathematics.
3. *Professional development (21).* There was a general consensus among administrators that professional development provided to principals, coaches and teachers participating in Reading First was highly effective. Positive comments described the intensity and amount of professional development, both at the beginning and advanced levels. It is important to note that at the beginning levels for teachers and principals, the training was not specifically unique to the Reading First program. AB 466 training for teachers and AB 75 training for principals was available throughout the state, regardless of Reading First participation. Many of the comments included in this category did not distinguish between beginning and advanced levels. LEAs tend to reference advanced levels of professional development as connected to AB 466 training. All LEAs are served by one of seven Reading Implementation Centers, which are networked through

[19] Throughout this report, every effort was made to use exact quotes. Some quotes were edited slightly to correct grammar or spelling, clarify an ambiguous word such as a pronoun, mask the identity of a school or district, or generally facilitate readability. In all cases, the original meaning is intact.

the California Reading Implementation Center at the Sacramento County Office of Education. These centers serve as professional development providers and are authorized to be the sponsors of the AB 466 program and the Reading First advanced programs as per the California Reading First plan.

Principals' Views: Focus Groups

There were six main findings related to the positive impact of Reading First from the perspective of principals who participated in the focus groups that occurred with significant strength and frequency. The frequency of the code is included in parentheses below, with frequencies from this data source ranging from 5 to 135. They are summarized below and representative quotes are provided in Table 5.1:

1. *Coaches (135).* Principals' perceptions of the value and impact of coaches far outweighed any other category of comments. Principals viewed coaches as highly valuable and as indispensable resources to the school community. Numerous comments described the importance of the coaches' role in assisting principals to monitor classrooms and hold teachers accountable for implementation. It is important to note that coaches have no official role or authority to monitor or evaluate teacher performance. The term, "monitoring," as used here, was derived from the principal focus groups and represent principals' perspectives. In fact, many principal distinguished the "monitoring" role of a coach from that of a principal. Coaches were generally described as keeping a watchful eye on instruction in the classrooms, providing assistance to teachers as needed, and fulfilling an important role of keeping principals informed so they could intervene if necessary. Some principals were careful to point out that coaches are teachers' peers and, in principal-coach meetings, discussions focused on needs across a grade level or across the school rather than on individual teachers who may be experiencing difficulty with the program. Principals also viewed coaches as instructional leaders who could collaborate well with teachers and work within the classroom in a way that principals could not. Again, this category of comments fell under the heading of positive perceptions but is discussed in more detail below under a separate theme heading of coaches.
2. *Professional development (85).* This set of comments focused on the intensity and depth of training provided to teachers through the workshops, ongoing site-based professional development and individualized coaching. Principals viewed this aspect of Reading First as having a significant impact on their school's fidelity of implementation. Again, though some comments did not differentiate beginning (i.e., AB 466 and AB 75) and advanced (Reading First and Coach) training, many of the comments were directed at the ongoing professional development that occurred through the coaching support and follow-up training that occurred through Reading First participation.

3. *Assessment and accountability (37).* Principals participating in the focus groups viewed the ongoing assessment requirements as having an important impact on teachers and how they implemented the program. Principals were positive about having required assessments and felt that it led to a much-needed accountability. The assessments prompted teachers to cover all essential components of the curriculum and think about the needs of individual students. In principals' views, Reading First provided support and opportunities for monitoring and accountability at the classroom level. Through the periodic assessments and the coaches' classroom-level involvement, principals felt well informed about classroom implementation and ongoing student progress and could intervene in a more proactive manner if there were problems.
4. *Coherence and focus (37).* Principals expressed positive views of the consistency in focus and practice that came from participation in Reading First. They often spoke of a "paradigm shift" in their schools. Having a clear and consistent instructional message was important to achieving this. Principals also indicated that the common curriculum and implementation across schools was helpful when students moved from one school to another.
5. *Improved instruction (35).* Principals reported a positive impact on the quality and content of instruction. They discussed improvement in individual teachers as well as a general improvement. Principals also discussed the positive change as teachers were required to fully implement the program for the entire length of time allotted to reading/language arts instruction.

