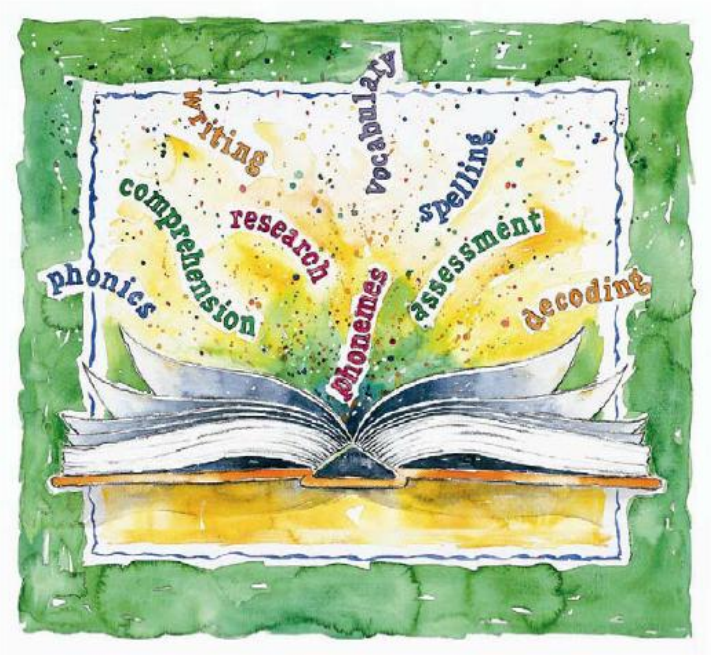
# The California Reading First Year 7 Evaluation Report

**November 2009**




**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

During the past seven years, California has witnessed a transformation of reading instruction that can be attributed not only to state and local policy and initiatives, but, in large part, to the Reading First program, an unprecedented federal program aimed at improving reading instruction in the United States. The federal Reading First initiative was authorized in 2001 as part of the No Child Left Behind Act and its purpose was to provide states with the means to improve reading instruction in grades K-3, in low achieving schools where many students are socio-economically disadvantaged. Since August 2002, when the State of California was notified that they would receive approximately $900 million over a six-year period, the state has launched a comprehensive and ambitious effort to build capacity at state and local levels through establishing policy and guidelines focused on improving reading achievement and funding large-scale professional development and ongoing support.

This final external evaluation of California's Reading First program provides a cumulative report of the achievement and implementation data collected over time. With this report, we reflect on the lessons learned and make recommendations regarding the continuation of policy and processes related to Reading First. The key findings, synthesizing those of prior reports, are summarized below. The body of the report contains the detailed analyses to support these conclusions. It should be kept in mind, as noted in Chapter 1, that this report examines Reading First in terms of its unique programmatic model for the State of California according to the California Reading First Plan. The findings for "Reading First" referenced below refer to this statewide study, a particularly significant fact in light of a national Reading First impact study which published findings that are not consistent with those presented here.

**Finding #1: Reading First has consistently led to achievement gains for students in historically low-achieving schools and with students of socio-economic disadvantage.** Reading First schools have realized significant reading achievement growth since the inception of the program in comparison to a statistical control group and non-Reading First schools. Various metrics have demonstrated a positive and significant impact of Reading First on achievement compared to non-Reading First schools and a statistical control group across all years of the program. The Reading First Achievement Index (RFAI), a measure of individual schools' achievement progress, steadily rose in Reading First schools over the past seven years. (Chapter 2)

**Finding #2: Level of implementation has consistently influenced reading achievement.** Higher implementation has been consistently associated with higher achievement in Reading First schools. The Year 3 through Year 7 reports showed significantly higher achievement for high implementing schools compared to lower implementing schools, as measured by the Reading First Implementation Index and

various achievement metrics. A meta-analysis of effect sizes conducted for the Year 6 Report (which did not include Year 7 data) found that the average Reading First (standardized beta) effect size in predicting all possible outcome variables from 2003-2008, after controlling for starting point and demographic factors, was 0.093 with a standard error of 0.006. This is approximately 15 standard errors higher than zero, where 2 standard errors above zero would be sufficient to claim a statistically significant effect with 95% confidence. What this means in practice, as a rule of thumb, is that a doubling in implementation corresponds to a doubling in achievement gains. (Chapter 2)

**Finding #3: Growth remains significant.** As of the Year 6 Report, which summarized all cohorts, the Reading First Achievement Index (RFAI), a composite of K-3 achievement metrics for Reading First schools that ranges from 0 to 100, had risen an average of 3.1 points per year, equivalent to 18.6 points over six years relative to a starting year. In Year 7, looking only at Cohort 2 and Cohort 3 schools, the RFAI has risen approximately 2.5 points per year; though a lower figure than in previous years, this gain is still statistically significant. (Chapter 2)

**Finding #4: The Reading First effect generalizes across student performance levels.** In Reading First schools, students have steadily moved into the "Proficient and Above" categories of the grades 2-5 California Standards Test (CST) achievement test, and the number of students in the "Below or Far Below Basic" categories has steadily decreased. The migration of students out of "Below and Far Below Basic" is more than twice what it is for non-Reading First schools. (Chapter 2)

**Finding #5: Reading First has had a significant impact on reading outcomes in grades 4 and 5.** As of Year 6, the above findings were replicated in grades 4 and 5, even though Reading First is a K-3 program. In Year 7, the differences between high implementing schools and the statistical control group were not statistically significant in all cases for Cohort 2 schools. However, Cohort 2 Reading First schools show significantly higher growth rates than non-Reading First schools in grades 4 and 5. This finding indicates a sustainable and replicable effect of the program once students no longer have grade-level access due to funding and programmatic limitations (K-3). (Chapter 2)

**Finding #6: Reading First has had a significant impact on reading achievement for English learners (ELs).** English learners in Reading First schools show higher rates of growth than English learners in non-Reading First schools across the state. English learners in high implementing Reading First schools show higher rates of growth than English learners in low implementing Reading First schools, and the implementation effect is more pronounced for English learners than for the student population as a whole. A corollary is that English learners in low implementing Reading First schools are at particular risk of low growth. (Chapter 2)

**Finding #7: Implementation declined in 2009.** Most schools in the Reading First program have implemented the program "adequately" but, in 2009, the average degree of implementation as measured by the Reading First Implementation Index (RFII) has declined. (Chapter 3)

**Finding #8: Principal participation and teacher program evaluations are strong predictors of achievement.** A Year 6 Report meta-analysis of effect sizes for individual program elements measured using the Reading First surveys found that school-level implementation by the principal and school staff, and teacher evaluation of Reading First, are the two strongest predictors of achievement gains. This suggests that active principal participation and, to some degree, positive teacher perceptions of the program are likely to increase program effectiveness. (Chapter 3)

**Finding #9: The Reading First program has built capacity at state and local levels of education.** Extensive professional development and ongoing support have developed a high level of expertise regarding effective reading instructional methodology among state, district, and school administrators as well as the teaching force. Reading First supported the development of an extensive network of reading coaches with a high level of expertise to support classroom instruction. (Chapter 4)

**Finding #10: Reading First has created a sustainable, comprehensive structure of reading/language arts instruction.** Participants reported positive regard for key elements of the program and the desire to maintain such features as collaborative planning time, a protected time block for reading/language arts, a common curriculum, coaching, professional development, and other features. (Chapter 4)

## Background

Reading First is a federal initiative aimed at improving reading instruction in America. Authorized in 2001 as part of the No Child Left Behind (NCLB) 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. This funding ceased as of the end of fiscal year 2008.

The Reading First program began in California during the 2002-03 school year[1], seven years ago. Its components include:

- Use of a state-adopted reading program
- Access to training programs authorized by state legislation and focused on research-based reading instruction, including Senate Bill (SB) 472 teacher and coach professional development and

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

Assembly Bill (AB) 430 principal professional development, as well as extended follow-up professional development

- Access to assessment tools that measure 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 7-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 seven years of Reading First funding and provides information regarding program efficacy.

*Chapter 1* provides an overview of Reading First and its history, data sources, and the research design. It also discusses demographic characteristics of three cohorts of Reading First schools and how they compare to non-Reading First schools, including teacher characteristics.

*Chapter 2* 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. It provides similar results for the English Learner subgroup.

*Chapter 3* provides Reading First Implementation Index (RFII) statistics. These measure fidelity of Reading First implementation and are computed for each school from data collected from surveys administered to every Reading First teacher, coach, and principal in California.

*Chapter 4* provides a synopsis of lessons learned over the course of the 7-year Reading First evaluation, with particular attention paid to the relative importance of various Reading First program elements.

Attached to this report are appendices (A – F), which give:

- State-level survey results for the teacher, coach and principal implementation surveys (Appendices A, B, and C, respectively)
- Additional charts and graphs showing trends in achievement to supplement Chapter 2 (Appendix D)
- The RFAI calculation description and formula (Appendix E)
- Listings of Reading First schools along with their RFAI and RFII scores for 2006-2009 (Appendix F)

## Two Data Examples from Grade 2: Cohort 1 and Cohort 2

The California Reading First evaluation has, since 2005, approached the problem of measuring program efficacy by comparing growth rates of schools with differing levels of program implementation. In order to display these growth rates graphically, each cohort of schools (defined according to the number of years its schools have received funding, also called its Years in Program or YIP), is assigned its own table and set of trend-line charts. Demographically dissimilar, each funding cohort or YIP has responded to the program in its own way.

As of 2008-09, the schools that were in the first funding cohort (2003) were dropped from Reading First. In 2009, they would have been in the program seven years (YIP 7), but since their scores only go through 2008 they can also be referred to as "YIP 6 Schools in 2008" or more simply "Cohort 1 Schools".[2] The majority of schools analyzed in 2009 were Cohort 2 schools ("YIP 6 in 2009").

The Cohort 1 schools, which included Los Angeles Unified School District (LAUSD), tended to be from large urban school districts. They proved to be quite responsive to Reading First by their second year. The Cohort 2 schools tended to come from more suburban and rural school districts and appear to have been less responsive to Reading First, though they have become more so in recent years. As a way of discussing the effect of Reading First, we compare the two cohorts in terms of the growth rates and trend-lines they have experienced in their grade 2 CST scores for Reading.

Table ES.1.0 presents the grade 2 CST starting and ending scores for schools that have been in the program six years as of 2009 (Cohort 2), where the starting year was 2003. Statistics are reported for the program as a whole, for the program broken out by high, medium, and low implementing schools, for a statistical control group, and for all of the non-Reading First elementary schools in the state.[3]

Table ES.1.1 presents similar scores for schools that were in the program six years as of 2008 (Cohort 1), where the starting year was 2002. Table ES.1.1 was presented in last year's Executive Summary. This table does not include a column for "Medium Implementation Schools", which was added in 2009.

[2] Cohort" is not quite accurate in this context, as it refers to funding provided to LEAs, not schools. For instance, LAUSD was a Cohort 1 LEA, yet some of its schools were added to Reading First years later and are thus assigned to different YIPs.

[3] Numbers reporting change since starting year were rounded and may not appear to be an exact difference between 2003 and 2009 figures.

**Table ES.1.0: CST Metric, 2009, Cohort = 2, YIP = 6, Grade = 2**

| **Year: 2009**<br>**Cohort: 2**<br>**Years in Program: 6**<br>**Grade: 2** | **Reading First Schools** | | | | | All Non-Reading First Elementary Schools |
|---|---|---|---|---|---|---|
| | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | Statistical Control Group (RFII = 25) | |
| Number of Schools | 289 | 73 | 153 | 63 | N/A | 4,025 |
| % Proficient and Above | | | | | | |
| 2003 | 20.7 | 20.9 | 20.7 | 20.8 | 20.8 | 43.0 |
| 2009 | 39.9 | 41.0 | 40.4 | 37.5 | 39.0 | 56.3 |
| Change Since Starting Year | **19.2**[bc] | **20.1**[bc] | **19.7**[bc] | **16.8**[bc] | **18.2** | **13.3** |
| % Below or Far Below Basic | | | | | | |
| 2003 | 46.0 | 45.0 | 46.8 | 45.2 | 46.0 | 25.5 |
| 2009 | 28.7 | 27.1 | 28.9 | 30.2 | 30.1 | 18.5 |
| Change Since Starting Year | **-17.3**[bc] | **-17.8**[bc] | **-17.9**[bc] | **-15.1**[bc] | **-15.9** | **-7.0** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 310.3 | 311.3 | 309.5 | 311.1 | 310.3 | 341.3 |
| 2009 | 333.5 | 335.1 | 334.1 | 330.1 | 330.8 | 357.6 |
| Change Since Starting Year | **23.2**[abc] | **23.8**[bc] | **24.7**[abc] | **19.1**[c] | **20.5** | **16.2** |

**Table ES.1.1: CST Metric, 2008, Cohort = 1, YIP = 6, Grade = 2**

| **Year: 2008**<br>**Cohort: 1**<br>**Years in Program: 6**<br>**Grade: 2** | Reading First Schools | | | | All Non-Reading First Elementary Schools |
|---|---|---|---|---|---|
| | All Reading First Schools | High Implementation Schools (Avg. RFII > 41.4) | Low Implementation Schools (Avg. RFII < 36.0) | Statistical Control Group (RFII = 25.0) | |
| Number of Schools | 253 | 28 | 96 | N/A | 4,057 |
| % Proficient and Above | | | | | |
| 2002 | 15.5 | 14.5 | 15.2 | 15.5 | 37.7 |
| 2008 | 35.8 | 36.6 | 34.8 | 33.0 | 51.2 |
| Change Since Starting Year | **20.3**[abc] | **22.1**[abc] | **19.6**[bc] | **17.5** | **13.5** |
| % Below or Far Below Basic | | | | | |
| 2002 | 54.1 | 53.9 | 54.9 | 54.1 | 30.7 |
| 2008 | 32.1 | 28.9 | 33.2 | 35.5 | 21.1 |
| Change Since Starting Year | **-22.0**[abc] | **-25.0**[abc] | **-21.7**[abc] | **-18.6** | **-9.6** |
| Mean Scale Score Per Student | | | | | |
| 2002 | 300.1 | 299.6 | 299.3 | 300.1 | 333.2 |
| 2008 | 330.0 | 333.2 | 328.9 | 326.2 | 352.4 |
| Change Since Starting Year | **30.0**[abc] | **33.6**[abc] | **29.5**[bc] | **26.1** | **19.3** |

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."
[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

We begin by looking in the "All Reading First Schools" columns in both tables, at the bottom row. We see that in 2009 (Table ES.1.0) the mean scale score gain was 23.2, whereas it was 30.0 in 2008 (Table ES.1.1). The Cohort 1 schools grew 7 scale score points further in six years than the Cohort 2 schools over the same duration. We see also that they ended up with roughly the same score. The Cohort 1 schools scored an average of 330.0 in 2008; the Cohort 2 schools an average of 333.5 in 2009. Thus, the relatively urban Cohort 1 schools entered the program with a greater performance deficit, but made up a greater distance to pull almost even with the Cohort 2 schools.

Moving to the middle columns, we see that for both cohorts the High Implementation schools showed higher growth rates than the Low Implementation schools and the statistical control group. However, this was much more pronounced in 2008 with the Cohort 1 schools than it is in 2009 with the Cohort 2 schools. The mean scale score change for Cohort 1 for high implementing schools is 33.6, significantly higher than the 26.1 scale score change for the statistical control group, a difference of 7 points. In Cohort 2 the high implementing schools grew 23.8 scale score points compared to 20.5 scale score points for the statistical control group, a difference of only 3 points, not statistically significant. The pattern is similar with the "% Proficient and Above" metric and the "% Below or Far Below Basic" metric.

The fact that the Cohort 2 schools have been less responsive to Reading First than Cohort 1 schools was studied and documented in the 2008 Year 6 Report. In that report, it was shown using a statistical meta-analysis of effect sizes that Cohort 2 Reading First effects have generally been smaller than those for the other cohorts.

Nonetheless, Table ES.1.0 also shows that Cohort 2 schools demonstrated statistically significant growth over six years, and that growth was significantly greater than the growth of the non-Reading First schools over the same period. This proves particularly true when we look at the "% Below or Far Below Basic" achievement metric. In non-Reading First schools, the movement of students out of the bottom performance levels was -7.0 (i.e., the percentage of students in the bottom categories dropped by 7 points); in Reading First schools the drop was -17.3, more than twice as much. This is a persistent pattern across the Reading First evaluation.

Stepping back, we see that after six years Cohort 2 Reading First schools moved approximately 20% more of their students into "Proficient" territory. For schools that in 2003 had among the lowest scores in the State, this is an impressive and significant achievement.

Figures ES.1.0 – ES.1.3 display the implementation-articulated trend-lines for each cohort that correspond to Tables ES.1.0 and ES.1.1.

**Figure ES.1.0: Cohort 2, CST % Proficient & Above, 2009, YIP = 6, Grade = 2**

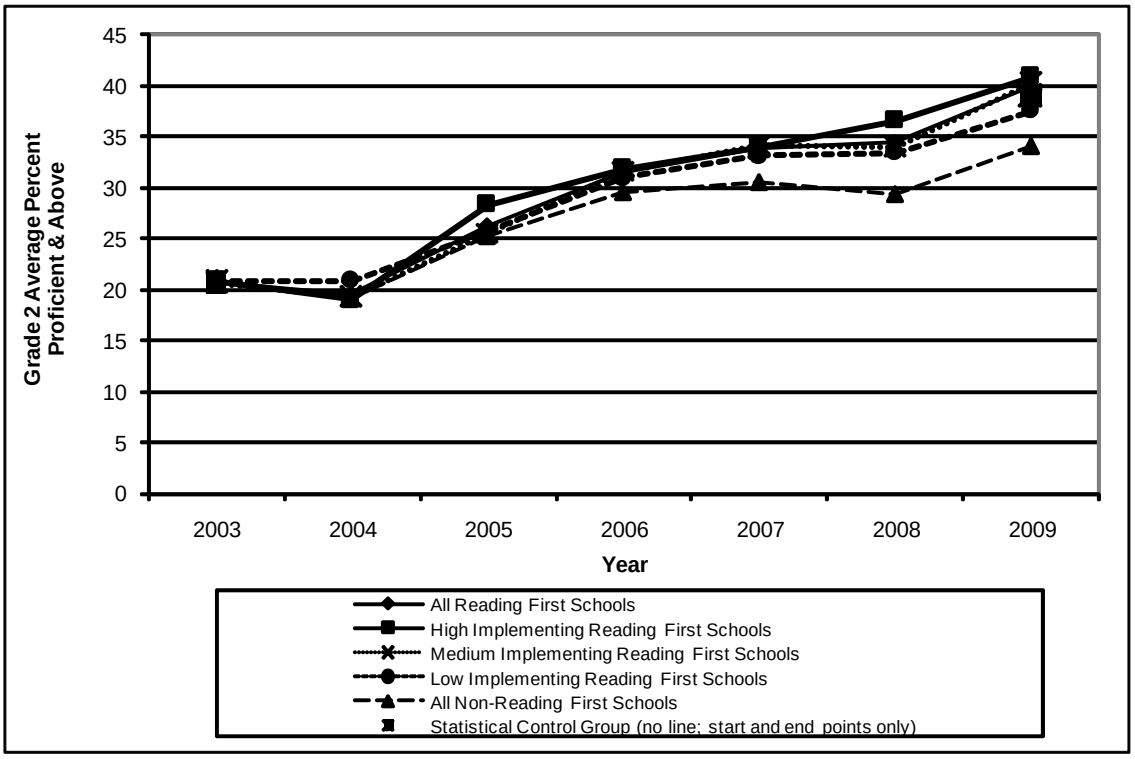



**Figure ES.1.1: Cohort 1, CST % Proficient & Above, 2008, YIP = 6, Grade = 2**

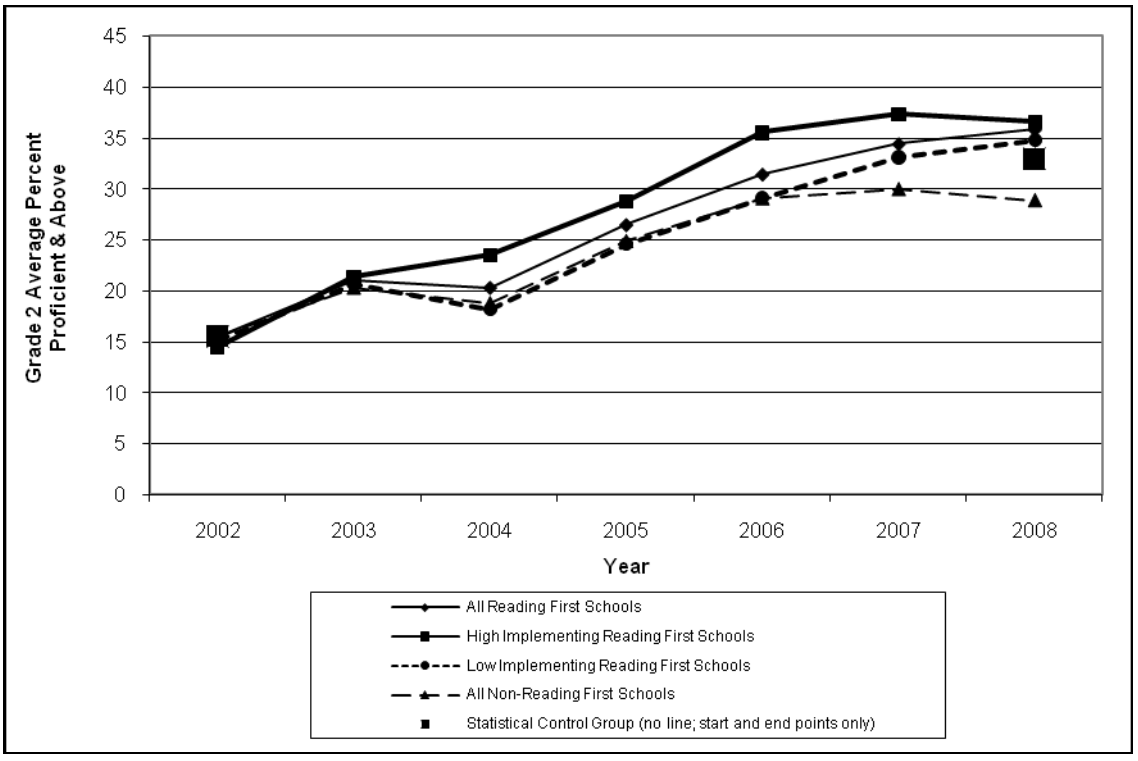


**Figure ES.1.2: Cohort 2, CST % Below & Far Below Basic, 2009, YIP = 6, Grade = 2**

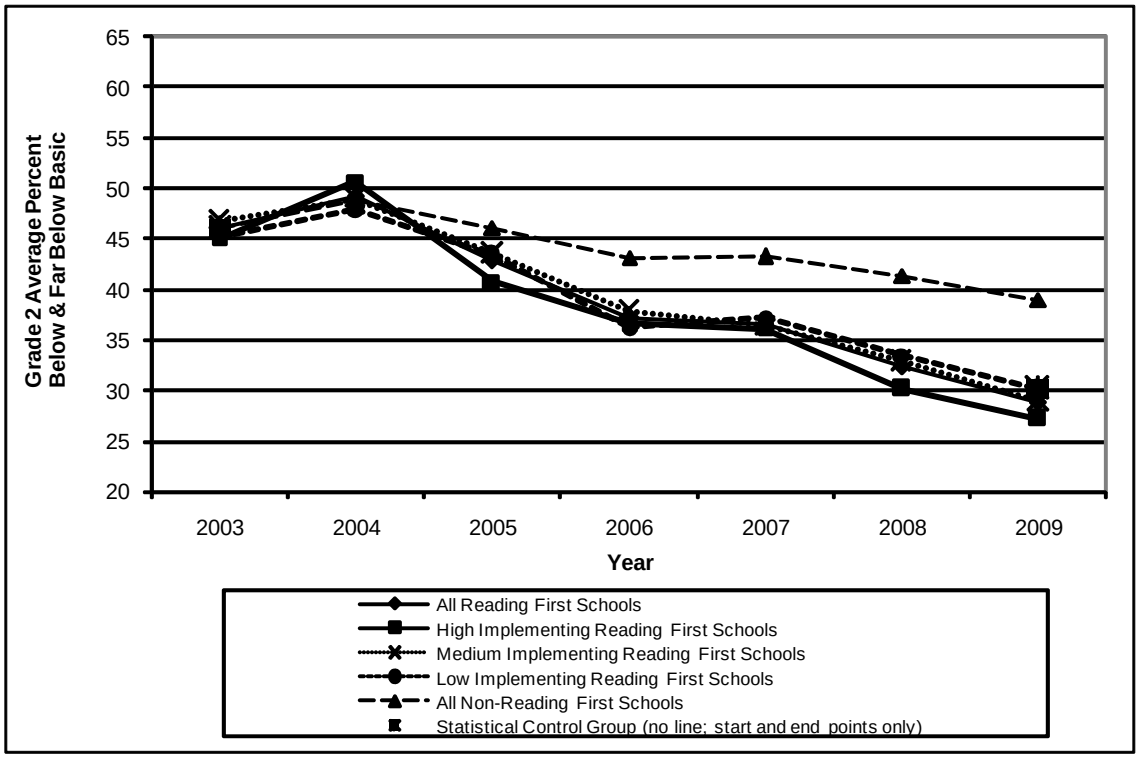
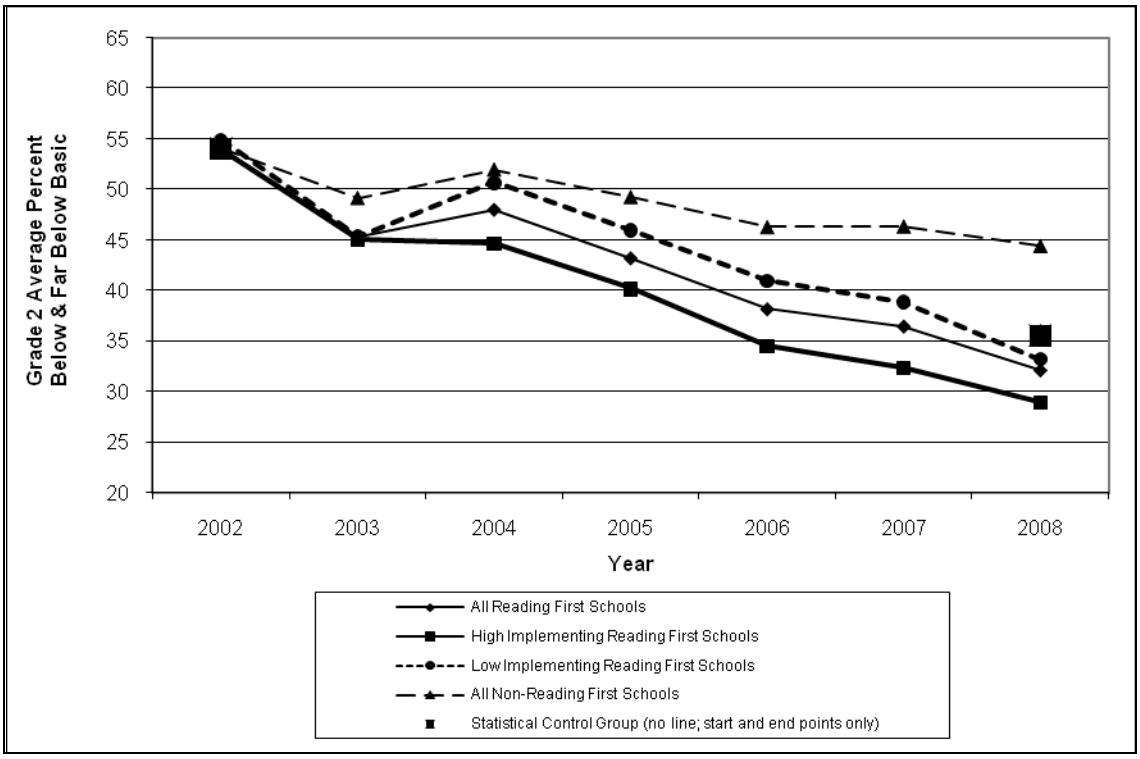