Table 5.1 Representative Quotes: District and School Administrators' Positive Perceptions of Reading First

	District Administrators	**Principals**
Positive Perceptions of Coaching Summary: Reading coaches are perceived as critical to the coherence of the Reading First program. Coaches have a direct and positive role in achieving implementation. The coach's role in the classroom, working with teachers on implementation compliments the monitoring role of the site principal.	*"Our coaches are the site instructional leaders. Our principals rely on the coaches' expertise and when we have openings for vice principals, often the reading coaches apply, and they get picked up immediately."* *"I don't think we could have implemented the program with fidelity without the coaches… They are the people who are really in the trenches with the teachers. Our coaches just rolled up their sleeves and they go in and help teachers. And we have protected their role with the union so they are not considered evaluators."*	*"The coach represents the face of Reading First to the staff and the parents. And in my experience, and from the comments I get from teachers, the coach is the most important part of the program. If you have a coach who is committed to establishing relationships with teachers, the principal and the staff, then things work out well. If the relationship component is not up to par, then the whole program suffers. And our school has had, as I said, ebbs and flows. Sometimes it's good, sometimes it's bad."* *"It's having that critical person to help you strategize. You work*

		together. The coach would come in, "I just did this, look at this, look at the assessments," and she ran the grade level collaboration meetings. We would be able to talk about the data, or talk about what she saw happening in one classroom, what I might have seen happening in another…the coach is the person who seems to be the keeper of the most current information."

Positive Perceptions of Coherence and Focus Summary: Respondents' views indicated that the Reading First program, and specific components, have brought a coherence focus, or consistent message, to the district or specific schools. Reading First has brought a paradigm shift to the schools.	*"When Reading First was introduced and it became very evident that teachers needed to plan together, not in isolation, we started seeing that collaboration taking place. We've seen it progressively increasing, and now, three years later, it's something that is ingrained as a part of the culture throughout the district."*	*"We've always had data analysis. We've always had staff development. We've always had support personnel on our campuses, but what Reading First has done is given us a new way to look at all three. So I think that we will keep it going, because the new way to look at those, that paradigm shift, has worked for us.* *"It is a mental and paradigm shift. The teachers have gone from being in a classroom behind closed doors with nobody monitoring to now people monitoring. Actually now they are very proud and want people in the classroom. And that in itself is a shift and has moved us to a 10 (full implementation). Getting the teachers to understand the value of the pacing schedule, getting them to understand why we're giving 2 and ½ hours for reading instruction in K-3, that is important in itself."*
Positive Perceptions of Professional Development Summary: The professional development provided to principals, coaches and teachers was viewed as highly effective.	*"What Reading First has really done is that it has brought in the resources to support a professional development infrastructure to ensure that teachers who may not be reading specialists, who may not have that particular pedagogical linguistic research background, are now competent.*	*"I see how the staff development has made them (teachers) teachers of reading. When you take that to a district level, you have the understanding that we have now reached a level that we have never been able to reach before in terms of expertise."* *"There was a time when we were considered a hard to staff school, but because of this wonderful training, when I interview teachers, I can honestly look them in the eye and say when you leave my school you will be one of the best trained teachers of reading in the entire district, and be one of the highest qualified teachers on the staff."* *(Regarding principals' training) "Going through our training has*

		really helped me to see which teachers may not be doing all the components to the level they should be."
Positive Perceptions of Assessment and Accountability Summary: Respondents had positive views of the assessment requirements of Reading First and how they allowed them as administrators to be more involved in monitoring instruction.		*I think a key piece is the issue of accountability, and from that standpoint Reading First has made all of us accountable in ways that we really haven't been before. The knowledge of assessment and how we use that assessment to drive our curriculum is really important… This is so critical to what we do and our kids are reading; our kindergarten kids are reading."*
Positive Perceptions of Improved Instruction Summary: Respondents discussed the noticeable improvements in the quality and nature of instruction as a result of participation in Reading First.		*"Reading First allows students to have better instruction because now the teachers are no longer working in isolation. Through collaborative meetings and ongoing professional development, the teachers are exposed to more. They go into each other's classrooms and observe lessons. I think it elevates their professionalism."* *"The pedagogical changes (methodology) transcend language arts. What the teachers are learning they are applying to other subjects. You're seeing the capacity for change in teaching and learning, because teachers are now understanding what true pedagogy looks like, and why they are doing it. There is purposeful instruction and that impact is invaluable."*

District and School Administrators' Perceptions of Reading First Implementation

Participants' positive views of implementation of the Reading First program was another theme that ran through the administrators and principals Year 4 data sources, with few, if any, negative comments. Findings are summarized in this section by respondent group and in the order of the weight of their occurrence. Table 5.2 provides a summary across district administrators and principals, with selected representative quotes.