**Figure ES.1.3: Cohort 1, CST % Below & Far Below Basic, 2008, YIP = 6, Grade = 2**



## Policy Recommendations

In this final report on California's Reading First program, we draw on seven years of data to present recommendations for sustaining the effects of the program. Consistent and cumulative evidence indicates the significant impact of the Reading First program on the students whose teachers, coaches and administrators participated in the program and received curriculum materials and support. Yearly reports have shown that the Reading First program has improved reading achievement for students in Reading First schools, students in high implementing schools, and subgroups of students including English learners and students of socio-economic disadvantage. Additionally, teachers, coaches and principals have reported the importance of key components of the Reading First program and their desire to sustain them. Also, we have presented evidence that many of non-Reading First schools have, at least in part, implemented program elements required of Reading First schools and demonstrated reading gains, showing that the impact of the program has reached beyond the cohorts of participant schools.

The policy recommendations listed below are based on evidence gathered in seven years of evaluation of the Reading First program. As the Reading First program comes to an end and we look to the future to consider how to continue the momentum of improved reading achievement in the state, we provide final policy recommendations that may assist the state in sustaining the benefits of the Reading First program in California in the coming years.

Maintain a Strong Focus on Reading Achievement

In all program activities, including professional development, classroom instruction and data analysis, the Reading First program has promoted reading achievement as its primary goal. California has established challenging Reading/Language Arts standards and most Reading First schools were vigilant in promoting and monitoring students' progress toward them. Initially, many were skeptical about the ability to achieve such goals and discussed in interviews and surveys that devoting such effort and time to teaching "scripted" reading curricula, adhering to "rigid" pacing plans, and conducting such extensive assessment would take time and effort away from valuable instructional activities, lessons and materials that had previously been in place. Some questioned the feasibility of improving achievement in schools that had chronic low achievement, and particularly with English learners. In the last few years of the Reading First evaluation, surveys and narrative responses of teachers, coaches and principals revealed that perceptions had changed significantly over time and were generally positive, despite specific suggestions for improving the program or curriculum. Many participants noted the improved reading skills of students as worth the effort. Participants also noted the importance of having a common focus across classrooms and throughout the district. The achievement gains realized in Reading First schools are an unmistakable

indicator that the focus on achievement has not only improved outcomes for students, but may also represent a transformation of reading instruction in California.

Fidelity of Implementation

The importance of strong implementation of research-based reading instruction via a state-adopted curriculum is a strong and consistent finding of the seven-year evaluation of California's Reading First program. The Reading First assurances have provided guidelines and the state has developed a strong infrastructure to support implementation. The 2008 Reading/Language Arts Framework and textbook adoption outlines the key elements of a research-based reading program, and may aid in ensuring continued implementation of research-based reading instruction. Our recommendation is that the state should continue to facilitate the supports that have been in place to ensure that the newly adopted programs will be fully implemented and used as intended. Such elements as the Reading Technical Assistance Centers (RTACs), coaching, initial and advanced levels of professional development and support of the use of ongoing data analysis procedures may assist in continuing the implementation of practices. Simply stated, a district's adoption of a new state-adopted Reading/Language Arts curriculum does not ensure fidelity of implementation. We strongly encourage California to consider how to continue to support fidelity of implementation and maintain the capacity built by Reading First.

Maintain Support Structures

Reading First has built capacity in the state through developing support mechanisms. Findings in the California Reading First evaluation reports have repeatedly highlighted the value and benefits of deep and ongoing professional development, a highly qualified coaching force, knowledgeable and involved site administrators, time for collaborative lesson planning, and the use of data to guide instruction. These elements have provided important and necessary support. Investing in maintaining the support structure that has been built with Reading First funding would support a continued focus on achievement and implementation.

Professional development is one support structure created through the Reading First program that has been highly important and effective in improving instruction and student outcomes. With Reading First, California has developed a network of knowledgeable providers of professional development. This program has shown that professional development must not only provide teachers with knowledge of research-based strategies, but it must also be specific to the curriculum. Additionally, the ongoing professional development and the advanced levels of training have helped to create a highly qualified teaching force. It is important for California to invest in maintaining this standard of quality for future teachers.

Coaching is another aspect that has been strongly supported through Reading First funds. The Year 5 report included a chapter highlighting the importance of the coaching force and their role in transforming reading instruction throughout the state. Through Reading First support, the state has developed a coaching force with expertise in research-based instruction, curriculum, data analysis, and collaboration. It is important for California to consider maintaining the investment in coaching to maintain a high quality of reading instruction in our state.

Retain "Program Coherence" as a Guiding Principle

Much of the success of Reading First resides in its program coherence. The assurances outlined essential elements that must be in place to accomplish the program goal of improving reading achievement. As districts and schools move to the use of newly adopted curricula, it is our recommendation that the state and local districts maintain program coherence. District-wide and school-wide use of curriculum and assessments promote a common focus and basis for professional communication. It is important for future efforts to be internally consistent, well-focused, cohesive, and based on rigorous standards.

Maintain a Focus on Improving Reading Outcomes for English Learners

The Reading First program has helped teachers to develop the expertise to adjust their instruction to meet the reading and language learning needs of English learners. Prior to Reading First, there was a widespread belief that English learners could not meet grade level standards due to the inherent challenges in learning to read while learning the English language. In this and previous evaluation reports, we have demonstrated that English learners in Reading First schools outperform English learners in non-Reading First schools in reading gains. The 2008 Reading/Language Arts Framework provides guidance for maintaining a focus on English language development through a comprehensive Reading/Language Arts program. We recommend that California continue to strive for excellence in providing appropriate Reading/Language Arts instruction to English learners through high-quality instruction.

Maintain Strong Policy

The Reading First program required states to adopt policy guidelines related to the Reading First assurances. These policies have thus far proven to be effective in raising the quality of reading instruction in California. We urge the state to maintain a strong policy to support continuation of the principles and practices promoted by Reading First, such as professional development, program coherence, leadership support, coaching assistance, and monitoring of student progress.

## Final Remarks about Reading First

Over seven years of implementation, California's Reading First initiative has transformed reading instruction in thousands of classrooms. Reading First funding was focused on improving reading outcomes for students in socio-economically disadvantaged areas and in schools with chronic low achievement. The cumulative reports of reading outcomes in California's Reading First schools show that Reading First has accomplished that goal. Reading achievement has risen steadily in Reading First schools according to various achievement metrics used in the evaluation reports in comparison to comparison groups and a statistical control group. There has been a steady migration of students out of the Below Basic and Far Below Basic achievement groups and into the Proficient and Advanced groups in Reading First schools. These findings also hold true for the subgroup of English learners and beyond the K-3 grades. This evaluation has yielded a school level Reading First Achievement Index, or RFAI. Over time, the RFAI has steadily risen and has proved useful as a measure of significant progress for schools participating in Reading First. Though California has not yet achieved the goal of ensuring that every student reaches proficiency by the end of grade 3, the results of this evaluation indicate a substantial step toward that goal. It is our recommendation that the state make every effort to sustain this trend through continued vigilance in improving reading instruction in the early grades.

It is impossible to understand the scope of the impact of Reading First without examining implementation in relation to achievement. This seven-year evaluation process has resulted in the ability to examine implementation at macro and micro levels. The Reading First survey data have yielded a Reading First Implementation Index, a measure of each school's level of implementation. A consistent finding of the past five years of reporting has shown a strong and positive correlation between implementation and achievement. There is no doubt that achievement rises when implementation of the program is strong.

This evaluation also examined specific elements of implementation. The findings demonstrate that the Reading First program has led to the development of a well-integrated structure and process of providing reading/language arts instruction in California. The program elements outlined in the Reading First assurances are integral parts to a whole that is more than the sum of its parts. The use of state-adopted curricula, professional development, coaching, ongoing data analysis and collaboration, leadership support, protected time blocks, and other program elements together form an integrated reading program that has had a strong impact on reading achievement in the state. It is important for state leadership and policy makers to consider the importance of sustaining these program elements as interconnected and essential ingredients of an effective reading/language arts program.

The findings of the California Reading First Evaluation have not ceased to be relevant to public policy discussions simply because the funding has ended. While this evaluation is specifically applicable only to

the domain of California schools participating in Reading First, we suggest that the findings accumulated over seven years of research may generalize beyond that domain. Therefore, we believe the following scenarios are plausible for schools that would adopt the principles of Reading First:

- Elementary schools would likely move a substantially higher percentage of students out of the Below and Far Below Basic performance levels.
- The application of Reading First principles in the upper grades (grades 4 – 12) would likely result in substantially improved scores for low-performing students.
- A substantially larger percentage of English Learners would score "Proficient" on the CSTs and may therefore be eligible for reclassification.
- Reading First-like assurances, such as extensive professional development, ongoing support, and others that ensure full implementation, may be similarly effective if adapted to other subject areas such as elementary school mathematics.

The evidence presented in the external evaluation reports supports the idea of sustaining program elements of Reading First. It also suggests that a similar support network and extensive reform effort might also apply beyond the K-3 grades, and perhaps beyond the domain of reading instruction. For that reason, we believe that Reading First-like program elements deserve serious consideration in State and Federal educational policy discussions, regardless of domain.

# TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF APPENDICES

# CHAPTER 1: INTRODUCTION AND DEMOGRAPHICS

## Overview of California's Reading First Program

Reading First is a federal initiative that was authorized in 2001 as part of the No Child Left Behind Act (NCLB). This program, intended to improve reading outcomes in the nation, promotes the use of instructional practices and curricula based on scientifically based reading research in grades K-3. On August 23, 2002, the State of California was approved to receive approximately $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. With Reading First funds, California has established a system to provide training, assist local educational agencies (LEAs) in acquiring curricular materials, monitor progress toward goals, and provide technical assistance to participating schools and school districts. This report provides an external evaluation of California's implementation of Reading First and student reading achievement for seven years of implementation from academic year 2002-03 to 2008-09, which is the last year of Reading First funding.

The California 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. A subcommittee of the Partnership, the Evaluation Advisory Group (EAG), including designees of the members, advises the external evaluator. The California Technical Assistance Center (C-TAC) has responsibility for the statewide technical assistance program and oversight of the Regional Technical Assistance Centers (R-TACs) in providing regional and local support to LEAs. The C-TAC also coordinates the statewide network of professional development programs for teachers and site administrators through the Reading Implementation Centers (RICs).

The California Reading First Plan is based on a series of Assurances that are implemented by the LEAs. With these assurances, California's Reading First program is designed to ensure full implementation with fidelity to a comprehensive research-based reading program. Here, we briefly describe the assurances and program elements designed to address them.[1]

[1] For a complete description of the program elements, we refer the reader to previous evaluation reports, available at: http://eddata.com/resources/publications/ and the state's Reading First plan, available at: http://www.cde.ca.gov/nclb/sr/rf/.

Vision Statement

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

Curriculum

Participant LEAs are required to use one of California's two state-adopted reading curricula: SRA/McGraw-Hill's *Open Court Reading 2000* or *2002* (OCR) or the Houghton Mifflin *Reading: A Legacy of Literacy 2003* (HM). The Reading First program has provided extensive support for LEAs in the implementation of the adopted curricula. In the 2004-05 school year, California's Reading First program began offering support for LEAs with "waivered" classrooms, that is, classrooms offering a bilingual instruction model using Spanish-language versions of the adopted curricula. California law (Proposition 227) mandates instruction in English for all students unless parents sign a waiver specifically requesting bilingual instruction. The two state-adopted Spanish language reading programs are: SRA/McGraw Hill's *Foro abierto para la lectura* and Houghton Mifflin's *Lectura: Herencia y futuro.* Students receiving bilingual reading instruction in Spanish and English must transition from bilingual instruction to English instruction, and take the English Standardized Testing and Reporting (STAR) English Language Arts Content Standards Test (CST) at the end of grade 2 and grade 3. Regardless of the LEA's selected curriculum, 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 in every Reading First school. Also, LEAs are beginning to plan and implement extensive intervention with those K-3 students who need an additional 30 minutes of instruction. The intervention materials are approved by the SBE as scientifically research-based.

Professional Development

LEAs must assure that all K-3 teachers in Reading First schools annually participate in 40-hour training focused on the adopted core reading program. Year 1 teachers attend a state-approved training as mandated in Senate Bill (SB) 472. For Years 2-6, the LEAs must provide advanced levels of professional development, either provided through trainings developed by the C-TAC and delivered through the Reading Implementation Centers (RICs), or provided by the LEA. In addition, LEAs must provide access to these trainings for their K-12 special education teachers who are teaching K-3 reading, using either the LEAs' adopted core or intensive intervention reading program. LEAs are encouraged to provide continuous training to principals with the use of the C-TAC developed administrator modules (1-3 hours)

on implementing the adopted reading program and providing instructional leadership. Training of LEA trainers on these modules is provided by the C-TAC.

Curriculum-Embedded Assessment

For program monitoring, LEAs are required (since 2005-06) to use curriculum-based assessments conducted every 6 to 8 weeks. Teachers, administrators, and coaches use the data to make instructional adjustments and to identify individual students who need extra assistance. The results of the End-of-Year (EOY) tests—the curriculum-based assessment administered at the end of the school year—are required to be submitted to the State by each school. The results of these assessments are used as part of the Reading First Achievement Index (RFAI; see Chapter 2 of this report).

Collaborative Teacher Meetings

All Reading First schools are required to hold regular grade-level meetings twice a month to provide an opportunity for teachers to work together to refine their implementation of the program. 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. This evaluation includes a district action plan for the subsequent year and each school's action plan for its first tri-semester based on student achievement data and principal, coach, and teacher recommendations. In addition, district personnel must assure that the Reading First program is well coordinated with other programs such as Title I, 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 and ensure that these experts are adequately trained. Coaches offer site-specific support for implementation of the LEA's adopted reading curriculum and effective instructional strategies. The C-TAC has provided these experts two Coach Institutes annually for in-depth training and a Leadership Program for selected experts in partnership with a California university. Additional training for new coaches is provided by the RICs, and support for both coach and coach coordinators is offered by the R-TACs.

Site Leadership

The site administrator's role is to support the full implementation of the school's adopted reading program and the state's Assurances. Administrators must attend the state's 40-hour AB 430 training program to become fully knowledgeable of the reading program and participate in 40 hours of aligned activities within a two-year period. LEAs are also required to provide on-going training annually and are encouraged to use the C-TAC provided administrator modules.

Program Coherence

Reading First schools must ensure that any supplemental programs or materials are fully aligned with the adopted reading program, if using Reading First funds. LEAs are encouraged to use the SBE-approved intervention and diagnostic assessment materials that offer extensive intervention. All categorical programs such as Language Acquisition, Title I, School Improvement, and Special Education programs, must be coordinated with the core program.

State Leadership

The CDE has designated key personnel to oversee and facilitate the administration of Reading First grants to LEAs, the contract with the external evaluator, and communications and legislation for the Reading First program. The SBE serves as the state educational agency for Reading First and works collaboratively with the CDE and the governor's office to develop and approve policy decisions regarding Reading First.

Technical Assistance

In addition to the statewide technical assistance programs provided by the C-TAC, the R-TACs, housed in county offices of education throughout the state, work directly with LEAs for full implementation of the Assurances. Some of their required activities include conducting classroom observations with LEAs' leadership team members; offering workshops on assessment, internal evaluation reporting, and interventions; and providing consultation on next steps to be taken by LEAs to meet goals of Reading First.

LEA Cohorts

California has now completed seven years of implementation of the Reading First program. LEAs have been added to the program in cohorts. The first year, 2002-03, can be characterized as a start-up year because LEAs did not have a full year in which to implement. Cohort 1 (329 schools) received funding and implemented the program for approximately five and one-half years, through 2007-08. LEAs in Cohort 2 (359 schools) were selected for funding in 2003-04 and implemented the program through 2008-

09. Cohort 3 (136 schools) was added in 2004-05. A small number of LEAs were added in 2006-07 to form Cohort 4 (20 schools). A total of 498 schools in 110 LEAs are included in this Reading First Year 7 report.

## California Reading First Year 7 Evaluation Study Design

The California Reading First Plan includes an annual external evaluation to study the implementation of the program and the resulting student achievement. Educational Data Systems (EDS[2]) has been the contractor for the Reading First evaluation study for each year of the program and has completed prior reports for Years 1 through 6. This current report represents the Year 7 evaluation report, and will include outcomes from the 2008-09 academic year and cumulative effects.

This report is guided by five research questions as stated in the scope of work for the external evaluation study. Two questions address program implementation:

1. How well did participating LEAs and schools implement their Reading First grants in accordance with California's Reading First plan?
2. What resources, support, and professional development activities are district-level administrative staff, school site administrators, and classroom teachers receiving in implementing the Reading First grants?

Three additional questions focus on the impact of Reading First:

3. What is the impact of the Reading First program on K-3 students in participating districts and schools?
4. What evidence is there that the Reading First program has improved the effectiveness of participating schools and districts?
5. Have any unintended consequences resulted from the implementation of the Reading First program?

The conceptual framework below provides an overview of the evaluation study design. It displays how the Reading First data can be organized into three types: a) school and district characteristics; b) achievement data; and c) implementation data. The school and district characteristics are described later in this chapter, with data drawn from state and federal databases, including the California Basic Educational Data System (CBEDS) file and the demographic sections of the California English Language Development Test (CELDT) and STAR files, and school demographic data from the National Center for

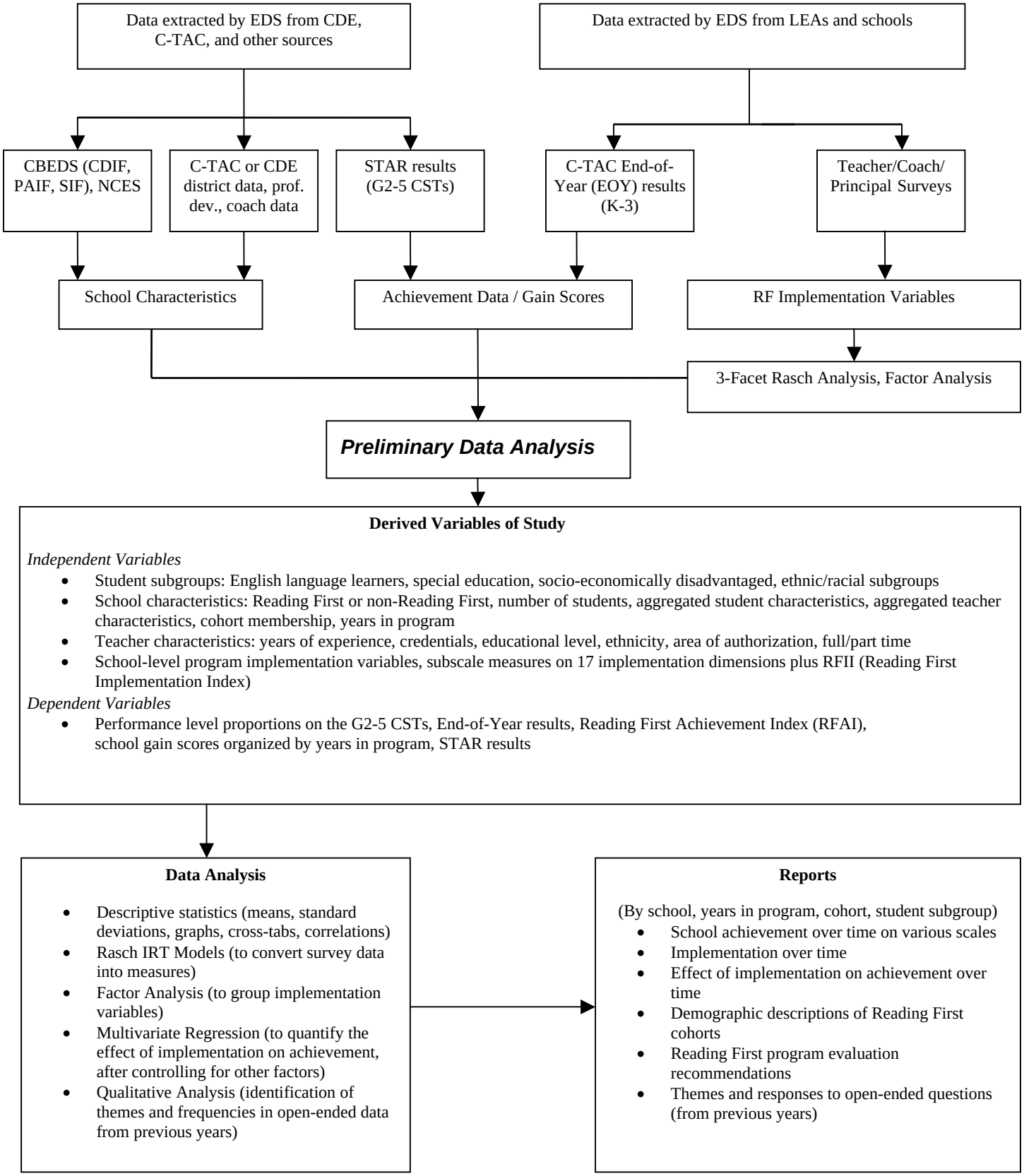
[2] EDS is a registered trademark of Hewlett-Packard Development Company, LP. However, in the context of this document, EDS refers exclusively to Educational Data Systems, Inc.

Education Statistics (NCES). The achievement data consist of school-level California Standards Test (CST) scores in a performance level metric and a scale score metric, school-level standardized test scores drawn from the California Achievement Test, CAT/6 (which ceased being administered in California in 2008-09), and C-TAC End-of-Year (EOY) scores (eight subtests for kindergarten and Oral Fluency for Grades 1-3) for both English and Spanish. The implementation data will, as before, be drawn primarily from the teacher, coach, and principal surveys that are administered to all Reading First schools annually.

The conceptual framework indicates the types of analysis employed. The achievement data are analyzed according to the percentage of students in a school at a given performance level and the average school scale score. An additional analysis yields the Reading First Achievement Index (RFAI), which combines the STAR and EOY data. To examine implementation, a multi-facet Rasch model is used to combine the teacher, coach, and principal surveys into a coherent measurement framework. The variables used and the analyses have been conducted in accordance with recommendations of the Reading First EAG. Unlike previous reports, the Year 7 report does not include qualitative analyses of open-ended survey responses.

While the Year 7 report summarizes findings from previous years, it does not include the qualitative analyses and other sub-studies covered in previous reports. We refer the reader to the Year 5 and Year 6 reports especially, and to the Reading First Supplemental Survey Report (March, 2008).

**Figure 1.1: Conceptual Framework – Year 7**



## Comparison Group

Past reports have included comparison groups against which to gauge the relative effects of the Reading First program. Past efforts included using "Reading First Eligible" schools, or those that would likely meet socio-economic and achievement criteria for Reading First if their LEA were included in the program. However, in the Year 3 report, it was demonstrated that these schools were too demographically dissimilar to Reading First schools to serve as a legitimate comparison group. The Year 4 report also discussed problems with creating a demographically matched group of schools due to differences in starting point for their achievement as compared to Reading First schools. An additional difficulty with using comparison groups is the statewide effort to improve reading instruction in non-Reading First schools. It is likely that state-adopted curricula, state-funded professional development, and other elements of Reading First are present in many non-Reading First schools, making it impossible to discern the true impact of the Reading First program. Indeed, a 2008 survey of LEAs eligible for but not participating in Reading First found that "Almost 60% of these LEAs use [Open Court and Houghton-Mifflin] programs exclusively in at least 67% of their schools[3]." Thus, the instructional materials and practices used in most eligible non-Reading First classrooms are likely to closely mirror those used in Reading First classrooms. This similarity in reading programs is matched by similar trends in student achievement, although Reading First schools have shown more substantial growth.

For a more complete discussion of the difficulties with constructing a valid comparison group of schools, the reader is referred to the Year 4 report. While this report does not use non-Reading First comparison schools, analyses are conducted using a statistically derived comparison group, the "statistical control group", as described in the Year 4 and Year 5 reports, and in Chapter 2 of this report.

## Demographic Characteristics of Reading First Schools

California's Reading First program began in the 2002-03 academic year. During subsequent years additional LEAs were funded. The Year 4 report distinguished between cohort groupings based on the year the LEAs received funding and "Years in Program" (YIPs), for school-level analyses. A small number of schools included in Reading First databases do not have the same years of participation as their assigned LEA cohort due to gaining and losing schools in cohorts for various reasons such as schools merging, closing, or replacing other schools dropped from the program. This is a relatively small number of schools, but for accuracy of school-level analyses, this report will use the YIP for achievement and implementation analyses in Chapters 2 and 3. For demographic analyses included in this chapter, we use LEA Cohorts to describe the characteristics of participants.

[3] See The Reading First Supplemental Survey Report, March 2008 at *www.eddata.com/resources/publications/*.

The following is a summary of the LEA cohorts, the typical YIP for that cohort, and the number of schools (a total of 498 in the 2008-09 academic year) for each cohort included in the current report. Cohort 1 LEAs did not receive any funding in 2008-09.

(a) Cohort 1, first funded in 2002-03, with 13 LEAs (0 schools in current report); YIP 6
(b) Cohort 2, first funded in 2003-04, with 60 LEAs (349 schools in current report); YIP 5
(c) Cohort 3, first funded in 2004-05, with 27 LEAs (129 schools in current report); YIP 4
(d) Cohort 4, first funded in 2006-07, with 10 LEAs (20 schools in current report); YIP 2

The demographic data included in this chapter are extracted from the STAR research file published on the CDE website.[4] In the STAR file, student-level data have been aggregated and presented at the school level. Therefore, the smallest unit of analysis in this chapter is the school. Other sources of data include the Professional Assignment Information Form (PAIF) file, the CBEDS file, and the NCES file.

Socio-Economically Disadvantaged (SED) Students in Reading First

According to the Reading First legislation, funding is earmarked for schools in the state with high numbers of students of low socio-economic status and a history of low achievement. Therefore, it is not surprising that the Reading First schools have a higher number of SED students compared to all elementary schools in the state. Table 1.1 displays the percentage of SED students in each cohort of Reading First for the starting year (varies by cohort) and for 2009. Table 1.1 also includes the 2004 and 2009 percentage of SED students in all elementary schools in the state.

Cohort 2 had 85.46% and Cohort 3 had 81.72% SED students in 2009. Cohort 4 had the lowest percentage of SED students, 80.84%.

English Learners (ELs)

In 2009, Reading First schools also had higher percentages of ELs than the category of All Elementary Schools. The percentage of ELs in Cohorts 2 and 3 was 55.65% and 59.62% respectively. Cohort 4, with 33.26% ELs, more closely resembled the statewide figure of 29.27%.

Students with Disabilities

In 2009, the percentage of students with disabilities was reported as 5.62% for Cohort 2, 8.77% for Cohort 3 and 6.47% for Cohort 4. This varies only slightly from the statewide percentage of 8.79%.

[4]The STAR research file used for the 2008-09 data was the version obtained by EDS on September 22, 2009, referred to as "P2."

Ethnicity Breakdown of Reading First Schools

Table 1.1 shows the percentage of students in each ethnicity category for each cohort, as compared to statewide figures. As compared to the All Elementary Schools category, Reading First schools in general had significantly higher percentages of Hispanic students and significantly lower percentages of White students. Cohorts 2 and 3 had significantly higher percentages of Hispanic students than Cohort 4.

**Table 1.1: Student Demographic Data, 2004-2009**

| | **Reading First Schools** | | | | | | **All elementary schools**[1] | |
|---|---|---|---|---|---|---|---|---|
| | **Cohort 2** | | **Cohort 3** | | **Cohort 4** | | | |
| | **2004** | **2009** | **2005** | **2009** | **2007** | **2009** | **2004** | **2009** |
| Number of Schools | 343 | 349 | 136 | 129 | 19 | 19 | 5977 | 6452 |
| SED (%) | 82.69 | 85.46 | 85.15 | 81.72 | 73.37 | 80.84 | 53.28 | 52.13 |
| EL (%) | 52.97 | 55.65 | 57.50 | 59.62 | 31.21 | 33.26 | 29.34 | 29.27 |
| Students with Disabilities (%) | 8.02 | 5.62 | 7.05 | 8.77 | 7.89 | 6.47 | 11.06 | 8.79 |
| African American (%) | 8.85 | 7.24 | 6.56 | 6.43 | 14.05 | 14.05 | 7.61 | 6.98 |
| American Indian (%) | .98 | .74 | .77 | .78 | 7.68 | 7.84 | 1.33 | 1.28 |
| Asian (%) | 4.57 | 3.99 | 1.14 | 1.08 | 1.68 | 1.74 | 7.46 | 7.30 |
| Filipino (%) | 1.66 | 1.35 | 1.26 | .98 | 4.63 | 3.53 | 2.33 | 2.33 |
| Hispanic (%) | 72.02 | 75.00 | 77.12 | 77.19 | 50.53 | 52.37 | 42.57 | 42.15 |
| Pacific Islander (%) | .81 | .85 | .54 | .60 | .63 | .58 | .66 | .60 |
| White (%) | 9.57 | 6.80 | 11.20 | 9.56 | 19.32 | 16.00 | 33.90 | 29.12 |

[1] The group "All Elementary Schools" *includes* Reading First schools in this chapter. In Chapter 2, "All Non-Reading First Elementary Schools" *excludes* Reading First schools.
Data source: California Standardized Testing and Reporting (STAR) research file. The number of schools included on this table may differ from other tables because STAR data is obtained beginning with grade 2 and therefore does not include schools with enrollment only for grade K-1.

Urban-Rural Distribution

Table 1.2 presents the prevalence of urban and rural designations in the existing Reading First cohorts, separately and combined. In last year's Year 6 Report, it was evident that most of the schools in Cohort 1 were designated as large or mid-sized cities. Here we see that Cohort 2 included primarily large, mid-size and both large and mid-size suburb categories. Cohort 3 consisted mainly of suburbs of large cities and rural designations, resulting in high levels of migrant students. Cohort 4 was evenly split between urban and rural designations.

**Table 1.2: Urban-Rural Distribution for Reading First Schools 2009**

| School Location | Cohort 2 | | Cohort 3 | | Cohort 4 | | All cohorts | |
|---|---|---|---|---|---|---|---|---|
| | No. of Schools | % of schools | No. of Schools | % of schools | No. of Schools | % of schools | No. of Schools | % of schools |
| City: Large | 126 | 36.3 | 11 | 8.7 | 1 | 5.0 | 138 | 27.9 |
| City: Midsize | 51 | 14.7 | 14 | 11.0 | 6 | 30.0 | 71 | 14.4 |
| City: Small | 17 | 4.9 | 13 | 10.2 | 2 | 10.0 | 32 | 6.5 |
| Suburb: Large | 96 | 27.7 | 58 | 45.7 | 1 | 5.0 | 155 | 31.4 |
| Suburb: Midsize | 17 | 4.9 | 0 | .0 | 1 | 5.0 | 18 | 3.6 |
| Suburb: Small | 3 | .9 | 4 | 3.1 | 0 | .0 | 7 | 1.4 |
| Town: Fringe | 16 | 4.6 | 0 | .0 | 0 | .0 | 16 | 3.2 |
| Town: Distant | 2 | .6 | 12 | 9.4 | 3 | 15.0 | 17 | 3.4 |
| Town: Remote | 3 | .9 | 0 | .0 | 0 | .0 | 3 | .6 |
| Rural: Fringe | 12 | 3.5 | 8 | 6.3 | 3 | 15.0 | 23 | 4.7 |
| Rural: Distant | 2 | .6 | 4 | 3.1 | 2 | 10.0 | 8 | 1.6 |
| Rural: Remote | 2 | .6 | 3 | 2.4 | 1 | 5.0 | 6 | 1.2 |
| *Total* | 347 | 100.0 | 127 | 100.0 | 20 | 100.0 | 494 | 100.0 |

[1] The percent of the schools in that cohort in a particular type of location.
Data source: National Center for Education Statistics (NCES). Urban-Rural data for four schools participating in Reading First in 2008-09 were not available.

## Teacher Qualifications in Reading First Schools

Table 1.3 provides information about Reading First teachers' credentials and teaching experience as derived from the CBEDS and PAIF research files. This table shows the percentage of teachers falling into each educational degree category by cohort and year, as well as teachers' average years of experience. The issue of teacher qualifications is an important one, given the focus of the NCLB on ensuring that schools are staffed with highly qualified teachers. To more easily compare cohorts to each other, 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.

**Table 1.3: Elementary Teacher Credentials and Experience 2004 – 2009**

| | Reading First Schools | | | | | | All Elementary Schools[2] | |
|---|---|---|---|---|---|---|---|---|
| | Cohort 2 | | Cohort 3 | | Cohort 4 | | | |
| | 2004 | 2009 | 2005 | 2009 | 2007 | 2009 | 2004 | 2009 |
| Number of Schools | 359 | 349 | 135 | 129 | 20 | 19 | 6186 | 6199 |
| PhDs (%) | .66 | .71 | .59 | .74 | 1.75 | 1.11 | .49 | .71 |
| Masters plus 30 or more semester units (%) | 13.73 | 15.69 | 16.28 | 15.30 | 14.06 | 16.61 | 14.48 | 16.87 |
| Masters (%) | 16.86 | 22.58 | 16.63 | 15.75 | 17.40 | 19.39 | 17.74 | 19.87 |
| Bachelors plus 30 or more semester units (%) | 49.36 | 49.25 | 47.05 | 54.13 | 52.09 | 50.27 | 53.16 | 49.38 |
| **Total Advanced Degrees (%)** | 80.61 | 88.23 | 80.56 | 85.91 | 85.30 | 87.38 | 85.87 | 86.84 |
| Bachelors (%) | 19.30 | 11.75 | 19.33 | 13.55 | 14.55 | 12.62 | 14.01 | 12.85 |
| Less than Bachelors (%) | .10 | .02 | .22 | .00 | .15 | .00 | .09 | .17 |
| **Total Bachelors or less (%)** | 19.40 | 11.77 | 19.55 | 13.55 | 14.69 | 12.62 | 14.10 | 13.02 |
| Fully Credentialed Teachers (%) | 93.73 | 98.48 | 92.05 | 96.24 | 98.03 | 96.94 | 97.14 | 97.15 |
| Weighted Teacher Qualification[1] | 2.26 | 2.44 | 2.31 | 2.34 | 2.36 | 2.43 | 2.34 | 2.42 |
| Average years teaching | 11.25 | 12.65 | 11.40 | 12.31 | 13.06 | 12.51 | 13.00 | 13.27 |

[1] The Weighted Teacher Qualification is computed as follows: The percentage of teachers with PhDs is given a weight of 5; the percentage of teachers with Masters plus 30 or more semester units is given a weight of 4; the percentage of teachers with Masters is given a weight of 3; the percentage of teachers with Bachelors plus 30 or more semester units is given a weight of 2; and the percentage of teachers with Bachelors is given a weight of 1. The weighted degree percentages are summed, and then divided by 100, to reach the Weighted Teacher Qualification. This index spans from 1 (lowest qualification) to 5 (highest qualification).
[2] In this chapter, the group "All Elementary Schools" *includes* Reading First schools. In Chapter 2, "All Non-Reading First Elementary Schools" *excludes* Reading First schools.
Data source: California Basic Educational Data System (CBEDS) file.

## Conclusions

This chapter yields the following findings:

- In past reports, we have discussed inherent difficulties and limitations in establishing adequate comparisons using non-Reading First schools. To address these issues, a "statistical control group" has been used since the Year 4 Report (2006) to assess the Reading First effect.
- The term "Cohorts" refers to the year a Reading First LEA (district) accepted funding. The term "Years in Program" (YIP) indicates the number of years a school within an LEA cohort has actually been implementing the program. For demographic analyses, this report uses cohorts. For achievement and implementation analyses, this report uses YIPs.
- Reading First schools had higher percentages of ELs than the figure for All Elementary Schools (29.27%). Percentages of ELs in cohorts ranged from 33.26% to 59.62%.

- Reading First schools had higher percentages of Hispanic students and lower percentages of White students than the All Elementary Schools category.
- Cohorts 2 and 3 had higher percentages of Hispanic students than Cohort 4. As described in the Year 6 Report, African American students were significantly over-represented in Cohort 1 compared to the other cohorts and the All Elementary Schools category.
- As described in the Year 6 Report, most of the LEAs in Cohort 1 were designated as serving large or mid-sized cities, while Cohort 2 ranged from large to mid-size fringe categories. Cohort 3 included mainly suburban and rural designations. Cohort 4 had an even mix of urban and rural LEAs.
- Schools participating in Reading First for two or more years have steadily increased their percentage of teachers with full credentials. Cohort 4, which entered the program in the 2006-07 school year, entered the program with a high percentage of fully credentialed teachers.
- Using a weighted teacher qualification index based on 2008-09 CBEDS data, Cohort 3 Reading First schools had lower weighted teacher qualification indices than the other cohorts and the All Elementary Schools category.
- In 2009, all cohorts had more than 95% of their teachers fully credentialed.

# CHAPTER 2: ACHIEVEMENT

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?* The chapter also looks at the degree to which Reading First, a K-3 program, influences achievement in grades 4 and 5.

While this is the last of the evaluation reports to study program effectiveness, it is not the definitive one. That would be the Year 6 Report, which conducted a "meta-analysis" – an average across 221 separate regression studies – of the California Reading First program from its inception in 2003 to 2008. The Year 6 Report contains the central finding of the complete 7-year Reading First evaluation, that the program is effective when implemented, a finding considered to be statistically robust and operationally meaningful. The Year 6 meta-analysis has not been updated with Year 7 data due to time and budgetary constraints and to the absence of Cohort 1 schools in the sample. However, we do not find any results in the Year 7 data that would substantially contradict or weaken the conclusions of the Year 6 meta-analysis. On the contrary, though the effect sizes are weaker for the schools studied in this chapter, they are consistent with findings from previous years.

The key findings in this chapter are:

- Schools that have been in the Reading First program for six years as of 2008-09 (generally schools that received funding as part of Cohort 2) have shown steady and significant growth since they entered the program. The Reading First Achievement Index (RFAI), a composite of K-3 achievement metrics for Reading First schools that ranges from 0 to 100, has gained an average of 2.5 points per year for this cohort of schools.
- That growth has exceeded that of all non-Reading First schools on all achievement metrics over the same period, especially in moving students out of Below and Far Below Basic.
- High Implementation schools generally show more growth than Low Implementation schools.
- All YIP 6 Reading First schools generally show more growth than the Statistical Control Group, though the difference is not statistically significant in all cases.
- The 2009 achievement gap between High and Low Implementation YIP 6 schools is similar to the gap they displayed in 2008.
- English Learners benefit from Reading First, especially in grades 2 and 3. In fact, English Learners in High and Medium Implementation schools score almost succeeded in closing the achievement gap by 2009.

- The Reading First effect appears to be absent for English Learners in grades 4 and 5, but this could be a statistical artifact of the process by which high-performing English Learners are reclassified as fluent starting around grade 4 and removed from the English Learner population.

These findings arise in the context of more general findings from the Year 6 Report:

- The average Reading First (standardized beta) effect size in predicting all possible outcome variables, after controlling for starting point and demographic factors, is 0.093 with a standard error of 0.006. This is approximately 15 standard errors higher than zero, where 2 standard errors above zero would be sufficient to claim a statistically significant effect with 95% confidence. This effect does not include "Years in Program" in the definition of Reading First implementation, which would effectively double the effect size.
- The Reading First effect is meaningful, being 58% as powerful a predictor variable as the average effect of percent SED, EL, black, and migrant students per school, which are widely considered to be important and meaningful demographic variables. On average, for every achievement loss of 10 scale score points associated with demographic factors, there is a 6 scale score point gain attributable to Reading First.
- Schools in YIP 6 in 2009 (Cohort 2) have generally shown lower Reading First effects than schools in YIP 6 in 2008 (Cohort 1) and those in YIP 4 in 2008 (Cohort 3).
- Reading First effects generalize to all performance levels of the Reading First student population and to the student population as a whole. On the California Standards Test (CST) metrics, the migration of students into "Proficient & Above" is matched or exceeded by a migration of students out of "Below and Far Below Basic." These migrations are confirmed by average student CST scale score gains on the order of 30 scale score points over a 6-year period. Reading First continues to be very effective with low-performing students, in contrast with non-Reading First schools.

Achievement results for Reading First schools are presented in terms of the Standardized Testing and Reporting (STAR) Program assessments – the California Standards Test (CST) – and the Reading First End-of-Year (EOY) curriculum-based assessments. As of this report, grades 4 and 5 CST results are included to assess the sustained effects of Reading First.

The objective of this aspect of the evaluation is to determine whether or not, and to what degree, the Reading First program is effective as implemented in California. What is meant by "effective"? According to the federal guidelines for Reading First, the program is effective to the degree it ensures "that every student can read at grade level or above not later than the end of Grade 3" (U.S. Department

of Education, 2002). There are several ways to examine the effect of Reading First on reading in California given the limitations of a non-experimental design.

- Measure the size of the achievement gains of the Reading First schools for grade 3 and other grades that are related to grade 3, such as grades 2, 4, and 5
- Compare Reading First schools to comparable non-Reading First schools
- Compare Reading First schools to a "statistical control group" using statistical methods to profile how a school that is similar to Reading First schools would perform without access to the program
- 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 the level of performance immediately preceding entry into Reading First (when implementation had not yet occurred) to the present (when the program has been in place and is presumably well implemented). A significant positive gain would suggest the program 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.

The second approach, comparing Reading First schools to comparable non-Reading First schools, was discontinued in Year 4 of the evaluation for reasons discussed in the Year 4 Report, although some specifically limited comparisons were given in the March 2008 Reading First Supplemental Survey Report.[1] Given the constraints of the study, it is not possible to identify non-Reading First schools that are not to some degree employing the same program elements that are required of Reading First schools, making comparisons between them problematic.

The statistical control group approach employed in the Years 4, 5, 6, and 7 Reports uses multiple regression to calculate the achievement gains that would be expected of schools that are similar to Reading First schools but do not implement the Reading First program. This approach relies on the existence of a school implementation measure, the Reading First Implementation Index (RFII) described in detail in Chapter 3. Mathematical in nature, the statistical control group gain scores are based on a calculated relationship between implementation and achievement that is used to extrapolate the performance of "non-implementing" schools, even though they do not exist *per se*. It offers the ability to control for school-level demographic variations.

[1] See *The Reading First Supplemental Survey Report, March 2008* at www.eddata.com/resources/publications/.

The fourth approach is statistically similar to the third, but it entails comparing a sample of Reading First "low implementing" schools with a sample of Reading First "high implementing" schools.

To these four approaches, the Year 6 Report added a fifth approach that became possible as data was accumulated through the years – average all the effect sizes calculated using multiple regression to derive an overall effect size, and determine whether that effect size is significantly greater than zero, a procedure known as "meta-analysis". Although the meta-analysis in the Year 6 Report has not been recalculated for Year 7, it remains the definitive answer to the program effectiveness question. It combines data from all cohorts, all metrics, all years through 2008, and grades 2 – 5. The data available in 2009 do not include scores for Cohort 1 schools (approximately 255) or grade 3 CAT/6 scores.

Based on these five approaches, Reading First is said to show evidence of being effective to the degree that:

1. Achievement gains in Reading First schools are positive for grades 2, 3, 4, and 5.
2. Reading First schools show higher achievement gains than non-Reading First schools for grades 2, 3, 4, and 5.
3. Reading First schools show higher achievement gains than what would be predicted from a statistical control group for grades 2, 3, 4, and 5.
4. High implementing Reading First schools show higher achievement gains than low implementing Reading First schools for grades 2, 3, 4, and 5.
5. The average of the effects of Reading First implementation across all achievement metrics, as calculated using multiple regression to control for confounding demographic factors, is significantly greater than zero, with 95% confidence.

## Measures of Achievement Gains

School progress or growth, also called achievement gains, is measured using the CSTs, the CAT/6 Mean Percentile Ranks (which were discontinued by the State in 2008-09), the Reading First End-of-Year (EOY) tests, and the Reading First Achievement Index (RFAI), which is a composite of the others and is used to make decisions about continued Reading First funding for LEAs. Each metric has unique characteristics described below.

The California Standards Test (CSTs). The CSTs are administered to all California students in grades 2 and above toward the end of the school year. We use the English language arts (ELA) component of the CSTs for grades 2, 3, 4 and 5. The inclusion of grade 4 commenced with the Year 5 Report; grade 5

commenced with the Year 6 report. Students that were in kindergarten when Reading First was first implemented by the Cohort 1 LEA's moved into grade 5 in 2008.

Within ELA, we study the percentage of students per school that fall within each of the two following performance categories, which are a consolidation of the five CST performance categories (Advanced, Proficient, Basic, Below Basic, Far Below Basic). We also study the average CST scale score of the students in those grades.

1. "Proficient and Above" means the percentage of students in a school that are in the Proficient and Advanced performance categories. This is the primary metric for measuring growth that is used for accountability purposes under NCLB.

2. "Below and Far Below Basic" means the percentage of students in a school that score in the bottom two performance categories. It is just as important to measure growth out of the bottom categories, as it is to measure growth into the top categories, making it possible to assess whether Reading First is effective for low-scoring students.[2] A negative change in the percent of students testing "Below and Far Below Basic" means that students are exiting that performance level and moving to higher performance levels. Thus, a negative "gain" in this context means that growth is occurring.

3. "Mean Scale Score" refers to the average CST score of the students in the grade. A scale score is a number ranging from approximately 200 to 500, which describes a student's performance on a test in a way that facilitates valid comparisons across years. Using scale scores (which are on an "equal interval" scale and use information at all parts of the scale equally) to measure growth reduces anomalies caused by the nonlinearities present in all percentage-based scales and particularly reduces anomalies caused by where a given student distribution happens to fall relative to, say, the "Proficient" cut-point. Aside from their desirable measurement properties, mean scale scores were introduced as a regular feature of the Reading First evaluation starting with the Year 5 Report to address a claim then being made that growth was somehow limited only to those students who move into the "Proficient and Above" category from below, or out of the "Below and Far Below Basic" category. This claim was that students who do not change performance level categories show no evidence of growth and were therefore unaffected by

[2] The "Basic" category was discontinued in the Year 5 Evaluation Report because change in the percentage of students scoring in this category is not interpretable. For instance, if a large migration of students into "Proficient and Above" is exactly matched by an exodus of students out of "Below and Far Below Basic," the net change in the "Basic" category would be zero, a phenomenon that has in fact been observed in previous reports. This could lead to the erroneous conclusion that Reading First has no effect on students in the "Basic" category, when in fact it has a large effect. Change in this category can also yield a false finding of Reading First effectiveness.

Reading First. The mean scale score metric makes it clear that growth caused by Reading First is pervasive across the Reading First student population.