District Administrators' Views

Within the theme of implementation, district administrators offered a perspective from the district level of implementation. Comments within this theme fell into only one category, that of administrators' direct involvement, with 52 segments coded. Many of the comments described site visits and "walk-throughs" to observe and monitor implementation. They explained that the program gave them, as administrators, license to set expectations for full implementation, even in the face of resistance from teachers. They also indicated that the program built capacity within the district to deliver professional development and provide support.

Principals' Views

From the principal focus groups, there were three main categories of comments, summarized below.

1. *High rating of school implementation (40).* In response to the question about their school's fidelity of implementation, principals in both sets of focus groups generally gave a high self-rating, indicating that they believed their school was fully implementing the program. When a principal indicated less than a full rating of 10, on a scale of 1-10, they generally indicated an 8 or 9, with an explanation of what component needed improvement and the steps to get there.
2. *Time in program affects implementation (37).* Principals indicated that time in the program affected their implementation, often describing the struggles they experienced in the first year or two. They described initial difficulties with pulling teachers on board or managing all the components of the program, and then discussed how much better their implementation was following the initial "growing pains."
3. *Continued funding essential (29).* The third category of comments had to do with the importance of the funding. Principals viewed the Reading First funding they received as vital and necessary to achieve implementation and continue the program. Some expressed concern that when the funding is discontinued, their district or school would have difficulty sustaining the level of professional development and support needed to achieve full implementation.

Table 5.2 Representative Quotes: District and School Administrators' Perceptions of Reading First Implementation

	District Administrators	Principals
Administrator Involvement Summary: Administrators were directly involved in implementation through site visits, consultation with principals, meetings, and coordination activities. Administrators felt enabled to set district expectations for implementation, even in the face of resistance	*"We deliver a lot of the training ourselves because we now have people who were trained to do that…"* *"In those visits, we visit every single K-3rd classroom… and we look for very specific elements of implementation related to Reading First, such as spelling cards and focused logs, and certainly implementation of the curriculum, teaching the standards, and engagement of students. We debrief before we leave with our administrative staff, the principal, vice principal and reading coach."* *"We dealt with a lot of issues, where teachers felt overwhelmed by Reading First. At the time it was uncomfortable perhaps to deal with some of the issues repeatedly, but it enabled us as district folks to clarify the purpose of Reading First, the goals, intent and need. It enabled us to have discussion with teacher leaders and ultimately we would put our conversation in writing, and send minutes out to all the teachers and administrators in the entire district."*	
High rating of school implementation Summary: In response to the question to rate their school's level of implementation, most indicated a level of 8-10 on a scale of 1-10		*"I would say we are an 8 or 9, and the reason is because every component of the program is not taught on the same level of rigorousness in every classroom. Part of that is where we are with professional development."* *"We're in our fourth year and we're now a 10. I would say our whole district is a 10. We have*

		our pacing guides, our lessons planned. I gave the teachers a schedule for the entire year and asked them to clear their calendars and be prepared to come to all the meetings. The first year, we were definitely not a 10. Teachers would not invite the coach into their rooms. They would say, "Why is that teacher on special assignment?" Now, it is, "Can the coach come and work with me?" or "Do you have five minutes? I want to show you something."
Time in program affects implementation Summary: The more time a school has been in the program, the better the implementation. It takes time for all the components of the program to develop and work together.		*"You have to give a good five years before it is really stabilized. Reading First is something that is good and it's moving forward. It's not 100% on everybody's plate, but it's getting there. We still have pieces that need to be developed… so that everybody at the school, and at the district level, and the community, owns it."*
Continued funding essential Summary: Principals expressed concern that if the funding were to be discontinued, they would be unable to sustain implementation. They viewed the funding as extremely effective in improving teaching and learning at their schools.		*"(If funding were discontinued) we would keep it going. We've seen that it (the program) has worked and we will keep it going. We'll find the money. If we need to do carwashes, we will."* *"The release time for teachers and the reading coaches are costly. We could not have been successful without the funding."*

District and School Administrators' Perceptions of the Reading First Coaching Model

District-level and site administrators (principals) expressed strong and positive views of reading or literacy coaches that existed in their district as a result of Reading First. This section treats "coaching" as a separate theme, and describes the categories of segments coded within it. Table 5.3 presents representative quotes within these categories.