The CST gain scores reported in the tables of this chapter are the 2009 percentage of students in a specified category minus the corresponding percentage in the year immediately *preceding* the first year of Reading First funding. The change in scale scores is calculated using the same time frame. The gain scores are averaged across a specified population of schools to produce the tabular statistics presented in this chapter.

CAT/6 MeanPR. In the spring 2005 administration of the California STAR assessment, the CAT/6 component was discontinued in all elementary grades except for grade 3. The CAT/6 was discontinued for all grades as of 2009. A description of the CAT/6 metric is given in previous reports. The CAT/6 played an important role in assessing Reading First in grade 3, not only by breaking out Reading, Language Arts, and Spelling, but also by serving as a cross-check on the grade 3 CSTs, which were at times anomalous relative to other grades. It also played an important role in ranking California students relative to their peers in other states, and as a component of the Reading First Achievement Index (RFAI).

End-of-Year (EOY) Test. As the name denotes, 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 a timed oral reading in which fluency 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 kindergarten and grade 1. It is also the only test used in this evaluation that is administered in Spanish to students in "waivered" Reading First classrooms (that is, classrooms in which instruction is conducted in Spanish using State Board of Education-adopted Spanish translations of the adopted reading programs by permission of a waiver). 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 (2005).

Reading First Achievement Index (RFAI). The RFAI has been a weighted combination of school-level percentages of students meeting various performance levels and benchmarks drawn from the CSTs, the CAT/6, and the EOY, with the heaviest weights placed on the CSTs. In 2009, due to its discontinuation, the CAT/6 was dropped from the RFAI. A simple equating procedure was used to make the 2009 RFAI as comparable as possible to the pre-2009 RFAI, described in Appendix E. The RFAI was first computed in 2004. Like the CST, each school RFAI can be interpreted as a percentage of students meeting a set of combined benchmarks and performance levels, although it is not based on a single benchmark or

performance level. The RFAI gain score for each school is its 2009 RFAI minus its RFAI at the end of its *first* year of Reading First implementation (unlike the CSTs, which refer back to the year immediately preceding the first year of Reading First implementation).

Grouping of Schools by "Years in Program" (YIP)

Starting with the Year 4 report, schools have been grouped by Years in Program (YIP) rather than LEA funding cohort for purposes of doing growth analysis. As explained in prior reports, there are cases where LEAs that received funding starting in one year added schools to Reading First in a later year. For purpose of measuring program effects, it is the year in which implementation actually began that is most important.

It is often found in educational research that intervention program effects vary over time and across cohorts. There are also changes in the behavior of tests over the years, which would influence the YIPs differentially. In the case of Reading First, both the YIPs and the achievement metrics have different characteristics depending on starting year. Cohort 1 (which would have been YIP 7 in 2009) is notably more urban than Cohort 2 (YIP 6 in 2009) and has had different rates of implementation. The grade 3 achievement metric experienced a statewide dip in 2004 which yields qualitatively different trend-lines for YIPs that started before the dip compared to those that started after.

In 2009, we focus on those Reading First schools that have been in the program for six years and five years, omitting schools from more recent YIPs (71 combined).

Comparison of Reading First to Non-Reading First Schools

Prior to the Year 4 Report, efforts were made to identify a sample of non-Reading First schools that would be comparable to the Reading First population and yet not contain Reading First-style program elements. These efforts were abandoned in Year 4 as it became increasingly clear that there was no way to control for the increasing similarity between the two groups of schools as regards their use of state-adopted reading programs, common professional development resources, and use of reading coaches. In place of a sample of comparable non-Reading First schools, we instituted the concept of the "statistical control group," described below. Nonetheless, we continue to report the gains of the non-Reading First elementary school population in California in order to provide an overview of the rest of the state and show how it has been trending since 2002. This provides an essential context for studying the Reading First gains, for we see that the Reading First upward trend is mirrored in the rest of the state. However, it

is emphasized that the non-Reading First group is demographically dissimilar to the Reading First group and that caution should be exercised when comparing them.[3]

Comparison of High, Medium, and Low Implementation Reading First Schools

One defining characteristic 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 and Appendices A, B, and C describe the teacher, coach, and principal surveys that were administered in all Reading First schools and used to compute a Reading First Implementation Index (RFII) statistic for each school with sufficient respondents. 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 – 2009 based on a survey administration in the spring of each year.

The RFII was used to divide Reading First schools into two groups labeled High Implementation Schools and Low Implementation Schools. For the Year 4 Report and those preceding, a high implementation school was defined as a school whose average RFII since entering the program is greater than or equal to 36.0, the average RFII in 2004. A low implementation school had an average yearly RFII less than 36.0. Based on advice from the Evaluation Advisory Group (EAG), the definitions were changed for the Year 5 Report and Year 6 Report. A high implementation school is one whose average yearly RFII is greater than 1 standard deviation above the original 36.0 cut-point, approximately 41.4. A low implementation school continues to be one whose average yearly RFII is less than 36.0.[4]

This change had the effect of introducing a more stringent definition of high implementation, but also of leaving out the schools between 36.0 and 41.4 from the high and low groups. Although these "Medium" schools were accounted for in the "All Reading First Schools" category, it was decided by the EAG in 2009 that statistics for the "Medium" schools ($36.0 < RFII < 41.4$) should be explicitly reported.

Another change, instituted in the Year 5 Report on the advice of the EAG, was to use a rolling 2-year average RFII instead of averaging all of a school's RFIIs. Therefore, each school's "final" 2009 RFII statistic is an average of its preliminary RFII (computed from the 2009 surveys) and its preliminary 2008 RFII (computed from the 2008 surveys) on the theory that a rolling 2-year average is more stable and reliable than the RFII computed from a single year's data, yet responsive to changes in school implementation practices.

[3] In the trend-line charts presented later in this chapter, the All Non-Reading First 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.

[4] An EAG recommendation to define "low implementing" schools as those with an RFII more than one standard deviation below the mean was not implemented because it was found that this yielded a very small number of low implementing schools, not sufficient for statistical comparisons.

Calculating Achievement for the Statistical Control Group

As discussed in prior reports, the statistical control group is defined using regression models to calculate the 2009 achievement score that a school similar to the Reading First schools (i.e., that has the same demographic and starting characteristics as the Reading First YIP under consideration) *would* obtain if it were *not* implementing the program. The school-level variables used to predict 2009 scores on each achievement metric are:

- School Starting Point. The average score of students at a school on a given achievement metric in 2003.
- YIP. The number of years the school has been in Reading First.
- RFII Average. The average of all the (preliminary) RFIIs calculated for the school since 2005.
- Percent SED. The percentage of socio-economically disadvantaged students in the school, based on participation in the school lunch program.
- Percent EL. The percentage of English Learners in the school, based on performance on the California English Language Development Test (CELDT).
- Percent Migrant. The percentage of students at a school classified as "migrant" in the STAR file.
- Percent Black. The percentage of students at a school classified as "black" in the STAR file.
- Number of Students. The total number of students at the school.
- Student/Teacher Ratio. The average number of students in a classroom.

These variables, selected as significant predictors of 2009 school scores, are used to construct regression equations. Each variable is given a regression coefficient. To calculate Statistical Control Group statistics, certain representative statistics are plugged into each regression equation and multiplied by the coefficients to yield a 2009 prediction. These "representative" statistics are generally the average of all the schools in the YIP (YIP 6 in this chapter, YIP 5 in Appendix D) for a given variable – the average starting point, the average percent SED, the average percent EL, the average percent migrant, the average percent black, the average number of students, and the average student/teacher ratio.

For reasons discussed in Chapter 3 of the Year 4 Report, we plug an RFII of 25 into the regression equation to signify a school that is not implementing the program. Thus, 25 is entered into the regression equation to calculate an expected 2009 achievement score and gain score for the statistical control group. As stated previously, the statistical control group is not a literal group of schools but an extrapolation based on a relationship between achievement and implementation derived statistically from the Reading

First schools. (Non-Reading First schools could not be used to compute this relationship since they do not take the surveys and do not receive an RFII.) For additional background reference regarding the detailed procedure for computing the statistical control group achievement statistics, the reader is referred to Chapter 4 of the Year 4 Report.[5]

## Achievement Results

The following pages present tables and trend-line charts showing starting scores, ending scores (2009), and gains on each of 13 achievement metrics for YIP 6.

Table 2.1 reports RFAI total gains for YIPs 5 and 6.

Tables 2.2, 2.3, 2.4, and 2.5, with accompanying trend-line charts, show total gains on the CST for YIP 6 for grades 2, 3, 4 and 5. Similar tables and charts are provided for YIP 5 in Appendix D.

Here are some notes on interpreting the data in the tables:

- Significance Tests. The statistics in the achievement tables provided in this chapter may be accompanied by superscripts "a", "b", and "c." These refer to tests for statistical significance. Significance tests answer the question, "How likely is it that the observed difference would have occurred by chance?" As noted below each table, the superscript "a" means that the group in question (the one with the superscript) has a gain score that is "significantly" higher than that of the Statistical Control Group at the 95% confidence level, which means that the probability of the difference occurring by chance is less than 0.05 (i.e., $p < 0.05$). The "b" means the group is significantly higher than the "All Non-Reading First Elementary Schools" group. The "c" means the new group average is significantly higher than its starting point, i.e., that the change is significantly larger than zero. Three pieces of information go into a significance test: the difference *between* groups, the amount of variation *within* each group, and the *number* of schools within each group. A large difference between groups, with little variation within each group, and a large number of schools within each group, will be more likely to yield a "statistically significant" difference.
- Rounding Errors. Sometimes we report a gain score that does not appear to exactly equal the difference between the starting score and the ending score for a given metric. The explanation is that the reported starting and ending scores have been rounded to one decimal place, whereas the reported difference or gain is computed at more than 8 decimal places. Thus the reported gain is (slightly) more accurate than the difference between the reported starting and ending scores.

[5] The California Reading First Year 4 Evaluation Report is available online at www.eddata.com/resources/publications/ .

- Downward Adjustment of the Trend-Lines. In the trend-line charts, the "non-Reading First Schools" trend-line has been adjusted downward to have the same starting point as "All Reading First Schools" to make it easier to compare their trend-lines. Ordinarily, the non-Reading First schools have much higher starting points than Reading First schools.

RFAI Gains (Table 2.1)

Table 2.1 reports starting points, ending points, and total RFAI gains for YIP 5 and 6 schools, starting from the first year of Reading First implementation. Because the RFAI is only administered to Reading First schools, there are no comparable statistics for non-Reading First schools.

**Table 2.1: RFAI Gains, YIPs 5 and 6**

| **Years in Program: 5, 6 Grades: K-3** | **Reading First Schools** | | | | |
|---|---|---|---|---|---|
| | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | Statistical Control Group (RFII = 25) |
| Years in Program: 6 | | | | | |
| Number of Schools | 294 | 74 | 156 | 64 | N/A |
| 2004 | 35.1 | 35.5 | 34.3 | 36.4 | 35.1 |
| 2009 | 49.9 | 52.5 | 49.6 | 47.8 | 47.1 |
| RFAI Gain | **14.9**[ac] | **17.0**[ac] | **15.3**[ac] | **11.4**[c] | **12.1** |
| Years in Program: 5 | | | | | |
| Number of Schools | 128 | 28 | 56 | 44 | N/A |
| 2005 | 34.7 | 37.1 | 35.8 | 31.9 | 34.7 |
| 2009 | 48.3 | 49.2 | 49.3 | 46.4 | 45.0 |
| RFAI Gain | **13.5**[ac] | **12.0**[c] | **13.5**[ac] | **14.5**[ac] | **10.3** |

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

The RFAI gains for YIPs 5 and 6 support the hypothesis that Reading First schools are growing and that they grow more quickly than the statistical control group. The gains for YIP 6 also show that high implementing schools grow faster than low implementing schools, a pattern not reproduced for the YIP 5 schools (as also was found in the Year 5 and Year 6 Reports). For YIP 5, we see that the low implementing schools actually grew faster than high implementing schools (RFAI gain = 14.5 versus 12.0).

The anomaly does not appear when we look at the YIP 5 Statistical Control Group gain (RFAI gain = 10.3, which is less than 14.5). Unlike the sample of Low Implementation schools, the Statistical Control Group controls for demographic variations among schools. It appears that the Low Implementation schools possess a demographic advantage over the High Implementing schools and so are not statistically

comparable to them. The Statistical Control Group removes this advantage and shows that when the schools are similar, implementation of Reading First is expected to produce higher growth rates.

CST Results for Grade 2 (Table 2.2 and Figures 2.2a – 2.2c)

Table 2.2 reports the starting and ending grade 2 CST scores of students in schools that have been in the program six years.

**Table 2.2: CST Metric, YIP = 6, Grade = 2**

| **Years in Program: 6**<br>**Grade: 2** | **Reading First Schools** | | | | | All Non-Reading First Elementary Schools |
|---|---|---|---|---|---|---|
| | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | Statistical Control Group (RFII = 25) | |
| Number of Schools | 289 | 73 | 153 | 63 | N/A | 4025 |
| % Proficient and Above | | | | | | |
| 2003 | 20.7 | 20.9 | 20.7 | 20.8 | 20.8 | 43.0 |
| 2009 | 39.9 | 41.0 | 40.4 | 37.5 | 39.0 | 56.3 |
| Change Since Starting Year | **19.2**[bc] | **20.1**[bc] | **19.7**[bc] | **16.8**[bc] | **18.2** | **13.3** |
| % Below or Far Below Basic | | | | | | |
| 2003 | 46.0 | 45.0 | 46.8 | 45.2 | 46.0 | 25.5 |
| 2009 | 28.7 | 27.1 | 28.9 | 30.2 | 30.1 | 18.5 |
| Change Since Starting Year | **-17.3**[bc] | **-17.8**[bc] | **-17.9**[bc] | **-15.1**[bc] | **-15.9** | **-7.0** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 310.3 | 311.3 | 309.5 | 311.1 | 310.3 | 341.3 |
| 2009 | 333.5 | 335.1 | 334.1 | 330.1 | 330.8 | 357.6 |
| Change Since Starting Year | **23.2**[abc] | **23.8**[bc] | **24.7**[abc] | **19.1**[c] | **20.5** | **16.2** |

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."
[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

The "% Proficient and Above" has risen from 20.7 percentage points in 2003 to 39.9 percentage points in 2009, a gain of 19 points over six years. This growth rate corresponds to a gain of 23.2 scale score points on the grade 2 CST over six years, 3.9 points per year. While not as large as the 5 scale score points per year growth rate enjoyed by Cohort 1 (see Year 6 Report, YIP 6, Grade 2), it is substantial. Similarly high growth rates occurred in High Implementation and Medium Implementation schools. Lower growth rates occurred in Low Implementation schools (though the High/Low differences are not as dramatic as in previous Reports).

As in previous reports, the Reading First schools strongly outgrow the non-Reading First schools, particularly evident in the movement of students out of the Below Basic and Far Below Basic

performance levels. However, the growth rates are not as strong as those reported in the Year 6 Report for the previous cohort of YIP 6 schools (Year 6 Report: gain of 30.0 scale score points; Year 7 Report: gain of 23.2).

Remember that these are *school-level* gains, with new students entering kindergarten each year. Since each student cohort can be assumed to start at roughly the same average level of ability in kindergarten, one can interpret this rate of growth to mean that Reading First schools are now moving each new cohort of students 23 scale score points further up the scale from kindergarten to grade 2 than they were, with similar cohorts, six years ago.

Due to common elements in statewide reading instructional implementation, we see that the rest of the state's elementary schools have also shown significant growth, but their growth lags that of Reading First schools by 7 scale score points (by 10 percentage points if we are talking about moving low performing students out of the bottom categories). We see that lower performing students are moving out of the bottom performance levels at a rate similar to that of the mid-range students moving into the top two performance levels, a pattern not seen in non-Reading First schools. This remains a key and important difference between Reading First and non-Reading First schools, a difference that has held up across all of our evaluation reports.

In comparing the Statistical Control Group and Low Implementation schools, we find that the Statistical Control Group (defined to have minimal implementation) performs slightly better than the Low Implementing schools. As discussed with Table 2.1, this is because the Control Group controls for demographic variables that are not controlled in the Low Implementation sample. In this case, it corrects for a demographic *disadvantage* of the Low Implementation sample. In deciding whether Reading First has been effective with this group of schools from 2003 to 2009, comparisons should be made with the Statistical Control Group.

Figures 2.2a – 2.2c show trend-lines for grade 2.

**Figure 2.2a: CST % Proficient & Above, YIP = 6, Grade = 2**

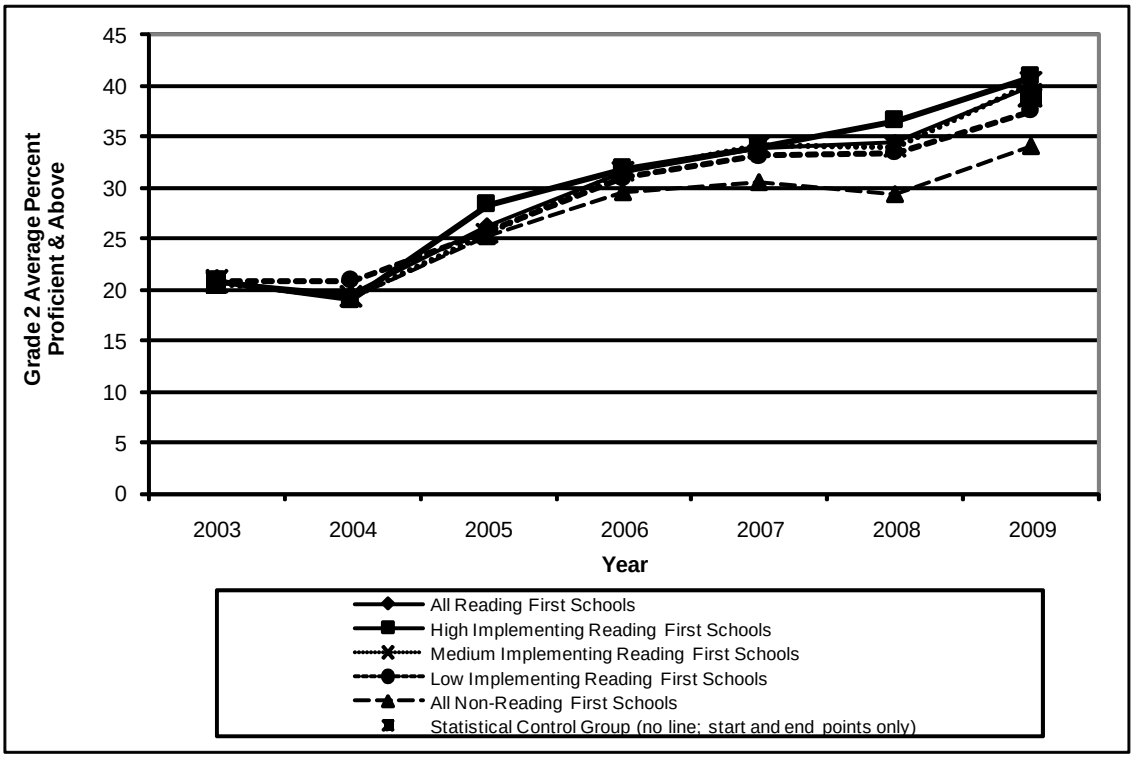



**Figure 2.2b: CST % Below and Far Below Basic, YIP = 6, Grade = 2**

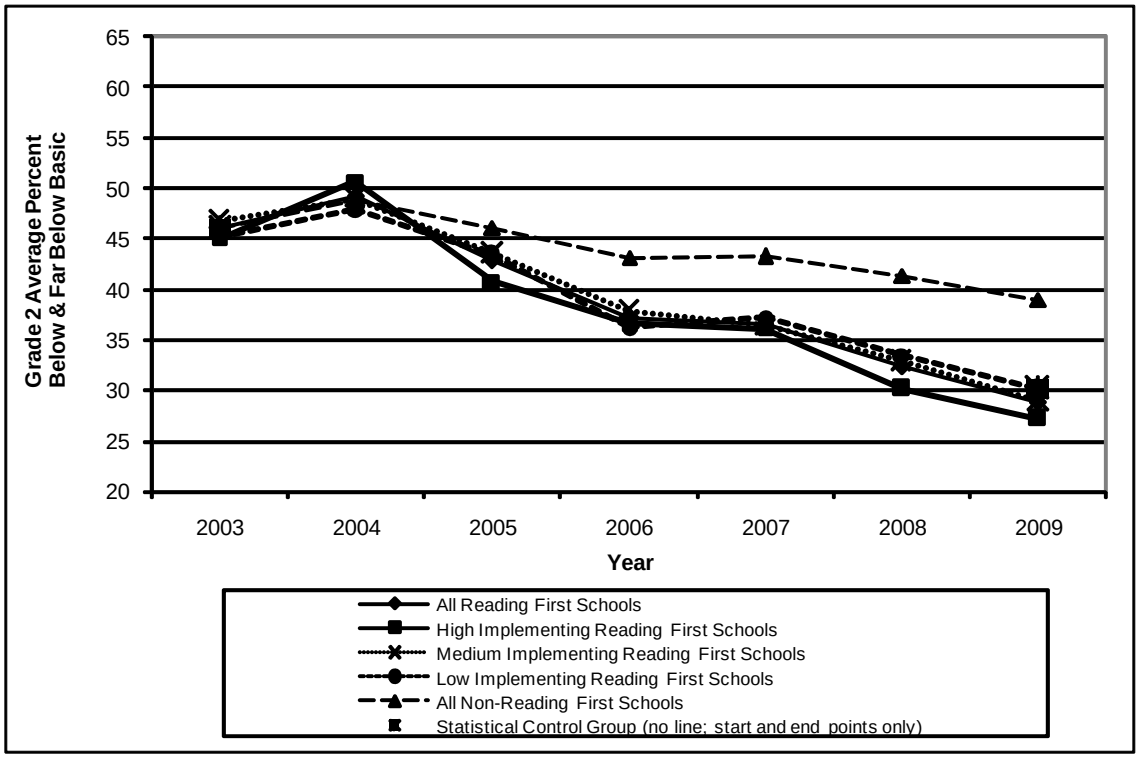
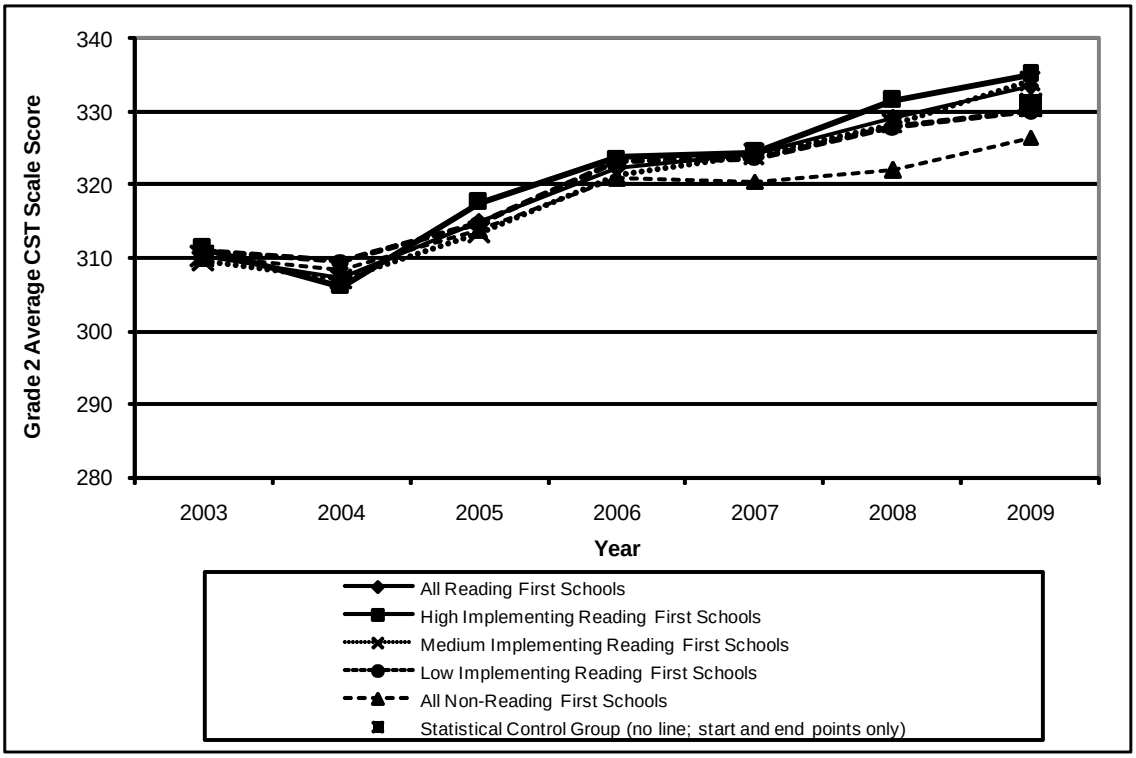


**Figure 2.2c: CST Mean Scale Score, YIP = 6, Grade = 2**




In addition to the patterns discussed above, we see that growth on the grade 2 scale score metric has, with the exception of 2004, been fairly steady. We see that the high and low implementation schools started at approximately the same location on the scale and fanned out according to their level of implementation. Such "fan" patterns are indicative of program efficacy. However, compared to trend-lines for Cohort 1 schools (YIP 6 in 2008), we see that the fan shape for Cohort 2 schools (YIP 6 in 2009) is much less pronounced (see the Year 6 Report). The gap between high implementing and low implementing schools was *reversed* in 2004 (indicating low program efficacy in the first year of implementation – a common finding which we have noted in previous reports) and only began to fan out positively in 2005 and 2006. In 2007 the implementation effect disappeared entirely (high and low implementing schools show almost identical scores), but in 2008 and 2009 the fan reopens. This pattern for Cohort 2 schools is analyzed in detail on pages 60-67 of the Year 6 Report. Unlike the other cohorts, the Reading First effect for Cohort 2 schools declined until 2008, at which point it seems to have reasserted itself and continues to the present. The reasons for this pattern, and why it is peculiar to Cohort 2 schools, are not known.

CST Results for Grade 3 (Table 2.3 and Figures 2.3a – 2.3c)

Table 2.3 reports gains, starting scores, and ending scores for grade 3 in YIP 6 schools.

**Table 2.3: CST Metric, YIP = 6, Grade = 3**

| **Years in Program: 6**<br>**Grade: 3** | **Reading First Schools** | | | | | All Non-Reading First Elementary Schools |
|---|---|---|---|---|---|---|
| | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | Statistical Control Group (RFII = 25) | |
| Number of Schools | 291 | 75 | 153 | 63 | N/A | 4016 |
| % Proficient and Above | | | | | | |
| 2003 | 16.8 | 17.1 | 16.4 | 17.3 | 16.8 | 39.9 |
| 2009 | 28.7 | 30.1 | 28.7 | 27.0 | 27.1 | 47.7 |
| Change Since Starting Year | **11.9[bc]** | **12.9[bc]** | **12.3[abc]** | **9.7[c]** | **10.3** | **7.8** |
| % Below or Far Below Basic | | | | | | |
| 2003 | 53.9 | 53.7 | 54.6 | 52.4 | 53.9 | 29.8 |
| 2009 | 40.4 | 38.4 | 40.5 | 42.5 | 43.3 | 24.8 |
| Change Since Starting Year | **-13.5[abc]** | **-15.3[abc]** | **-14.1[abc]** | **-9.9[bc]** | **-10.6** | **-5.0** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 298.3 | 299.0 | 297.0 | 300.5 | 298.3 | 334.5 |
| 2009 | 318.3 | 320.7 | 318.4 | 315.1 | 315.1 | 346.1 |
| Change Since Starting Year | **20.0[abc]** | **21.7[abc]** | **21.4[abc]** | **14.6[c]** | **16.8** | **11.6** |

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."
[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

While absolute gains in grade 3 "% Proficient and Above" are not as large as those for grade 2, they are large relative to non-Reading First schools. Movement out of the bottom categories is particularly large. Unlike grade 2, the growth rate of the YIP 6 schools here is actually larger than the previous cohort of YIP 6 schools in the Year 6 Report (Year 6 Report: gain of 15.2 scale score points; Year 7 Report: gain of 20.0 scale score points).

As with grade 2, the Statistical Control Group outperforms the Low Implementation schools due to a demographic disadvantage of the Low Implementation schools.

Figures 2.3a – 2.3c show trend-lines for grade 3.

**Figure 2.3a: CST % Proficient & Above, YIP = 6, Grade = 3**

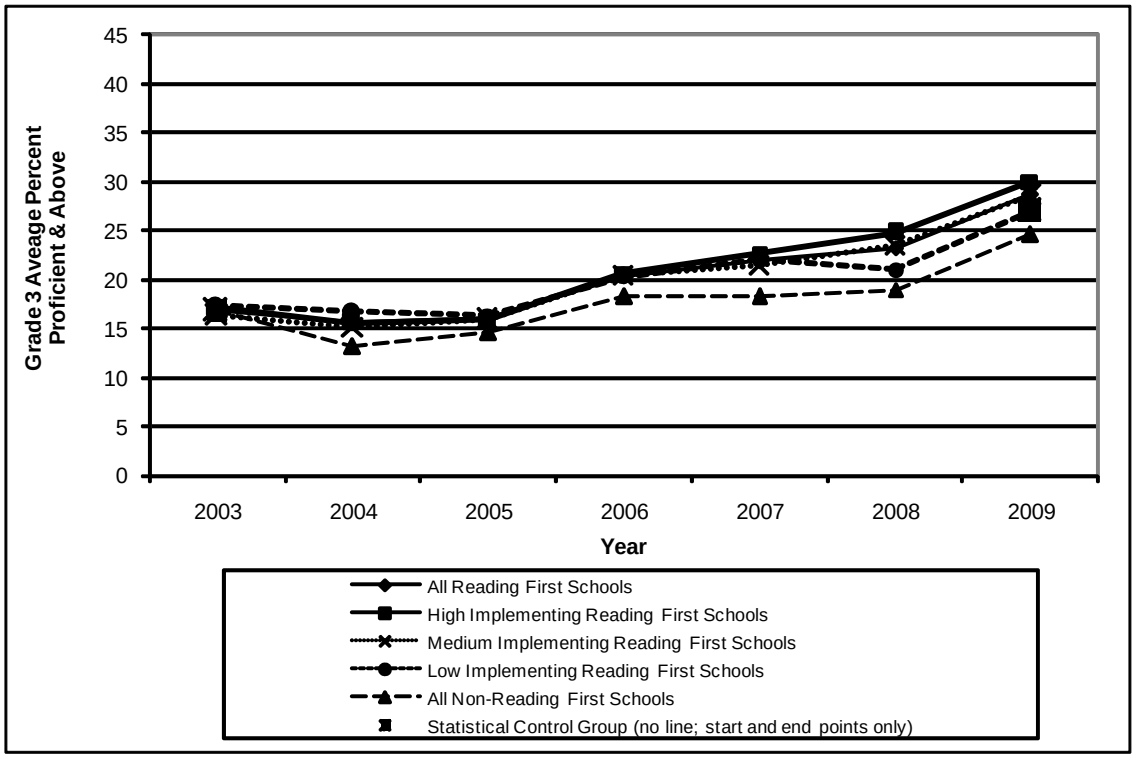



**Figure 2.3b: CST % Below and Far Below Basic, YIP = 6, Grade = 3**

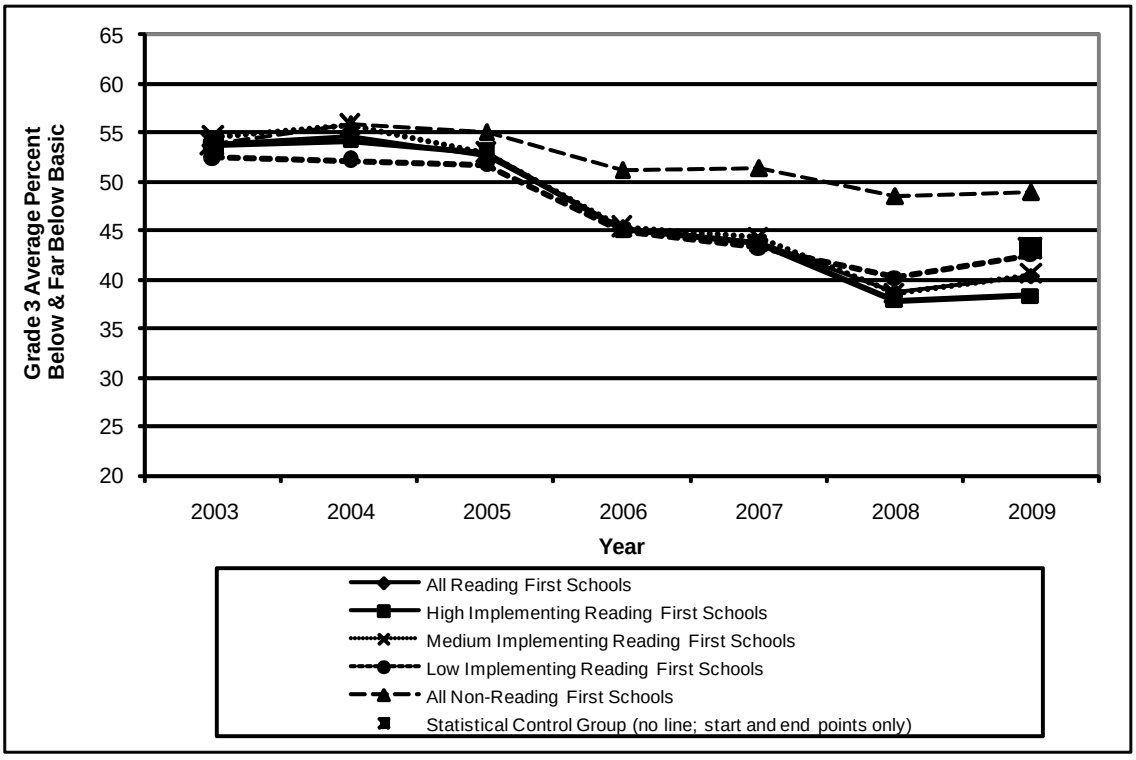
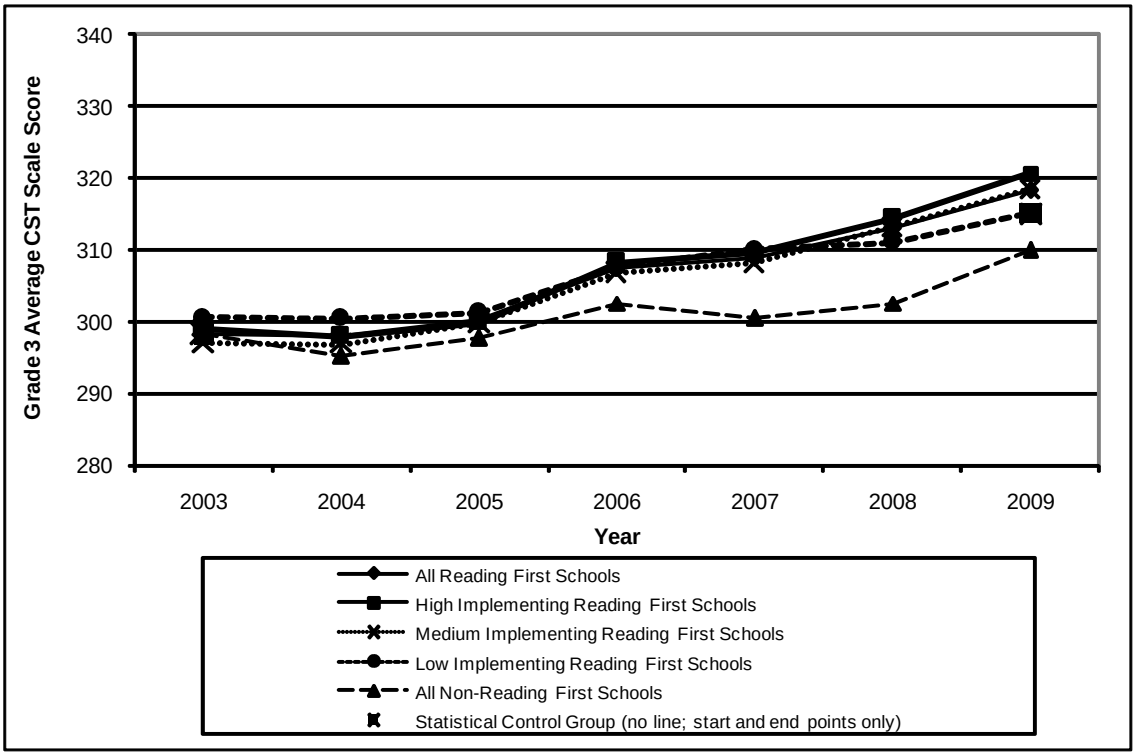


**Figure 2.3c: CST Mean Scale Score Per Student, YIP = 6, Grade = 3**




Figures 2.3a – 2.3c reveal a number of important patterns that are not readily apparent in the statistics of Table 2.3. The most obvious, noted in previous reports, is that the grade 3 CST scores dipped in 2004, creating a "U" shape for both Reading First and non-Reading First schools. We see that even though the CST trends for Reading First schools are somewhat flat relative to grade 2, they are substantially more positive than those for the non-Reading First schools. After 2004, the trends are steadily positive.

The "fan" pattern has similarities with grade 2. The fan is reversed in 2004, the first year of implementation, narrows in 2005 and 2006, then starts widening in the "correct" orientation (high implementing on top) in 2007 and 2008. In 2009, the fan is at its widest, suggesting that Cohort 2 schools are just now warming to the program and becoming effective at implementing Reading First principles. In this regard, Cohort 2 has been much slower to respond and engage with the program than Cohort 1 schools, where the fan widened dramatically in the second and third years of implementation. We expect that if the program were to be continued in some form beyond 2009, the high implementing Cohort 2 schools would show continued improvements beyond what they have shown thus far.

CST Results for Grade 4 (Table 2.4 and Figures 2.4a – 2.4c)

Table 2.4 reports the CST results for grade 4 which have been collected for YIP 6 schools.

**Table 2.4: CSTs, YIP = 6, Grade = 4**

| **Years in Program: 6**<br>**Grade: 4** | **Reading First Schools** | | | | | All Non-Reading First Elementary Schools |
|---|---|---|---|---|---|---|
| | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | Statistical Control Group (RFII = 25) | |
| Number of Schools | 279 | 73 | 149 | 57 | N/A | 3990 |
| % Proficient and Above | | | | | | |
| 2003 | 21.4 | 22.4 | 20.5 | 22.4 | 21.4 | 46.0 |
| 2009 | 46.6 | 47.1 | 46.5 | 46.1 | 45.5 | 65.0 |
| Change Since Starting Year | **25.2**[bc] | **24.7**[bc] | **26.0**[bc] | **23.7**[bc] | **24.1** | **19.0** |
| % Below or Far Below Basic | | | | | | |
| 2003 | 39.3 | 38.6 | 40.3 | 37.6 | 39.3 | 20.2 |
| 2009 | 20.9 | 20.3 | 21.3 | 20.7 | 21.7 | 12.4 |
| Change Since Starting Year | **-18.4**[bc] | **-18.3**[bc] | **-19.0**[bc] | **-16.9**[bc] | **-17.6** | **-7.9** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 317.3 | 318.4 | 315.9 | 319.6 | 317.3 | 347.4 |
| 2009 | 345.5 | 346.4 | 345.3 | 345.1 | 343.8 | 372.2 |
| Change Since Starting Year | **28.2**[bc] | **28.1**[c] | **29.4**[abc] | **25.5**[c] | **26.5** | **24.9** |

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."
[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

The grade 4 growth patterns are similar to those for grades 2 and 3, and Reading First schools compare favorably with the non-Reading First schools. The average scale score growth is 28.2 points over six years, even larger than the 26.9 point growth reported for the YIP 6 schools in the Year 6 Report. However, the differences between High and Low Implementation schools, and between All Reading First schools and the Statistical Control Group, are small and not statistically significant.

What makes this table highly significant is that only grades K-3 classrooms are funded by Reading First. There is no grade 4 Reading First program: yet as we saw last year, and continue to see this year, the CST scores are almost what one would expect if Reading First extended to grade 4. This supports the hypothesis that Reading First students have been able to carry with them the skills, reading habits and conceptual understanding that they developed in the earlier grades, and that rigorous instruction in the lower grades lays the groundwork for large gains in the higher grades.

Non-Reading First schools also show substantial gains over this period, but the gains are smaller and less uniform across the population as can be seen in Figures 2.4a – 2.4c.

**Figure 2.4a: CST % Proficient & Above, YIP = 6, Grade = 4**

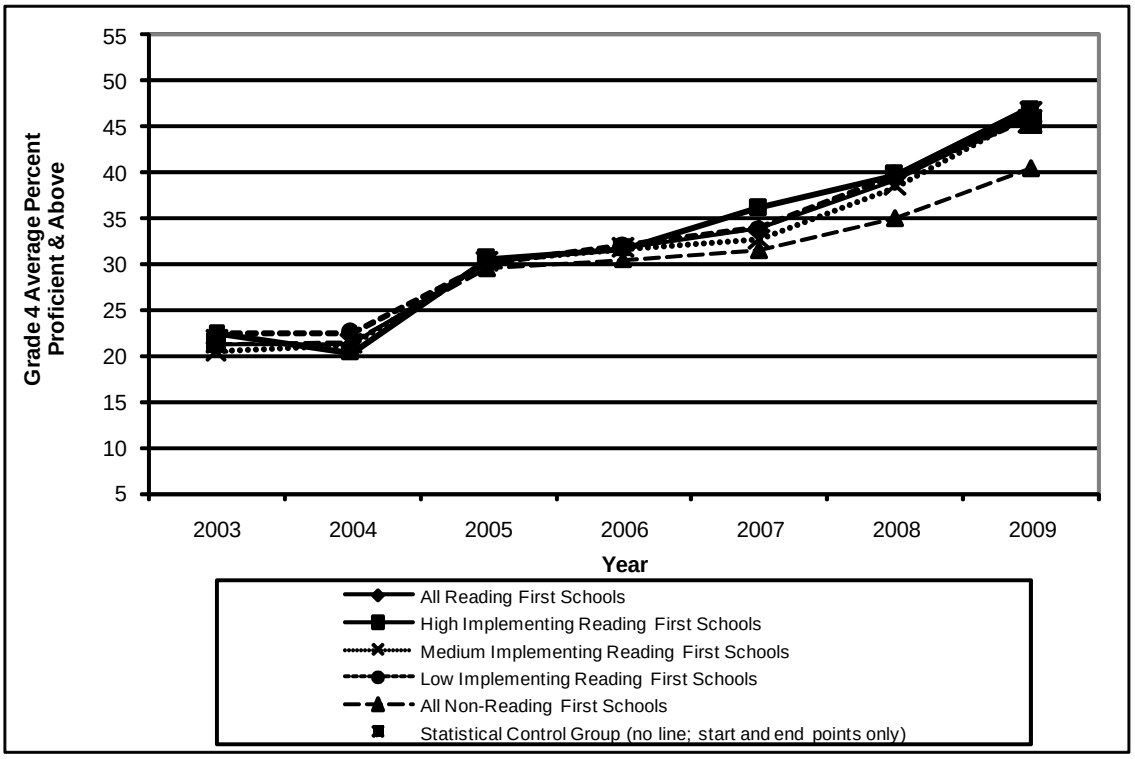



**Figure 2.4b: CST % Below and Far Below Basic, YIP = 6, Grade = 4**

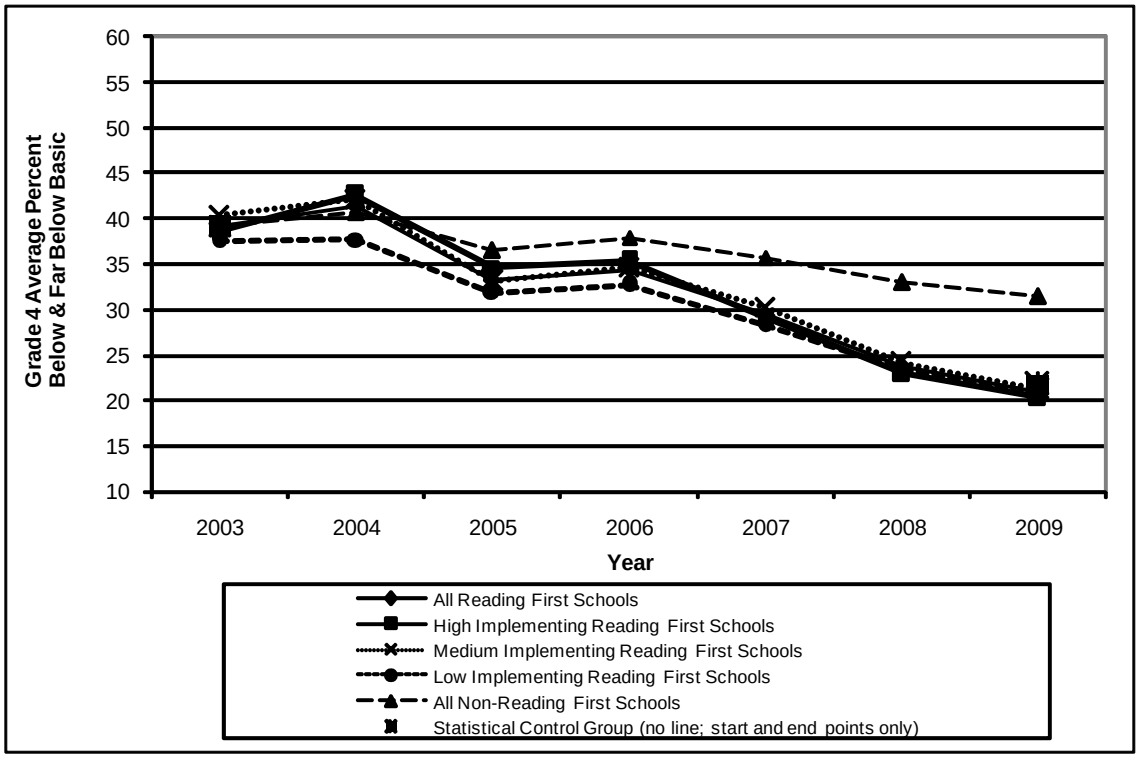
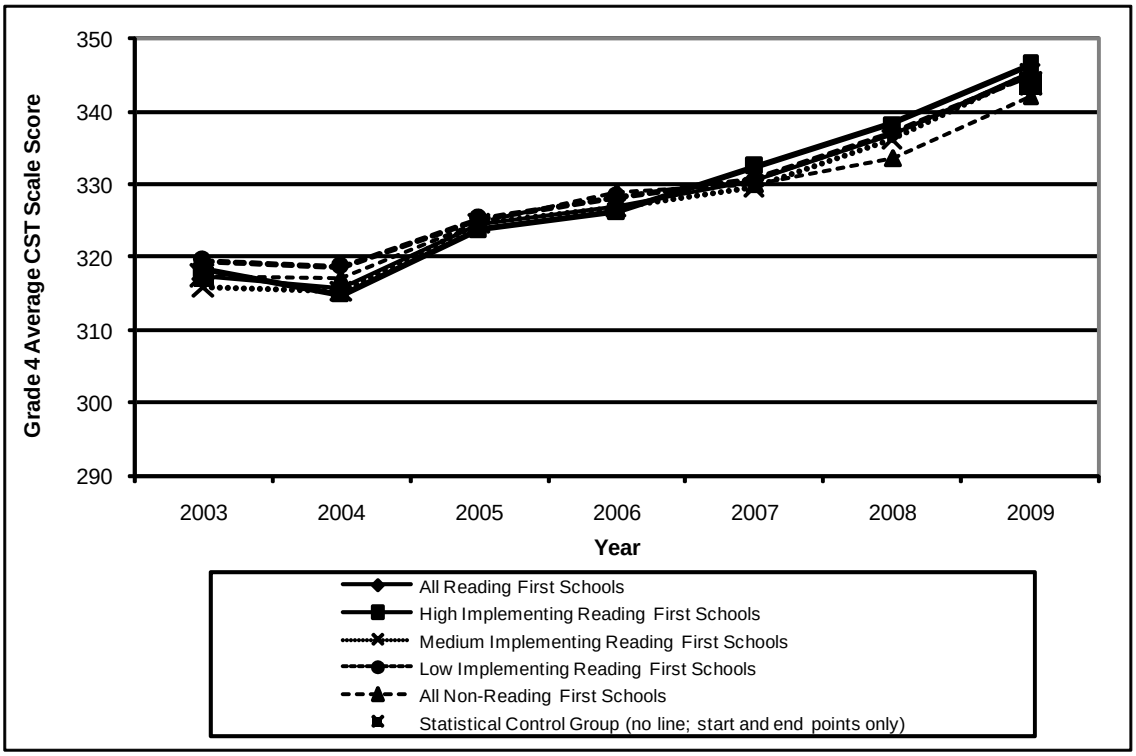


**Figure 2.4c: CST Mean Scale Score, YIP = 6, Grade = 4**




Figures 2.4a – 2.4c show strong positive trends that compare favorably with those for non-Reading First schools. However, the gap between high and low implementing schools is generally small and for some years reversed, especially in Figure 2.4b. This is to be expected in part because students trained in Reading First classrooms since kindergarten did not really begin entering grade 4 until 2008. They only began entering grade 5 in 2009. And it is precisely in the span from 2007 – 2009 that we see the Reading First effect begin to emerge, especially in 2.4c, and high implementing schools perform better than low implementing schools.

This supports the hypothesis the students raised in Reading First in grades K-3 do better in the higher grades.

CST Results for Grade 5 (Table 2.5 and Figures 2.5a – 2.5c)

Table 2.5 reports the CST results for grade 5 which have been collected for YIP 6 schools. 2008 was the first year that students who had been in Reading First since kindergarten moved into grade 5. The present cohort of students also had the experience of being in Reading First classrooms since kindergarten.