District Administrators' Views

1. *Coaches' role is critical to success (34).* District administrators described the role of the reading/literacy coaches in their districts as essential for accomplishing the coherence or focus of

the district and for achieving implementation. They discussed the communication between the district personnel and coaches as helpful and commented on the dynamics between principals and coaches.

Principals' Views

1. *Coaches are a valuable resource (110).* The high frequency category of comments in this category supports the conclusion that principals view coaches as indispensable. Their comments described the role of coaches as providing knowledge and assistance on an ongoing basis to teachers. Because coaches were free of teaching duties, they were able to work closely with teachers, providing very specific assistance and support. Coaches assisted with the acquisition of materials, demonstrated how to deliver specific aspects of lessons, answered questions on an ongoing basis, interacted with parents, gave feedback and validation to teachers, and assisted principals with supporting teachers having difficulties.
2. *Coaches collaborate with teachers and principals (81).* Coaches were viewed as highly collaborative and this role was very important to implementation. Principals viewed coaches as being able to bridge the communication between the site administrator and the classroom. Because coaches were trusted colleagues, teachers were more likely to take direction and support from coaches than from principals. Coaches ran effective collaborative meetings to discuss student data and instructional methodology.

Table 5.3 Representative Quotes: District and School Administrators' Perceptions of Reading First Implementation-Coaching Model

	District Administrators	Principals
Coach role is critical to success Summary: Administrators viewed coaches as essential to achieving success with implementation and changing student achievement.	*"We deliver a lot of the training ourselves because we now have people who were trained to do that…"* *"The coaches are the center of Reading First. In the district, making sure the program is carried out, the coaches hold the collective expertise. That is a knowledge base we have built in the district. There are so many beneficial aspects- the fact that they are non-evaluative, they provide support- so they can be an ally to the teacher, while the principal provides the supervision and the pressure."*	

Coaches are valuable resources Summary: Coaches were viewed as important resources to sustain high quality instruction. Coaches provided assistance and support in very specific ways to teachers regarding the curriculum and assessments.		*"The coach comes in, observes the teacher, and notices something in particular that could be better. Then, she models for the teacher and then has the teacher do it for her to observe again. So, it's on the job staff development. It takes away that layer of (teacher resistance), 'Well, I don't know if it would work with my kids,' because the coach has shown her it works right there in her own territory. You know, the classroom is the teacher's territory. Only the coaches can do that- go into their territory and intervene"*
Coaches collaborate with teachers and principals Summary: Coaches were viewed as highly skilled not only in methodology, but also in how to work collaboratively with teachers and principals.		*"The coach and I talk every day. It's nothing really planned- she just drops in my office. We discuss, 'How's it going? Who are you seeing? Who are you working with?' We sometimes visit classrooms together, but even if I don't go, I know she is keeping things going."* *"The coach is there for planning with the teachers. And, they look at data and talk about what trends she (the coach) sees across classrooms. She is able to troubleshoot, get people the things they need, materials, books, or just give them ways to use certain parts of the program better."*

District and School Administrators' Perceptions of Elements to be Sustained

District and site administrators were asked about what elements of the Reading First program they viewed as likely to be sustained in their district and school. Sustaining of the coaching model was the key finding in this theme. Both respondent groups indicated that coaching was critical to their success and they were hopeful that it would be funded with district funds when the grant program discontinues. Findings are not summarized here by respondent group because there was only one category of codes that occurred with consistency. Table 5.4 provides selected representative quotes.

Regarding the sustaining of the coaching model, there was high agreement that coaching was very important to success and should be sustained. Because the hiring of coaches, or out-of-the-classroom teaching experts, is costly, many respondents expressed concern that their district or school would have difficulty continuing the coaching model with current funding structures. Some described the coach as so critical to their success that they would attempt to use categorical funds to maintain the coaching model if the Reading First program were discontinued.