**Table 2.5: CSTs, YIP = 6, Grade = 5**

| **Years in Program: 6**<br>**Grade: 5** | **Reading First Schools** | | | | | All Non-Reading First Elementary Schools |
|---|---|---|---|---|---|---|
| | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | Statistical Control Group (RFII = 25) | |
| Number of Schools | 275 | 72 | 146 | 57 | N/A | 3963 |
| % Proficient and Above | | | | | | |
| 2003 | 17.9 | 17.9 | 17.5 | 19.1 | 17.9 | 41.8 |
| 2009 | 39.3 | 40.2 | 38.8 | 39.5 | 38.6 | 57.6 |
| Change Since Starting Year | **21.3**[bc] | **22.2**[bc] | **21.3**[bc] | **20.4**[bc] | **20.7** | **15.8** |
| % Below or Far Below Basic | | | | | | |
| 2003 | 43.9 | 43.6 | 44.5 | 42.6 | 43.9 | 23.4 |
| 2009 | 24.6 | 22.8 | 24.9 | 25.9 | 26.1 | 15.2 |
| Change Since Starting Year | **-19.3**[bc] | **-20.9**[abc] | **-19.6**[bc] | **-16.7**[bc] | **-17.8** | **-8.2** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 311.2 | 311.5 | 310.5 | 312.8 | 311.2 | 339.1 |
| 2009 | 336.6 | 338.3 | 335.9 | 336.0 | 335.9 | 361.2 |
| Change Since Starting Year | **25.4**[bc] | **26.9**[bc] | **25.5**[bc] | **23.2**[c] | **24.7** | **22.1** |

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."
[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

The patterns for grade 5 are similar to those for grade 4. Growth has been substantial. Movement into the top performance levels is similar to movement out of the bottom performance levels, a pattern not seen in the non-Reading First schools. But while the Reading First schools out-perform the non-Reading First schools, the contrasts with the statistical control group are not significant except for the "percent Below and Far Below Basic" achievement metric. The implementation contrasts were much stronger in the Year 6 Report, due most likely to special characteristics of the cohort of students that entered kindergarten in 2003, a cohort that seems to have been unusually responsive to Reading First. The succeeding cohorts have been less responsive.

Again, it is worth pointing out that there is no special program being implemented in grade 5 that would differentiate it from the rest of California schools. Yet there does appear to be a Reading First effect that becomes evident as Reading First-trained students enter the higher grades.

**Figure 2.5a: CST % Proficient & Above, YIP = 6, Grade = 5**

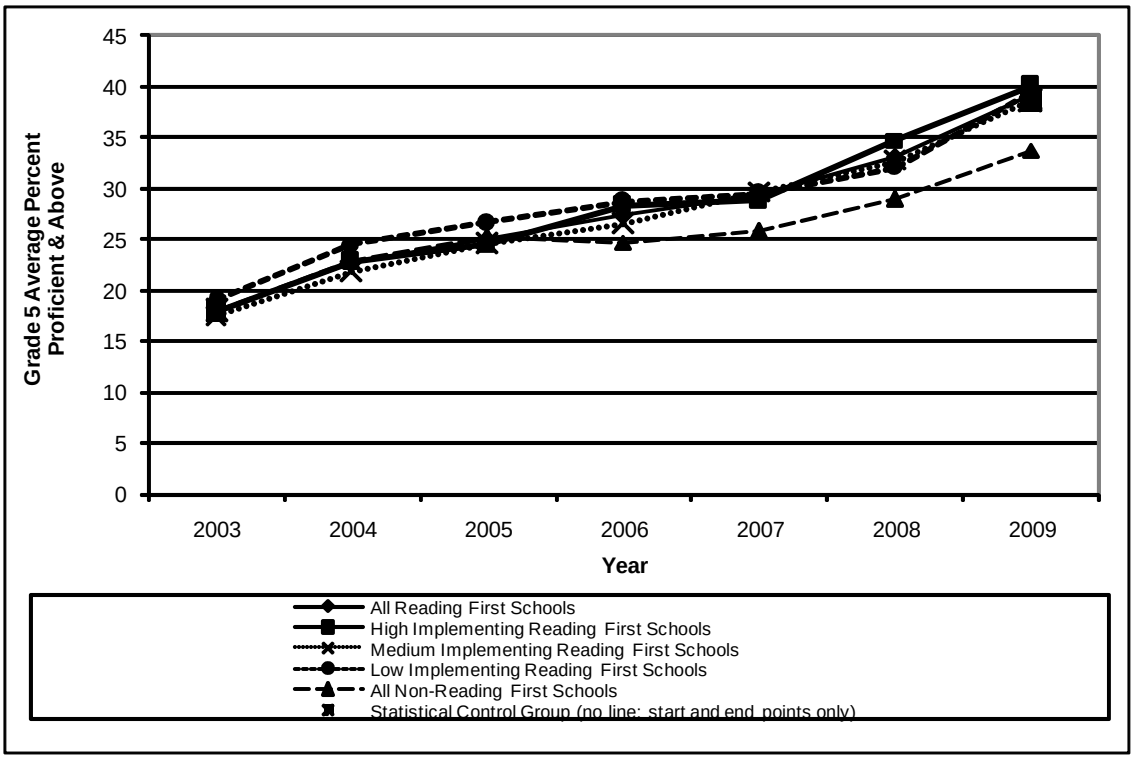



**Figure 2.5b: CST % Below and Far Below Basic, YIP = 6, Grade = 5**

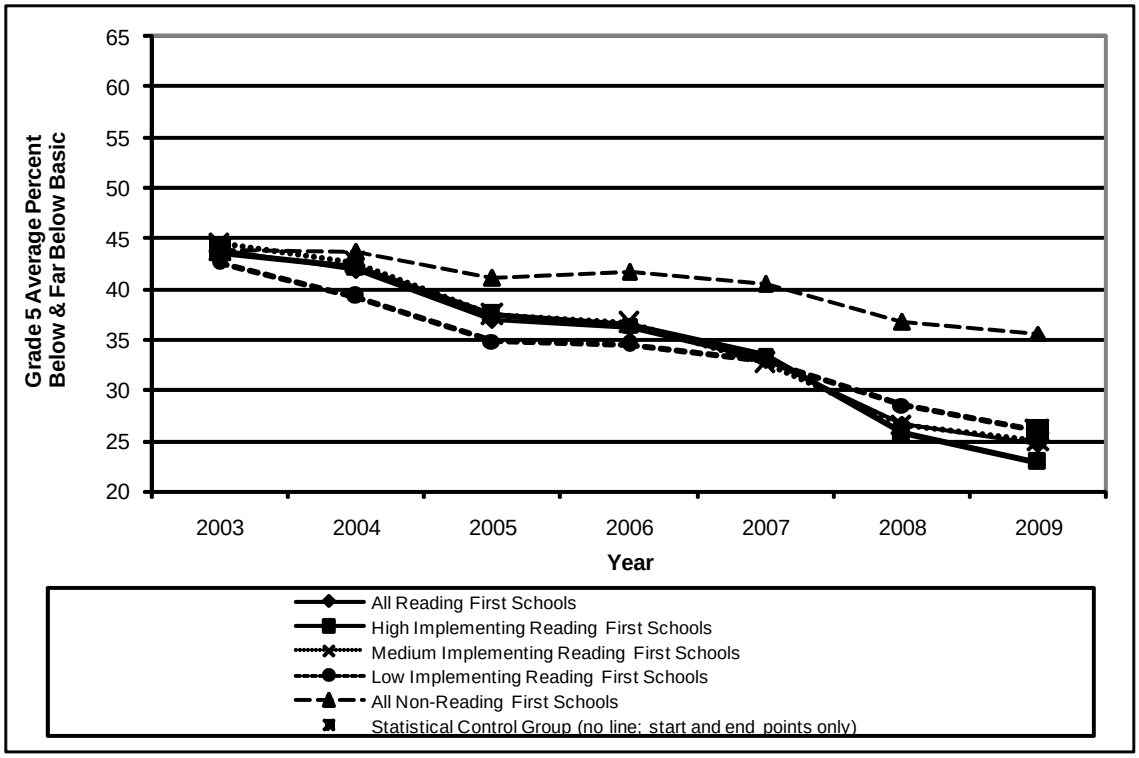
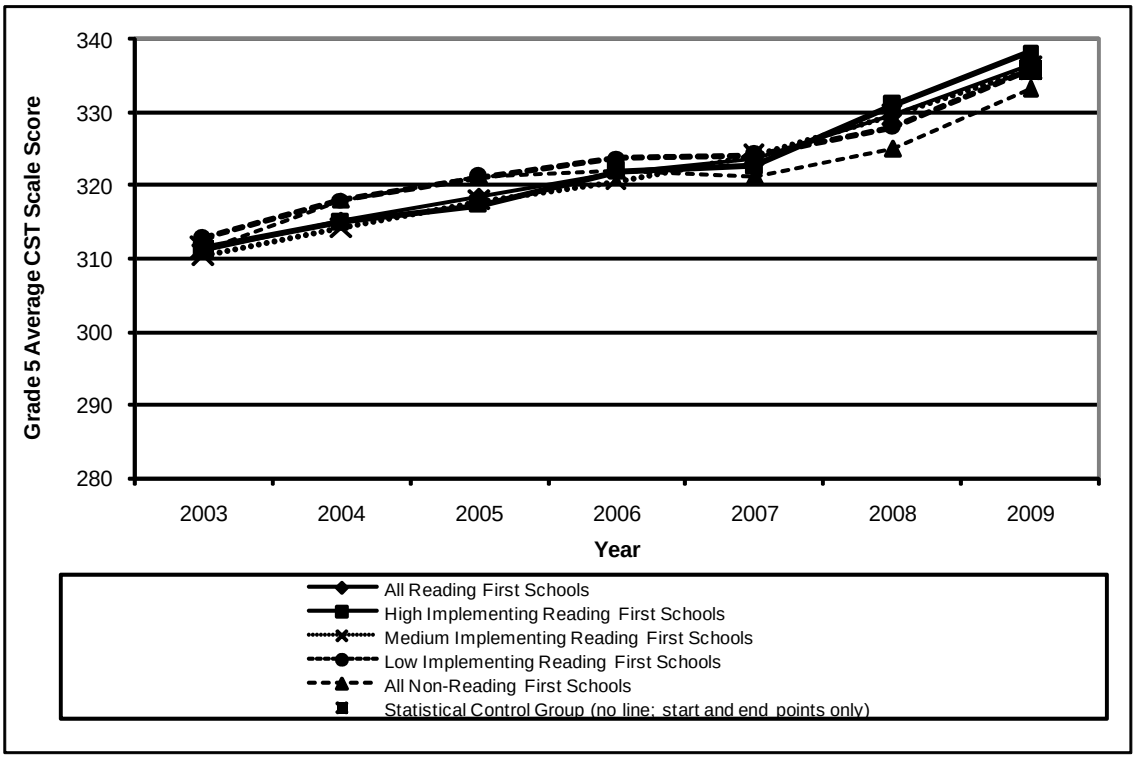


**Figure 2.5c: CST Mean Scale Score, YIP = 6, Grade = 5**




Figures 2.5a - 2.5c show very clearly that the low implementing schools performed better than the high implementing schools until students who were raised on Reading First since kindergarten or first grade began entering grade 5 classrooms. The implementation fan pattern is reversed until 2007, at which point it starts to widen with high implementing schools on top. This strongly reinforces the pattern found in grade 4 and supports the hypothesis that Reading First is preparing students for work in the higher grades.

The grade 4 and 5 effects in this and previous reports strongly support the strategy of focusing on reading and language arts in the early grades by providing funds, professional development, coaching, and curricular coherence. This is consistent with extensive research that documents the importance of a strong foundation of early reading development, a concept that is also central to the national Reading First initiative, but is delivered through a unique model in California (e.g., Foorman & Torgesen, 2001; National Reading Panel, 2000; Snow, Burns & Griffin, 1998).

## Achievement Results for English Learners (ELs)

Achievement gains for the English learner (EL) subgroup of students are presented in this section. English Learners are identified according to their performance on the California English Language Development Test (CELDT), the results of which are recorded in the California STAR file. In YIP 6 schools (Cohort 2), 55% of the students are classified as English Learners.

Two achievement metrics are reported: the percentage of EL students per school that are in the Proficient or Advanced CST performance categories (Percent Proficient and Above), and the average CST English Language Arts scale score of EL students in the grade (Mean Scale Score). Percent Below or Far Below Basic are not included due to data missing from the STAR file. The number of schools reported in this section is lower than that reported earlier in this chapter because some schools lack CST data for the English learner subgroup. This is especially noticeable for the group of non-Reading First schools. Statistical Control Group statistics were not calculated for this set of analyses.

CST Results for Grade 2 English Learners (Table 2.6 and Figures 2.6a – 2.6b)

Table 2.6 reports starting and ending grade 2 CST scores of English Learner (EL) students in schools that have been in the program six years (YIP 6, generally from Cohort 2). The first column of achievement gains duplicates the "All Reading First Schools" data that is reported in Table 2.2. The gains in the four columns headed "English Learner Students" were computed using *only* data for the EL subgroup.

**Table 2.6: English Learners, CSTs, YIP = 6, Grade = 2**

| | Reading First Schools | | | | | |
|---|---|---|---|---|---|---|
| | | English Learner Students | | | | |
| **Years in Program: 6 Grade: 2** | All Reading First Schools All Students | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | All Non-Reading First Elementary Schools |
| Number of Schools | 289 | 278 | 70 | 148 | 60 | 2180 |
| % Proficient and Above | | | | | | |
| 2003 | 20.7 | 14.4 | 14.1 | 14.5 | 14.6 | 23.4 |
| 2009 | 39.9 | 35.9 | 37.5 | 36.7 | 32.1 | 41.1 |
| Change Since Starting Year | **19.2** | **21.5[bc]** | **23.4[bc]** | **22.1[bc]** | **17.5[c]** | **17.7** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 310.3 | 301.5 | 302.6 | 300.7 | 301.9 | 315.5 |
| 2009 | 333.5 | 328.3 | 330.8 | 329.3 | 323.1 | 335.8 |
| Change Since Starting Year | **23.2** | **26.9[bc]** | **28.2[bc]** | **28.5[bc]** | **21.2[c]** | **20.3** |

[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."

[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

Table 2.6 reinforces the patterns observed in previous reports. Reading First is even more effective with English Learners than with the population as a whole. English Learners gained 26.9 points from 2003 to 2009 while the entire population gained 23.2 points. The pattern is particularly evident in High and Medium Implementation schools. English Learners in Medium Implementation schools had a mean scale score of 300.7 in 2003, a 10 point deficit relative to the population as a whole. By 2009 that deficit had closed to around 4 points, a gain of 28.5 points. In other words, English Learners in the YIP 6 schools have almost caught up with the population as a whole. (The Year 6 Report shows that the previous cohort of schools had completely caught up to the whole population by 2008.) On the other hand, English Learners in Low Implementation schools lag the whole population by 2 points. This shows that Reading First is helping to close the achievement gap so long as it is strongly implemented, in effect delivering on the promise to "leave no child behind." It is important to remember, however, that the whole Reading First population consists largely of English Learners, so the comparison of the English Learner subgroup with the whole population is not particularly useful in measuring the achievement gap.

Table 2.6 also shows that English Learners in Reading First schools have significantly larger gains (26.9 vs. 20.3 scale score points) than English Learners in non-Reading First schools.

Figures 2.6a and 2.6b illustrate Table 2.6 with trend-lines.

**Figure 2.6a: English Learners, CST % Proficient & Above, YIP = 6, Grade = 2**

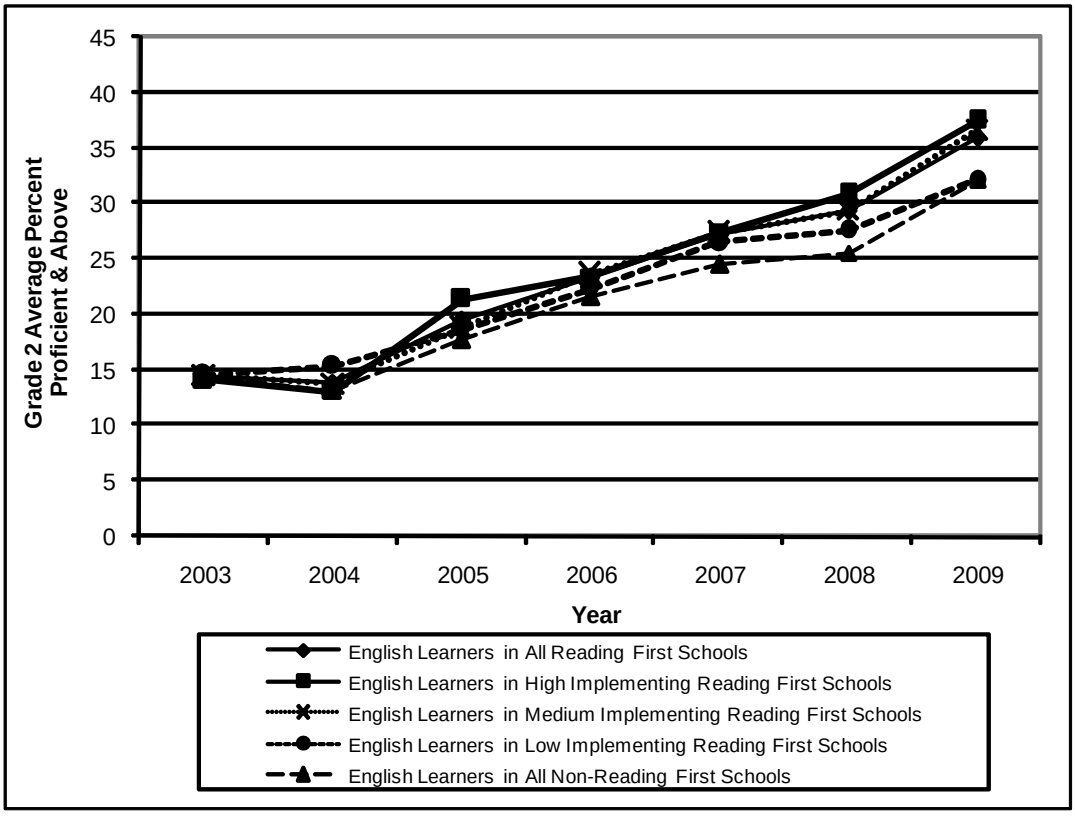
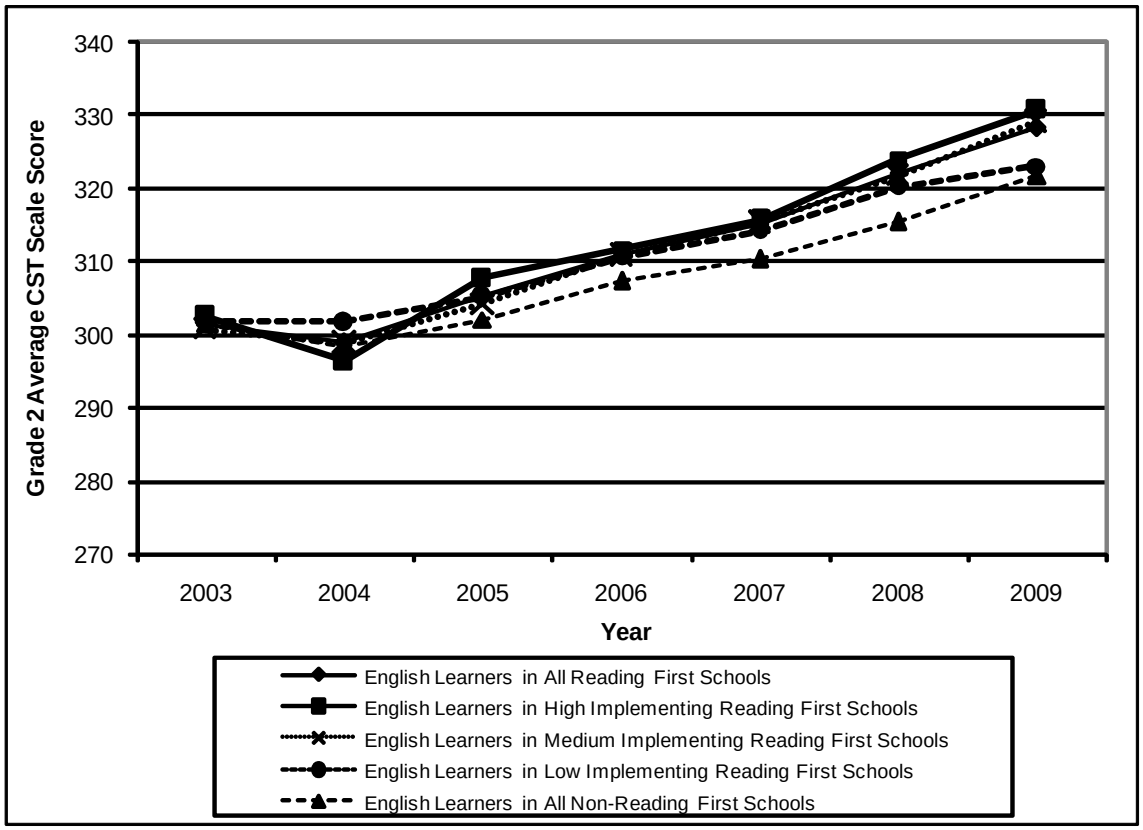


**Figure 2.6b: English Learners, CST Mean Scale Score, YIP = 6, Grade = 2**




Figures 2.6a and 2.6b show that with the exception of 2004, high implementing schools have posted higher scores and have increased the size of the gap with low implementing schools and non-Reading First schools. The "fan" shape has in fact widened in 2008 and 2009, supporting the idea raised in earlier sections that the 2009 YIP 6 schools have only gained traction in the last two years or so.

CST Results for Grade 3 English Learners (Table 2.7 and Figures 2.7a – 2.7b)

Table 2.7 shows CST results for grade 3 English learners in YIP 6.

**Table 2.7: English Learners, CSTs, YIP = 6, Grade = 3**

| | Reading First Schools | | | | | |
|---|---|---|---|---|---|---|
| | | English Learner Students | | | | |
| **Years in Program: 6 Grade: 3** | All Reading First Schools All Students | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | All Non-Reading First Elementary Schools |
| Number of Schools | 291 | 274 | 69 | 146 | 59 | 2055 |
| % Proficient and Above | | | | | | |
| 2003 | 16.8 | 9.7 | 9.3 | 9.5 | 10.7 | 16.6 |
| 2009 | 28.7 | 18.1 | 18.2 | 17.9 | 18.3 | 22.7 |
| Change Since Starting Year | **11.9** | **8.3[bc]** | **8.8[c]** | **8.4[bc]** | **7.6[c]** | **6.1** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 298.3 | 285.4 | 284.7 | 284.6 | 288.3 | 300.0 |
| 2009 | 318.3 | 303.6 | 304.6 | 303.8 | 302.1 | 311.1 |
| Change Since Starting Year | **20.0** | **18.2[bc]** | **19.9[bc]** | **19.2[bc]** | **13.7[c]** | **11.1** |

[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."

[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

Unlike Table 2.6, Table 2.7 shows that English Learners did not score as high as the whole Reading First population (18.2 vs. 20.0 scale score points). It is not immediately clear why English Learners lag the whole population more in grade 3 than in grade 2. It may be that high-performing English Learners are starting to be reclassified as fluent in grade 3, which would lower the average score of the remaining English Learners. However that may be, the pattern remains that English Learners in high and medium implementing schools post higher gains than English Learners in low implementing and non-Reading First schools.

**Figure 2.7a: English Learners, CST % Proficient & Above, YIP = 6, Grade = 3**

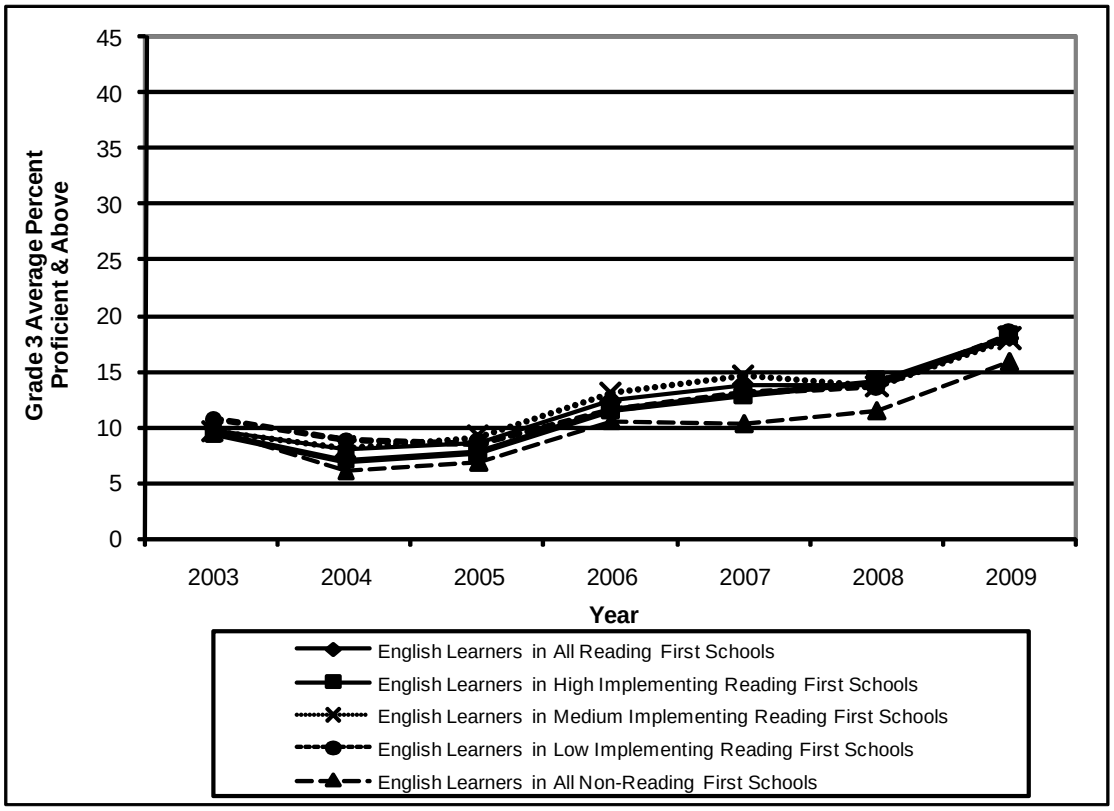
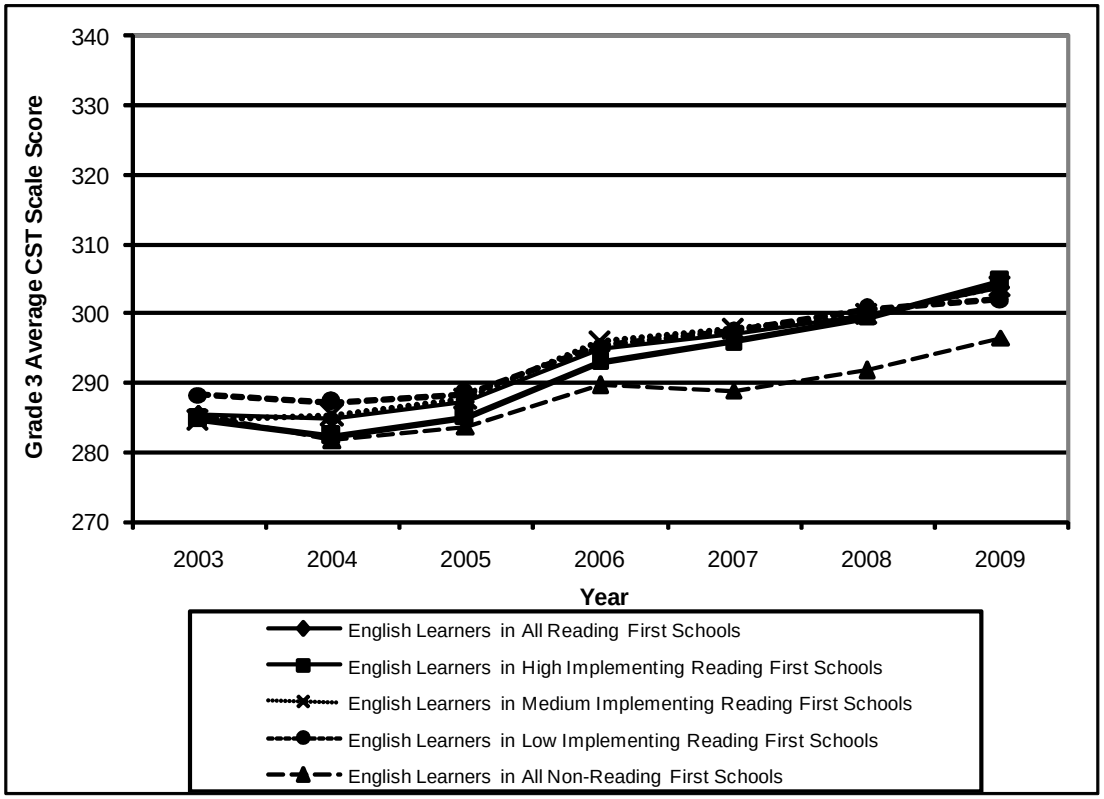



**Figure 2.7b: English Learners, CST Mean Scale Score, YIP = 6, Grade = 3**



Figures 2.7a and 2.7b, a bit like the grade 3 trend-lines in Figures 2.3a and 2.3b, show a "reverse fan", the high implementing schools lagging the low implementing schools in 2003 but gradually making up the difference over time. They pull even in 2008 and exceed the low implementing schools in 2009. It is important to note that a "reverse fan" can be just as much a sign of program efficacy as a regular "fan" pattern; what matters is the overall size of the gain and the relative steepness of the trend-lines. For whatever reason, the high and medium implementing schools started off behind, but grew more quickly than the low implementing schools and caught up with them.

Figures 2.7a and 2.7b also show higher growth rates for all the Reading First schools than the non-Reading First schools.

CST Results for Grade 4 English Learners (Table 2.8 and Figures 2.8a – 2.8b)

**Table 2.8: English Learners, CSTs, YIP = 6, Grade = 4**

| | Reading First Schools | | | | | |
|---|---|---|---|---|---|---|
| | | English Learner Students | | | | |
| **Years in Program: 6<br>Grade: 4** | All Reading First Schools All Students | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | All Non-Reading First Elementary Schools |
| Number of Schools | 279 | 262 | 68 | 139 | 55 | 1877 |
| % Proficient and Above | | | | | | |
| 2003 | 21.4 | 10.9 | 11.2 | 10.3 | 11.9 | 17.1 |
| 2009 | 46.6 | 29.6 | 28.5 | 29.3 | 31.9 | 36.1 |
| Change Since Starting Year | **25.2** | **18.7[c]** | **17.4[c]** | **18.9[c]** | **20.0[c]** | **19.0** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 317.3 | 304.7 | 304.4 | 303.8 | 307.3 | 313.8 |
| 2009 | 345.5 | 325.9 | 325.2 | 325.1 | 328.5 | 332.4 |
| Change Since Starting Year | **28.2** | **21.2[bc]** | **20.8[c]** | **21.3[c]** | **21.3[c]** | **18.6** |

[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

In Table 2.8, the Reading First effect disappears for English Learners. Previous reports have noted this grade 4 pattern and offered explanations. One possible explanation is that high-scoring English Learners start being reclassified to "fluent" in grade 4 in many districts based on CST performance, causing the remaining English Learners to have lower scores on average. Another explanation is that English Learners are especially sensitive to low implementation, and Reading First is not implemented at all in grade 4.

Figures 2.8a and 2.8b show what the lack of an effect looks like.

**Figure 2.8a: English Learners, CST % Proficient & Above, YIP = 6, Grade = 4**

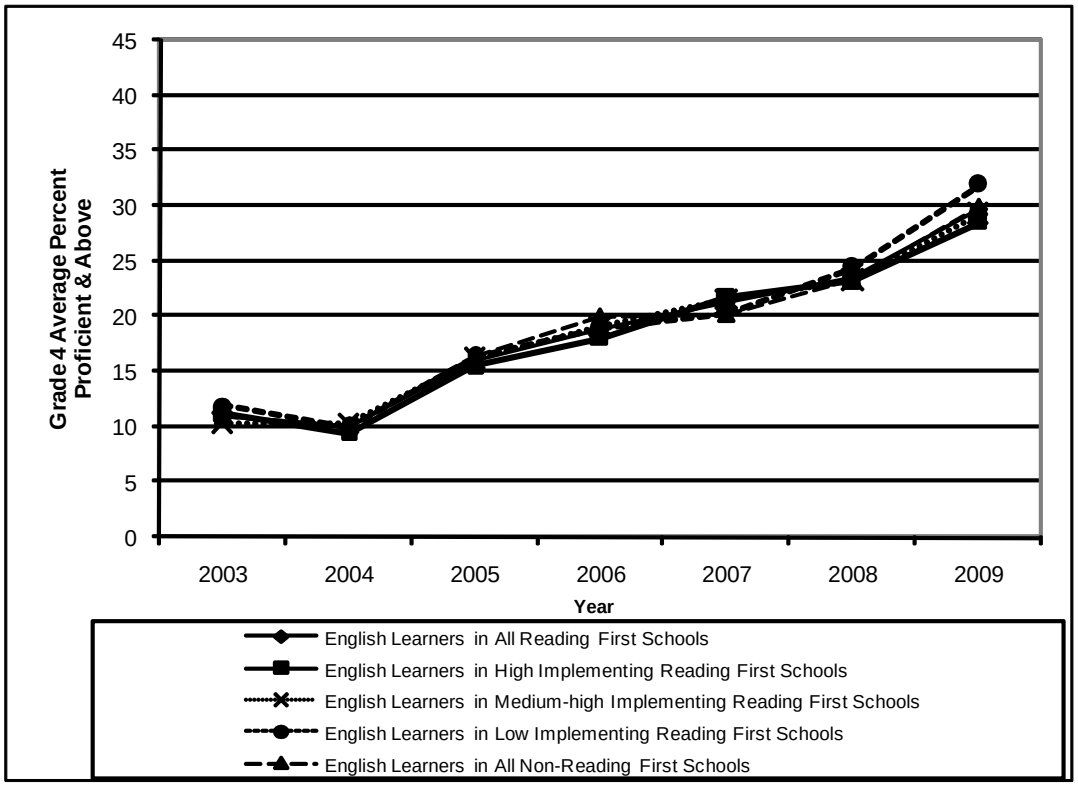



**Figure 2.8b: English Learners, CST Mean Scale Score, YIP = 6, Grade = 4**

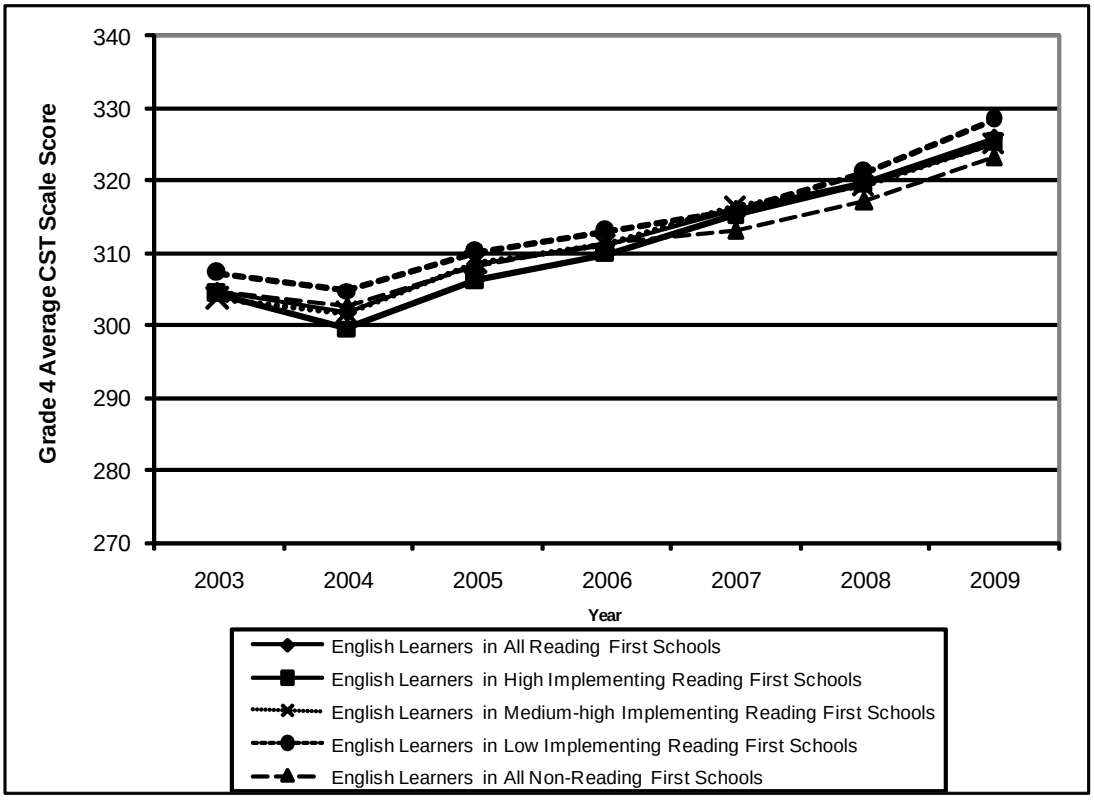


Figure 2.8a shows virtually no differentiation in English Learner achievement across levels of implementation. English Learners in Reading First schools are indistinguishable from those in non-Reading First schools for grade 4. The trend-lines in Figure 2.8b, on the scale score metric, are slightly more pronounced. Here, English Learners do have significantly higher growth than those in non-Reading First schools, but not to a degree comparable with the other grades.

CST Results for Grade 5 English Learners (Table 2.9 and Figures 2.9a – 2.9b)

Table 2.9 provides CST gain scores for EL Reading First students in grade 5.

**Table 2.9: English Learners, CSTs, YIP = 6, Grade = 5**

| Years in Program: 6 Grade: 5 | Reading First Schools | | | | | |
|---|---|---|---|---|---|---|
| | | English Learner Students | | | | |
| | All Reading First Schools All Students | All Reading First Schools | High Implementation Schools (RFII > 41.4) | Medium Implementation Schools (36.0 < RFII < 41.4) | Low Implementation Schools (RFII < 36.0) | All Non-Reading First Elementary Schools |
| Number of Schools | 275 | 244 | 64 | 129 | 51 | 1682 |
| % Proficient and Above | | | | | | |
| 2003 | 17.9 | 5.9 | 7.0 | 5.3 | 6.1 | 9.8 |
| 2009 | 39.3 | 17.6 | 17.8 | 17.0 | 19.1 | 21.0 |
| Change Since Starting Year | **21.3** | **11.7[c]** | **10.8[c]** | **11.7[c]** | **12.9[c]** | **11.2** |
| Mean Scale Score Per Student | | | | | | |
| 2003 | 311.2 | 295.8 | 296.3 | 295.0 | 297.0 | 302.5 |
| 2009 | 336.6 | 311.8 | 313.3 | 310.7 | 312.6 | 316.2 |
| Change Since Starting Year | **25.4** | **16.0[bc]** | **17.1[c]** | **15.6[c]** | **15.7[c]** | **13.7** |

[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

The pattern for grade 5 repeats the EL pattern for grade 4. High implementing schools do not out-perform low-implementing schools. English Learners lag the Reading First population as a whole. Reading First English Learners out-perform their peers in non-Reading First schools only on the scale score metric.

**Figure 2.9a: English Learners, CST % Proficient & Above, YIP = 6, Grade = 5**

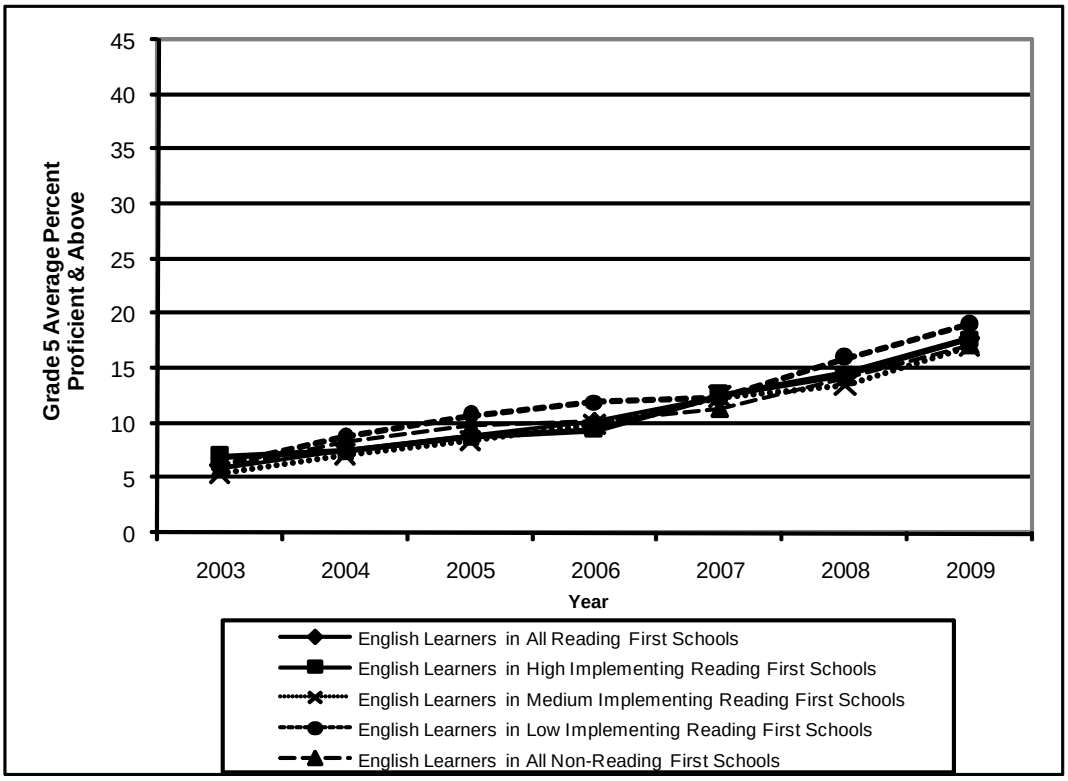
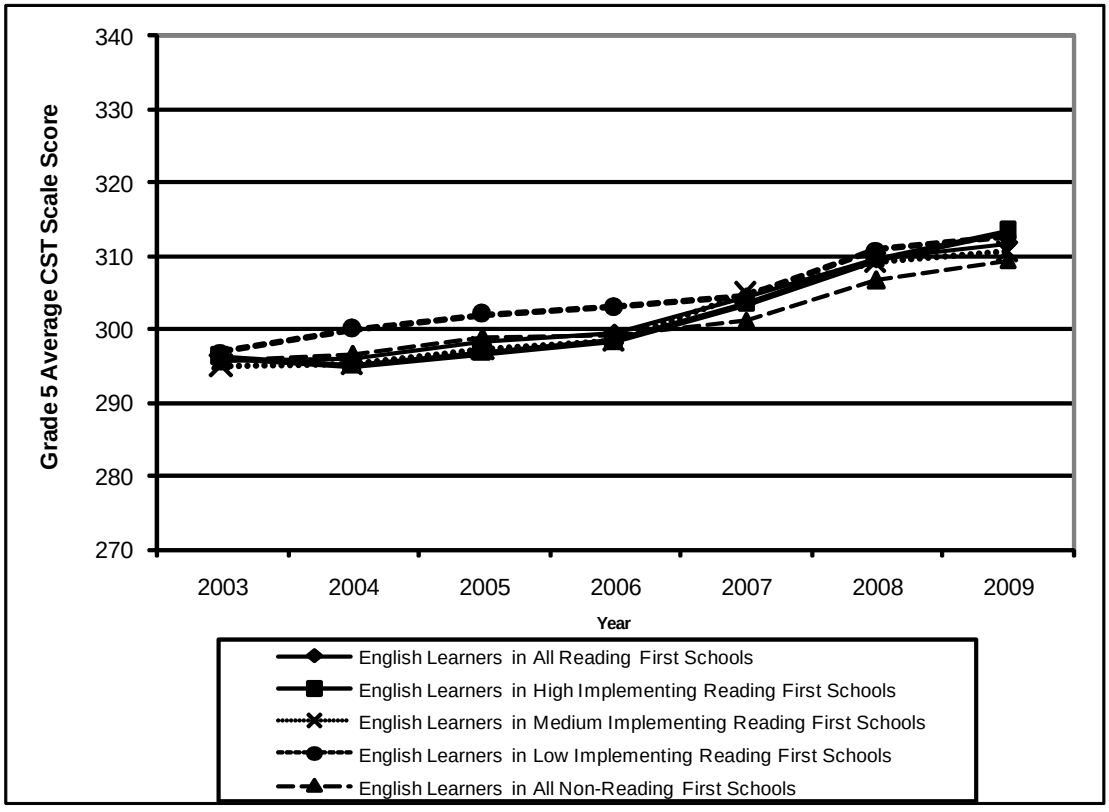



**Figure 2.9b: English Learners, CST Mean Scale Score, YIP = 6, Grade = 5**



The grade 5 trend-lines are similar to those for grade 4. The Year 6 Report Chapter 2 meta-analysis and Chapter 6 on English Learners explore the grade 4 and grade 5 questions in some detail. They discuss the hypothesis that the relative lack of Reading First effect for English Learners might be a statistical artifact of LEA reclassification criteria, the criteria used to decide whether to designate a student as "fluent" and remove him or her from the English Learner population. Generally these criteria, which vary widely across LEAs, are applied starting in grades 3 and 4 and put a lot of weight on whether a student scores Proficient on the CST.

Regardless of whether this hypothesis is applicable in this case, the statistical artifact is a serious one and impossible to ignore. It leads to the following paradox when evaluating a program. The more successful the program is, the larger the percent of students that will score Proficient on the CSTs. To the degree reclassification depends on scoring Proficient, these students will no longer be counted as English Learners and will enter the general population. This has two effects: a) the remaining English Learners who define the English Learner population will be those who are not able to score Proficient; b) the scores of the non-English Learner population will be dragged down by the influx of recently reclassified English Learners who will tend to be only marginally Proficient. Therefore, the more successful the program, the lower the average scores of the English Learner population will be, and the lower the average scores of the non-English Learner population as well. That is the paradox.

The fact that the Reading First effect is more apparent on the scale score metric than the Proficient and Above metric, as evidenced in the 2.8 and 2.9 figures, argues in favor of this hypothesis. The Proficient and Above metric is the most vulnerable to the statistical artifact since "Proficient" is the criterion generally used to reclassify English Learners. The average scale score metric, on the other hand, includes scores of students on the lower extremes of the scale where the reclassification artifact is less powerful. (Unfortunately, the STAR file lacks data on movement of English Learners out of the Below and Far Below Basic categories, which is crucial to answering this question.) We consider it likely, therefore, that the diminished grade 4 and grade 5 Reading First effect for English Learners is indeed the effect of reclassification. However, we have no definite way to test this hypothesis and in the absence of better statistical control it remains little more than speculation.

## Conclusions

The conclusions in the Year 7 Report reinforce and extend those of the Year 6 Report. We began the chapter by stating that Reading First would be said to show evidence of being effective to the degree that:

1. Achievement gains in Reading First schools are positive for grades 2, 3, 4, and 5.
2. Reading First schools show higher achievement gains than non-Reading First schools for grades 2, 3, 4, and 5.
3. Reading First schools show higher achievement gains than what would be predicted from a statistical control group for grades 2, 3, 4, and 5.
4. High implementing Reading First schools show higher achievement gains than low implementing Reading First schools for grades 2, 3, 4, and 5.
5. The average of the effects of Reading First implementation across all achievement metrics, as calculated using multiple regression to control for confounding demographic factors, is significantly greater than zero, with 95% confidence.

The Year 7 Report, taken in conjunction with the more comprehensive Year 6 Report, finds that properly interpreted the answer is a qualified "yes" to all five questions though the differences are not all statistically significant for this cohort of YIP 6 schools, and though the grade 4 and grade 5 effects are weak for English Learners (for reasons that may involve English Learner reclassification artifacts). The fifth test – averaging of all Reading First effects since the beginning of the program (but not including Year 7) – was conducted in the Year 6 Report and constitutes the final word regarding Reading First program effectiveness. It found that when averaged across all years and metrics up to and including 2008, the Reading First effect has been quite significant statistically, more than 15 standard errors greater than zero where two standard errors greater than zero would be sufficient to claim "significance". Statistically the Reading First effect is real. The effect is meaningful, as well. Reading First is approximately 60% as powerful in impacting achievement as such well-established demographic variables as percent of SED, EL, black, and migrant students per school. The Year 6 Report also showed how a doubling of implementation can effectively double school achievement gain scores.

The Year 7 Report, though it focuses on a cohort of schools that has been slower to respond to Reading First than other cohorts, does nothing to shed doubt on the central findings of the Year 6 Report. It confirms the findings of the previous evaluation reports and supports the hypothesis that students who were in Reading First programs in grades K-3 are better prepared for higher grades than students who were not.

We conclude by restating from the Year 4 and Year 5 Reports an idea that has implications for all schools in California. Reading First implementation, and thus Reading First exclusivity at the school site, is a significant predictor of positive cross-year gains. This fact supports the hypothesis that the upward trend in reading scores in Reading First schools since 2002 is the result of the program and not some other factor. Because the rest of the state K-3 schools have shown similar, though less dramatic, upward trends over the same time period, and because many non-Reading First schools have been found to be using Reading First-style program elements, it is likely that the statewide trend in non-Reading First schools has being driven by the same program elements that are driving the Reading First gains. This validates efforts to make such program elements available to all California elementary schools, not just those in Reading First.