Table 5.4 Representative Quotes: District and School Administrators' Perceptions of Reading First Implementation-Elements to be Sustained

	District Administrators	Principals
Sustain the coaching model Summary: District and site administrators were in high agreement that the coaching model should be sustained, due to its importance. However, many expressed concern that funds would not be available to do so when Reading First is discontinued.	*"I thoroughly believe that we need an instructional coach for keeping the elbow-to-elbow support structure."* *"The important thing is that we have the expectation that coaches will be paid by each school out of categorical funds, and we will continue with training (for coaches).*	*"We're starting to think about what will happen if we have no more Reading First funding. I would like to see the coaches continue to be a part of our model."* *"I would be so overwhelmed if I could not have that coach, or that extra support."*

Comparison of High Implementing/High Achieving and High Implementing/Low Achieving Schools

One purpose of the qualitative study was to explore differences between schools that, according to survey data, were implementing the program with a high degree of fidelity, but in fact differed in their achievement results. As described earlier in this chapter, district administrator interviews included questions about specific High RFAI and Low RFAI schools in their district. The principal focus groups were composed of principals from the same districts as the administrators, with one group representing High RFII-High RFAI schools and another High RFII-Low RFAI. In this section of the report, we examine similarities and differences between these two groups of schools.

Though district administrators were able to pinpoint differences between the high and low achieving groups of schools, there were few discernable differences in the perceptions of the principals involved in the focus groups. Below, the district administrators' perceptions are discussed first, followed by the similarities and differences between the principal groups. Because themes were not parallel across respondent groups, representative quotes are not provided in table format. Instead, when it provides further insight into a category, a quote is provided. When the summary is sufficient, no quote is provided.

District Administrators' Views of Differences between High Implementing/High Achieving and High Implementing/Low Achieving Schools

District administrators provided some insight into differences between the higher and lower achieving schools in their districts. There were four categories of findings under the theme of implementation. They

are discussed in the order of frequency from high to low. The range of frequency in the high-low distinction was 8 – 89.

1. *Differences in implementation (89).* Administrators generally reported differences in implementation, with higher implementation at the higher achieving schools, though it was not consistent with all schools or all districts. In some cases, the administrator could not articulate specific differences due to implementation issues, and instead, cited demographic or staffing concerns. When implementation concerns were expressed, they included issues with teachers being resistant to instructional methodology or assessments, principals who were not fully supportive or involved, principals with less experience, lack of monitoring at the classroom level, or inexperienced coaches. To illustrate, two quotes are provided:

   "Knowing the curriculum is one thing and knowing the program is one thing, being aware of all the assurances and how you are supposed to deliver it is one thing. But then being able to be an artful teacher in engaging students is another. And those two have to be married. So I think it is very complex, but it boils down to the adults working together with a clear understanding of where they're heading and then aligning all of the support and resources towards that. And some schools do that better than others."

   "Bottom line: There are some schools (from the lower achieving schools list) where I'm really not surprised, because they don't think their kids could learn. The teachers think they were teaching literacy skills, but you'd go into the classrooms and you could see that they're just doing 'literacy' activities and not really teaching the curriculum. The other schools (higher achieving) just really embrace the notion of having all students achieve to the same standard and so they'll figure out a way to support that child who is having trouble. The difference is sort of intangible, it has to do with the commitment, the passion, the understanding and the skills teachers need to really buoy the lower kids up to the level of proficiency."

2. *Differences in site administrators' leadership (81).* District administrators cited some instances of differences in site administrators' leadership skills or styles. In some lower achieving schools, administrators were viewed as more actively involved in Reading First. For example, in the higher achieving schools, principals were more aware of instructional needs, professional development opportunities, and how to motivate teachers. One administrator noted: *"You have those principals who ask about available resources or, you know they know how to, instead of being passive about what's available for them, they will go out and find resources and they'll bug us until we could provide the kinds of support or professional development they really need.*"

Some of the comments indicated that administrators were puzzled by the differences in the schools because they saw the leadership at both sets of sites as strong.

3. *Stability or experience level of personnel (58).* Generally, higher achieving schools were viewed as having more highly skilled teachers or administrators while lower achieving schools had more inexperienced personnel, with newer teachers described as being overwhelmed by the complexity of the program and having difficulty getting through all sections of the curriculum. However, there were numerous exceptions. For example, a new principal was described as particularly enthusiastic about the program and therefore, had more success with achieving teacher buy-in. Veteran teachers were sometimes described as "headstrong" and resistant to take on new methods. Turnover in personnel was cited as a factor: "*There was, in at least two of the schools with lower achievement gains, significant change in staff personnel. And what that means is, though in an individual classroom you may be having a high implementation of the curriculum, the strength of the teacher collaboration isn't there, resulting in a lack of differentiation that's focused around student mastery of skills."*
4. *Differences in school organization (48).* To some extent, higher and lower achieving schools differed in organizational structures, including time and consistency of collaborative meetings, coaching availability, participation in professional development, principals' classroom visits, and other factors. For example, *"Two of the schools that experienced lower achievement gains actually did not have a well established and effective internal professional development system or plan."*

Comparison of Principals' Views from High Implementing/High Achieving and High Implementing/Low Achieving Schools

There were few differences in the perspectives of principals from the higher and lower achieving schools with regard to implementation. Both groups reported positive perceptions of the coaching model as reported in previous sections of this chapter. Both groups also reported positive perceptions about the professional development that teachers received. There were two categories of comments in which some differences occurred. They are discussed below.

1. *Reading First assessments promote accountability.* In the higher achieving focus groups, there were 47 instances of this code as compared to 31 in the lower achieving groups. This code indicated that principals viewed the required assessments as highly important to improving achievement due to the accountability that it creates. When teachers and principals are reviewing and discussing data on an ongoing basis, they feel more accountable to produce results. Across these two groups of principals (higher and lower achieving schools), the comments did not differ significantly in nature, only in quantity.

2. *Monitoring.* The category of monitoring had to do with the extent to which principals perceived they monitored the program through classroom visits, communicated with teachers and coaches, viewed assessment data, and intervened when implementation faltered. Compared to 28 segments coded with this category in the higher achieving schools group, only 9 were coded in the lower achieving schools group.

Teachers' Views of Reading First from Year 3 Surveys

This section summarizes the teacher responses on the Year 3 survey open-ended question, "Describe any unintended positive or negative consequences of your district's adopted reading program." Though the prompt specifically asked teachers to comment on their district's adopted reading program, the actual responses provide insight into teachers' views of Reading First in general. Most comments addressed respondents' views of the adopted curriculum materials (i.e., OCR or HM), but many respondents took the opportunity to include more general comments about various aspects of the Reading First program.

Due to the time-intensive nature of qualitative analysis, it was not possible to analyze all qualitative data sources from the Year 4 data set prior to the writing of this report. The external evaluators were able to conduct this analysis of the Year 3 teacher surveys while the Year 4 data collection was under way. The Year 4 survey open-ended responses will be considered for inclusion in a subsequent report.

Rationale for Examining Teachers' Views of Reading First

Teachers' perceptions of the Reading First program provide a classroom-level glimpse of implementation. Their ratings on the survey figure heavily in the evaluation of implementation as reported in Chapter 3 but, in this chapter, their responses to the open-ended question are used to provide a descriptive analysis of teachers' perceptions about implementation of the program.

For this analysis, findings were categorized into main themes and are summarized with a comparison of positive and negative examples, accompanied by an analysis of the Reading First expectations derived from the national Reading First guidelines and California Reading First plan.

Teachers' Perceptions of the Curriculum

Positive Responses	Negative Responses	Reading First Expectation
Summary: Teachers reported positive regard for the two adopted curricula in general, and specifically with regard to the phonics component, the materials and the unity or coherence it provided to have a common curriculum across grades and across schools.	Summary: Comments within the theme of curriculum of notable frequency that reflected negative perceptions included teachers' perceptions that the curriculum lacks opportunities to review skills when students struggle or that the program is inappropriate for struggling readers. They also expressed displeasure with the writing instruction and the literature and stories included in	Reading First guidelines indicate that instruction should be grounded in reading research. A complete Reading/Language Arts program would also incorporate spelling, writing and language development. The first quote included here represents a teacher who grasps the complexity of covering the core skills instruction for a diverse classroom. For some teachers,

Representative quotes: *"This program forces me to stay sharp in all aspects of reading/language arts because it is a very interactive, multi-level program. I can't afford to take a moment off or the high kids will catch me in a moment. This is good. Other programs have made teachers complacent, teaching from habit rather than thought."* *"I feel that this is the strongest and most complete phonics instructional lessons that I have ever seen, and our students really need this and have excelled in this area because of it."* *"I feel that because our school is committed to teaching the curriculum the way it was designed and in its entirety that our school is being affected positively as children move from grade level to grade level with a curriculum and terminology that builds on the previous year."*	the texts. Representative quotes: *"The pacing guide does not allow us time to reteach or adjust teaching based on the needs of our students."* *"I feel the students need a bit more variety in literature instead of the same books over and over."* *"I feel the writing component is weak. We needed to add much to this part."* *"I would like to see more spiral review on phonics and spelling."*	whose prior training differed from current methodology, seeing student results cements the importance of the research behind the program. For other teachers, the idea of changing from their prior model is not comfortable or desirable. The comments regarding the lack of opportunity to use high quality literature or stories of the program indicate misperceptions about the Reading First program. Teachers have historically read books to and with their students and the Reading First guidelines neither prohibit nor discourage the use of external literature.

Teachers' Perceptions of the Impact of Reading First on English Learners

Positive Responses	Negative Responses	Reading First Expectation
Summary: A moderate number of comments expressed the view that EL students were successful in the curriculum and the teachers felt supported in meeting the needs of these students. Representative quotes: *"Every student in my classroom is reading at or above expected grade level regardless of native language!"*	Summary: There were more negative than positive comments regarding the impact of the Reading First program on EL students. These comments focused on the difficulty of the vocabulary and language included in the curriculum as well as more general concerns about the difficulties encountered by a fast-paced and rigorous program. Representative quotes: *"My concern is for second language learners. The content may be too high for the students. There aren't necessarily any stories or vocabulary words that*	The professional development offered through the Reading First program includes components regarding differentiating instruction for struggling learners, including English learners. The positive responses included here illustrate teachers who maintain high expectations for their English learner students and adjust their teaching to ensure they meet them. The first negative example included here represents a teacher who may need additional support to understand how to respond to the individual needs of English learners while maintaining high expectations. The second

"I am amazed at how well the students do in our school considering their language and socioeconomic level. I see a lot of progress in all students with this program. The program challenges my more advanced students and brings up my students that struggle at times."	*these second language learners are familiar with."* *"The program does not meet the needs of a language learner. It is assumed that a language learner can understand the vocabulary presented just by listening to the stories or because words are repeated throughout the theme."*	example illustrates a lack of understanding of how to differentiate or adjust instruction to make the high expectations achievable for English learners. A key component for English Learners would be to supplement instruction through explicitly teaching vocabulary and language concepts during small group time.

Teachers' Perceptions of the Pace, Schedule and Rigor of Reading First

Positive Responses	Negative Responses	Reading First Expectation
Summary: Though most comments about the fast pace of the program were negative, some teachers expressed positive views of having a common plan across the district. When students move from school to school, they do not lose academic ground and are familiar with the concepts and structure of the curriculum.	Summary: Most comments in this category were negative, expressing the view that the curriculum or pacing plan moves too quickly and does not allow for differentiation of instruction when students do not grasp concepts. Other comments, particularly in kindergarten, indicated that the pace moves too slowly, as students do not learn the letters and sound quickly enough. Numerous comments indicated that full implementation of Reading First limits teachers' ability to teach other subjects, that reading instruction takes away from social studies, science or the arts. Representative quotes: *"When a child does not understand something, we need to move on in order to stay with our pacing schedule. I just think that is silly. We don't meet students' needs, just work as a whole."* *"The pacing plan creates a rigid implementation schedule that compels the teacher to strive toward quantity rather than quality of learning."*	Providing reading/language arts instruction for two to three hours per day represents a significant portion of the school day and is a dramatic shift in the time structure of schools. The tension between full implementation of the reading program and other subject areas is a real and important issue. Though science and social studies content is integrated into the reading curricula, teachers' views that it is insufficient to fully address content standards may be valid concerns and warrant further consideration as the program continues. Comments such as those included here regarding the inability to differentiate instruction due to the pacing demands indicate a need for further support to assist teachers in supplementing instruction during small group instruction and through ongoing monitoring and support.

References

Brantlinger, E., Jimenez, R., Klingner, J. K., Pugach, M. & Richardson, V. (2005). Qualitative studies in special education. *Exceptional Children, 71* (2), 195-207.

Brent, E., Slusarz, P. & Thompson, A. (2002). *Qualrus: The Intelligent Qualitative Analysis Program Manual*. Columbia, MO: Idea Works.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273-285). Thousand Oaks, CA: Sage.