## References

Cooper & Hedges, Eds. (1994). *The Handbook of Research Synthesis.* New York: Russell Sage Foundation.

Dhar, Heimbichner, McMillan & Moulton (2008). *The Reading First Supplemental Survey Report, March 2008.* Prepared by Educational Data Systems for the California Department of Education. Available at www.eddata.com/resources/publications/.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children, Learning Disabilities Research and Practice, 16(4), 203–212.

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.

National Reading Panel. (2000). Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: National Institute of Child Health and Human Development.

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.

Snow, C. S., Burns, S. M., & Griffin, P. (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.

U.S. Department of Education (2002). The No Child Left Behind (NCLB) Act, Title I, Part B, Subpart 1, Section 1201. Signed into federal law January 2002.

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.

# CHAPTER 3: IMPLEMENTATION

This chapter presents data gathered from surveys of Reading First participants used to address the question: *How well has the Reading First program been implemented in each participating school and district?* Principal, reading coach, and teacher surveys provide a global perspective on implementation in Reading First schools as well as information about specific dimensions of program implementation 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, Educational Data Systems (EDS) developed three surveys – one each for Reading First teachers, coaches, and principals – and administered them annually from 2004 to 2009. Because participation in the evaluation process is part of the commitment that local education agencies (LEAs) made when they applied for funding, the response rate on the surveys has been high. In 2009, a total of 8,852 usable surveys were received from teachers, 465 from reading coaches, and 476 from principals, totaling 9,763. An estimated 9,639 classrooms were funded. That, plus one principal and coach per 498 schools, suggests a total population of 10,635 possible respondents, yielding an estimated response rate of 92%, which is in line with past years.[1] Results of the surveys can be found in Appendices A – C of this report.

This chapter primarily discusses the analysis of the survey data to compute a Reading First Implementation Index (RFII) for each school. This index is used to evaluate the overall implementation at the school level.

Key points in this chapter are:

- 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).
- The RFII can be interpreted as a (theoretical) percentage of times that teachers rate their schools "more than adequate" on relevant survey questions.
- Most schools in the Reading First program have implemented the program "adequately" but, in 2009, the average degree of implementation has declined.
- It is possible for Reading First schools to significantly increase their implementation of the program. It is likely that the majority of schools could substantially improve their achievement scores by doing so.

[1]For response rates and specific information from previous years, the reader is referred to past reports available at: www.eddata.com/resources/publications/.

- School-level implementation (institutional support of the teacher by the principal and coach) and teacher evaluations of Reading First are the two strongest predictors of achievement gains. This positive predictive effect is strong enough to offset large demographic effects.

## Measuring Reading First Program Implementation

To fully evaluate the effectiveness of an educational program, it is not enough to look at student achievement gains alone. It is necessary to examine achievement gains in relation to the degree of implementation of the program elements, or implementation fidelity (Dane & Schneider, 1998; Ruiz-Primo, 2006). 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 has an impact on achievement, the ultimate desired program outcome. If achievement gains bear no relation to the degree of program implementation, no evidence of program efficacy can be claimed (Schiller, 2001).

Fidelity of implementation is defined as "the degree to which an intervention [or program] is implemented as planned" (Gresham, Gansle & Noell, 1993). Studies of implementation have found significant correlations between degree of implementation of an educational program and student outcomes (Dane & Schneider, 1998; Leinhardt, Zigmond & Cooley, 1981). Therefore, the monitoring of implementation fidelity provides evidence regarding the extent to which the program elements are being applied according to design so that those responsible for program oversight can determine whether adjustments are needed to improve effectiveness (Power, Blom-Hoffman, Clarke, Riley-Tillman, Kelleher, & Manz, 2005).

In this chapter, we use survey data to quantify the degree of implementation occurring within each Reading First school. For each school, multiple respondents completed the survey, providing the perspectives of the site principal, the reading coach, and participating teachers. A school that may report a low level of use of curricular materials, neglects professional development, or inadequately allocates or employs instructional time, for example, would not be considered to be implementing the program. 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.

Rationale for Using a Survey

To directly measure the presence, absence, or degree of implementation of Reading First in *all* participating schools and districts is a daunting task. There is no statewide database that would definitively reflect Reading First implementation, and it is impossible within the scope of this evaluation to conduct observations at all sites. In 2009 there were 498 Reading First schools in California, and over

800 Reading First schools in previous years. 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). However, the State has specifically solicited in its Request for Proposals an implementation measure for *all* Reading First schools. To obtain information about implementation from all Reading First schools and districts, teachers, principals, and reading coaches in all Reading First schools were asked to complete a comprehensive survey constructed to gather information about the presence, absence, and degree of utilization of the critical elements that define the implementation of the Reading First program. Anecdotal information received from teachers and coaches indicates that it took 20 to 30 minutes to complete the survey.

The advantage of using a survey is that it is feasible to administer and analyze results from all schools, 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 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 may be a strong incentive for them 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 have questions tailored 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 impairs 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 issues have been discussed at length in previous reports and accepted survey analysis models have been used to ameliorate these potential limitations throughout the six years of the survey use.[2] To summarize, the above issues are addressed as follows:

1. Schools are measured relative to each other rather than against an absolute standard.
2. Teachers complete the survey anonymously, enhancing their ability to report truthfully about the program. Because in most schools there is only one principal and one reading coach, their responses are not entirely anonymous, though school code numbers and not school names are used in the analysis process. A school's implementation measure pools together the teacher, principal, and coach responses.
3. Questions are worded so that their "correct" answers are not immediately obvious, increasing the chance that respondents select truthful answers.
4. There are numerous opportunities for cross-verification of findings across respondents within a school. Respondents not only report their own use of program elements but also rate other respondent types (coaches rate teachers, teachers rate coaches, etc.).
5. The implementation survey provides data that are used for making program adjustments and no "high-stakes" funding decisions rest on results. The "significant progress" regulations[3] approved in fall 2007 were based entirely on achievement data.
6. Equating methods are used to equate responses across respondent groups and across program years.
7. The potential ambiguity at the question level is addressed by using statistical methods to group items into coherent "dimensions" that cluster together statistically and are validated by experts in the California Technical Assistance Center (C-TAC) and the Evaluation Advisory Group (EAG).

The reliability (Cronbach-alpha) of the Reading First Implementation Index has been established in previous reports and has ranged from .90 to .92 (a reliability of 0.85 is widely considered sufficient). Additionally, the *validity* of using the RFII as a measure of school-level implementation has been previously established. The fact that the RFII is a significant predictor of achievement growth is itself

[2] The reader is referred to previous annual reports at www.eddata.com/resources/publications/ for details about the development of the survey and analysis procedures.
[3] Information on "significant progress" is available at: http://www.cde.ca.gov/pd/ca/rl/rdfst06achievedef.asp .

evidence that it has a strong non-chance component. 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 consider the RFII to be sufficiently valid and reliable to be used for measuring implementation at the school level.

Changes to the Survey

From year to year, it has been necessary to make minor changes to the survey to reflect programmatic changes or to clarify ambiguous items. In each round of changes, equating procedures have been employed to allow for cross-year comparisons. The changes over time are summarized in this section.

Individual questions throughout the survey underwent editorial modifications, often to clarify routing from section to section on the web survey. In 2005, based on a change in the Reading First program to include Spanish curricular materials for waiver classrooms (instruction in Spanish), the teacher survey was expanded to include additional questions involving the receipt and use of the Spanish versions of curricula. In 2006, further revisions were made to clarify which curricular materials were referenced in specific questions. In 2007, very minor wording changes clarified some items thought to be potentially confusing or no longer relevant in a program that has been in place for several years. In 2008, the addition of a special education survey necessitated some changes to the teacher survey to facilitate routing on the web survey, but there were no changes made to items included in the calculation of the RFII. In 2009 the special education section was dropped, as well as the open-ended questions, to reflect the reduced Year 7 evaluation budget. In each round of revisions, efforts have always been made to retain enough "old questions" to link the different survey administrations together.

Calculating the Reading First Implementation Index (RFII)

Previous reports have described in detail the steps by which the RFII was constructed and how it is calculated. In short, the procedure is as follows:

Using an Item Response Theory (IRT) program called Facets, subsets of questions across the three surveys are used to generate measures on 17-19 dimensions.[4] IRT equating designs rely on common items that serve as links across forms and survey administrations. In 2008, the item difficulty calibrations which are the basis of survey equating were refreshed to correct for the effects of item "drift" over time.

[4] 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.

Most of the items showed little change in difficulty over time, but some types of items, in particular those asking about usage of program materials, had become easier over time, most likely because of increased familiarity and practice. The 2009 calibrations copied those for 2008. Item analysis was performed individually for each of the 19 dimensions.

Three of the 19 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).

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%

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). It is called the "Preliminary RFII". Based on advice from the EAG, as of 2007 the Preliminary RFII of a school in a given year is averaged with its Preliminary RFII from the preceding year (if one exists) to come up with a "Final RFII." Thus, the Final RFII assigned to each school in 2009 is an average of its 2009 Preliminary RFII and its 2008 Preliminary RFII. It was hoped that this 2-year rolling average approach makes each school's Final RFII more robust to changes in the sample of teachers in each school who take the survey while allowing it to be reflective of the school's recent implementation history. For purposes of this report, unless otherwise stated all references to the 2009 RFII signify the Final 2009 RFII, not the Preliminary 2009 RFII.

## Implementation Results

Distribution and Interpretation of the RFII

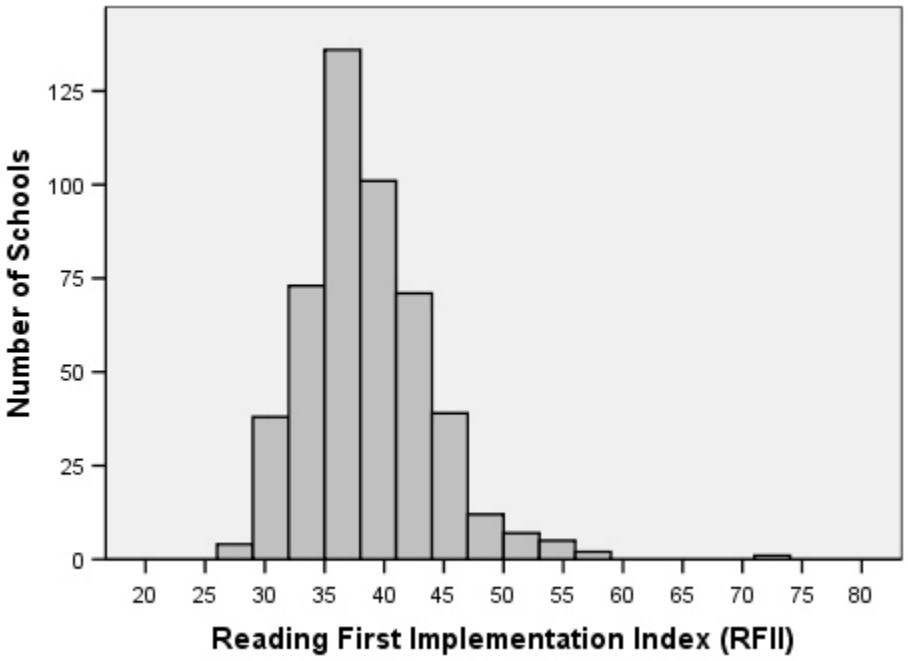
Figure 3.1 shows how the RFII was distributed across all Reading First schools in 2009. The mean 2009 RFII was 38.5; the standard deviation around the mean was 5.3. The mean of the Preliminary 2009 RFII (which does not include the 2008 RFII) is 37.7 with a standard deviation of 5.8. Thus, there has been a definite decline in level of implementation in 2009, about which more will be said.

An RFII of 38.5 can be practically interpreted as follows: Reading First *teachers* on average found their schools to be "more than adequate" 38.5% of the time (i.e., on 38.5% of the relevant items). However, interpreting the RFII as a percentage of items is not strictly correct. The RFII is actually based on a

statistical probability that teachers in a school will rate their school "more than adequate" across the test. It is a theoretical statistical parameter used to explain the data, not a literal count of responses.

Note the emphasis on teachers; the RFII was intentionally calibrated relative to teacher perceptions of "more than adequate implementation." Teachers tended to give lower scores to their schools than coaches and principals. While most of the dimension measures in Table 3.1 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 the table.

**Figure 3.1: Distribution of Schools – 2009 Reading First Implementation Index (RFII), N = 490, Mean = 38.5, S.D. = 5.3**




Dimensions of Implementation

Table 3.1 shows the dimensions derived in the RFII calculation process as well as the RFII itself, and their means for each year from 2004 to 2009. In the bottom row, we see the mean school RFIIs for 2004 through 2009 for all schools in the Reading First population. The 2009 RFII had a mean of 38.5 and a standard deviation of 5.3. The eighteen dimensions are listed along with the sections of the implementation surveys they most depend on, and the number of items in each dimension. Three of these dimensions, set in bold type, were used to calculate the RFII. The means in the columns by year may be interpreted as the average 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 column labels include the N counts of Reading First schools (that had RFIIs) for each year. For the Year 6 and Year 7 Reports the two columns under the heading "Effect on Achievement" report the results of a meta-analysis conducted in 2008 for the Year 6 Report where it is discussed in detail in Chapter 2. They allow us to quantify the effect of each dimension on achievement gains. Effect sizes are reported as a "standardized-beta coefficient," which is the number of standard deviations that the dependent or outcome variable (i.e., achievement) in a regression equation increases for each one standard deviation increase of the predictor variable. The table is sorted in descending order according to this effect size. All of these effects, except for Coach Professional Development, are significantly greater than 0 with 95% confidence.

To provide context, an additional column called "Effect Relative to Demographics" is provided. This tells how large the implementation dimension effect is relative to the average effect sizes of the demographic variables that are included in the regression equations. The demographic variables are: percent of Socio-Economically Disadvantaged (SED) students in the school, percent of English Learners (EL), percent of blacks, percent of migrant students, and number of students in the school (included even though it is an institutional variable). Their average effect size, generally negative, is converted to a positive number called the "Mean Absolute Demographic Effect Size." Thus, each dimension's effect size is divided by this demographic effect size to get its "Effect Relative to Demographics." A value of 1.00 means that the dimension predicts (and causes, for the most part) achievement to the same degree that the demographic variables do, i.e., it is a powerful, meaningful effect. The last column reports the correlation between each dimension and the 2008 (not 2009, which lacked Cohort 1 schools) RFII, thus their alignment. Correlations range from -1.00 to +1.00, where 0.00 means no relationship.

The "Effect on Achievement" and "Correlation" columns offer insights into how to improve achievement at the school level: focus on those dimensions that have the largest effect sizes and the largest correlations with the RFII, but where the school's measure on that dimension is low relative to the state average. We set aside those dimensions that contain the word "Evaluation" as these are not properly elements of the program, just opinions of it (though it is certainly true that a positive perception of the program improves implementation and achievement). We then refer to the indicated sections of the Reading First survey (Appendices A, B, C) and study the items contained there. By conforming teacher and school practice to these items, it is possible to generate meaningful gains in student achievement.

In interpreting the dimensions, note that some are contained within others. For instance, "School Implementation Overall" is composed of items from all the implementation dimensions.

**Table 3.1: All Schools, N (2009) = 490, Mean for Each Dimension, 2004-2009[1, 2, 3]**
**Sorted by Effect on Achievement**

| | | | % of the time teachers rated their school "More than Adequate" | | | | | | Effect on Achievement | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 04-08 | 04-08 | |
| | Dimension (refer to cited sections of questionnaires, Appendices 1-3, for relevant items) | # Items, 2008 | Mean (N = 628) | Mean (N = 808) | Mean (N = 856) | Mean (N = 885) | Mean (N = 863) | Mean (N = 490) | Std-Beta Effect | Effect Relative To Demogr. | Correlation with RFII |
| 1 | Teacher RF Evaluation (Section I, Teacher) | 4 | 14 | 14 | 16 | 15 | 17 | 18 | 0.098* | 1.03 | 0.56 |
| 2 | School Implementation, Instruction (Section D, Teacher) | 28 | 34 | 36 | 40 | 40 | 41 | 40 | 0.087* | 0.75 | 0.75 |
| 3 | **School Implementation Overall (Impl. Sections)** | **210** | **39** | **40** | **43** | **43** | **43** | **43** | **0.083*** | **0.64** | **0.96** |
| 4 | Teacher Implementation (Section F) | 33 | 48 | 50 | 54 | 54 | 54 | 52 | 0.075* | 0.46 | 0.64 |
| 5 | Principal RF Evaluation (Section I, Principal) | 6 | 23 | 24 | 23 | 20 | 22 | 26 | 0.071* | 0.38 | 0.21 |
| 6 | Coach RF Evaluation (Section I, Coach) | 6 | 20 | 19 | 24 | 23 | 25 | 29 | 0.071* | 0.38 | 0.26 |
| 7 | Principal RF Understanding (Section H, Principal) | 17 | 17 | 19 | 20 | 20 | 19 | 19 | 0.070* | 0.34 | 0.16 |
| 8 | School Implementation, Materials (Section C, Teacher) | 175 | 36 | 37 | 41 | 41 | 45 | 46 | 0.070* | 0.35 | 0.73 |
| 9 | Evaluation of Professional Development (Section B, Teacher) | 5 | 11 | 14 | 15 | 15 | 17 | 17 | 0.069* | 0.32 | 0.49 |
| 10 | Coaching Implementation (Section F, Coach) | 32 | 46 | 48 | 50 | 49 | 49 | 47 | 0.067* | 0.27 | 0.70 |
| 11 | **Overall RF Understanding (Section G, Teacher; Section H, Coach, Principal)** | **17** | **23** | **25** | **26** | **26** | **26** | **25** | **0.066*** | **0.25** | **0.26** |
| 12 | Implementation, Assurances (Section C, Principal) | 11 | 44 | 48 | 46 | 45 | 41 | 49 | 0.066* | 0.24 | 0.49 |
| 13 | Principal Professional Development (Section B, Principal) | 3 | 48 | 46 | 57 | 56 | 57 | 63 | 0.065* | 0.22 | 0.15 |
| 14 | Coach RF Understanding (Section H, Coach) | 17 | 36 | 39 | 38 | 39 | 31 | 30 | 0.064* | 0.21 | 0.10 |
| 15 | Teacher Professional Development (Section B, Teacher) | 9 | 38 | 36 | 35 | 34 | 30 | 29 | 0.062* | 0.16 | 0.50 |
| 16 | Teacher RF Understanding (Section G, Teacher) | 17 | 27 | 29 | 30 | 30 | 30 | 29 | 0.062* | 0.16 | 0.23 |
| 17 | **Teacher Coach Professional Development (Section B, Teacher, Coach)** | **11** | **40** | **37** | **35** | **34** | **28** | **27** | **0.060*** | **0.10** | **0.46** |
| 18 | Coach Professional Development (Section B, Coach) | 7 | 58 | 56 | 48 | 33 | 39 | 32 | 0.058 | 0.07 | 0.24 |
| **19** | **RF Implementation Index (RFII)** | **238** | **36** | **36** | **39** | **39** | **39** | **38** | **0.082*** | **0.63** | **1.00** |

1-Dimensions 3, 11, and 17 are in bold because they are weighted contributors to Dimension 19, the RFII. The 2009 statistics are across 490 schools from the point of view of teachers for most dimensions. Dimensions 6, 10, and 18 are from the point of view of coaches. Dimensions 5, 7, and 13 are from the point of view of principals.
2-The two columns called Effect on Achievement are effect sizes derived using a meta-analysis of 221 regressions, conducted in 2008. The left column is the "standardized-beta coefficient." The asterisk "*" means the effect is statistically significant at the 95% confidence level. The right column is the RFII dimension effect relative to the Mean Absolute Demographic Effect for that dimension. A value of 1.0 means the dimension has the same predictive power as the demographic variables, on average.
3-The statistics in the right column report each dimension's correlation with the RFII. The closer to 1.00, the more it captures what is meant by "implementation" as embodied by the RFII.

Trends in Implementation

The year 2009 saw a drop from 863 schools with RFIIs to 490, reflecting the defunding of the Cohort 1 LEAs that occurred in 2008-09. This cautions us to be careful in drawing conclusions about trends that include 2009. However, Table 3.2 (the mean preliminary RFII of schools that had an RFII in 2009) shows us that the overall decline in the RFII is not a sampling artifact.

**Table 3.2: Mean Preliminary RFII, Schools that have an RFII in 2009**

| | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|
| Mean | 35.78 | 36.79 | 40.05 | 40.02 | 39.40 | 37.76 |
| SD | 6.37 | 5.38 | 6.15 | 5.42 | 5.86 | 5.82 |

Using the "Preliminary RFII" statistic (not averaged with the previous year), Tables 3.2 shows us that the RFII for these Cohort 2-4 schools began around 36 in 2004, peaked at 40 in 2006 and 2007, and has declined since then. This trend is borne out by anecdotal reports from the R-TACs.

If we study the trends for each dimension (acknowledging that 2009 is harder to interpret given its reduced sample size), we see that a number of dimensions peaked around 2006 and 2007 and started declining in 2008 and 2009. Examples are: Teacher Implementation, Principal Understanding, Coaching Implementation, Overall Understanding, Coach Understanding, and Teacher and Coach Professional Development. The decline in Professional Development is especially pronounced. There is also some decline in Overall Understanding, which suggests either that some teachers have forgotten what they have learned, or they are choosing to teach differently.

There is a striking decline in the coaching variables. Coaching Implementation has declined steadily since peaking in 2006. Coaching Professional Development has strongly declined ever since 2004, the first year of the RFII. Coach Understanding of Reading First principles has dropped 9 points in the last two years. Nonetheless, coach evaluations of Reading First have climbed steadily. These patterns are a strong warning that a central component of Reading First – coaching – has been under-supported for several years and that with the end of the program it is fading rapidly.

There are other dimensions that have been trending upwards more or less steadily since 2004. These include: Teacher Evaluations of Reading First, Principal Evaluations (except for a dip in the middle years), Coach Evaluations, Evaluations of Professional Development, use of program Materials, Implementation of Assurances (with a dip in the middle years paralleling Principal Evaluations), and Principal Professional Development.

Two themes emerge involving principals and evaluations. Principals appear to have had second thoughts about Reading First in the middle years of the program. But by 2009 (faced with loss of funding) they have become much more involved and supportive. The other theme is that evaluations of the Reading First program have steadily improved with time. As teacher comments revealed, there was a great deal of skepticism from 2003 - 2005, when the program began. The improving perceptions of the program are a strong indicator that it has been accepted as a useful and trusted component of elementary school education in the Reading First schools.

Effectiveness of Reading First Components

Table 3.1 also tells us which dimensions of Reading First have the biggest effect on achievement. It turns out that Overall Reading First Understanding and Teacher/Coach Professional Development, two of the three components that make up the RFII, have a relatively small impact on achievement – at least as they are realized in the survey. This is easily explained in the case of the Teacher/Coach Professional Development dimension, which has relatively few items, causing high measurement error which obscures the relationship to achievement. Also, the role of professional development has inevitably declined as the program matures and teachers move through the various levels. There are few teachers who have not received the initial 40 hours of AB 466/SB 472 training. This "leveling out" in training would cause the effect size to diminish.

As it happens, 70% of the weight of the survey resides with School Implementation Overall, a composite implementation measure that combines the various implementation dimensions and is the third most powerful predictor. The most powerful implementation dimension is School Implementation, Instruction. This dimension is built from items that ask about principal support of the teachers, planning time, the pacing schedule, grade-level meetings, and the principal's involvement in these meetings. It has to do with how well the school and principal supports the teaching staff. Related to this is Principal Understanding of Reading First, a dimension that was not used when the RFII was being constructed. This also turns out to be a strong predictor of student achievement, no doubt in combination with the School Implementation dimension. Thus we find that well-informed and active involvement of principals is essential to the success of the program.

This is good news from a school improvement perspective. It means that so long as principals are well informed and participate aggressively, they can help raise achievement dramatically across the elementary school reading program as a result of actions undertaken at the administrative level. The principal matters.

The most powerful non-implementation predictor of school achievement is "Teacher Reading First Evaluation" – how teachers evaluate the program at their school. While this statistic has never been particularly high, it has increased fairly steadily from 14 in 2004 to 18 in 2009. When teachers are

positive, schools grow rapidly. When they are negative toward the program, schools suffer. It is tempting to dismiss any causal relationship here, to say that when teachers feel negative about the program it is because of problems with the program, not with their own pre-dispositions toward the program; and that when teachers feel positive about the program, it is only because they are getting better achievement results. But it is also possible that teacher perceptions of the program – teacher "buy-in", in other words – have their own effect on the program's efficacy. We already know that the program is effective when implemented. It is not much of a reach to say that teachers are more likely to implement the program if they view it positively, and that they are less likely to implement it if they view it negatively. If that is true, the high predictive power of Teacher Evaluations of Reading First has at least some causal component.

The conclusion is simple and intuitive. While Reading First has potential to promote school growth, even in very challenging circumstances, it cannot happen without teacher buy-in and without strong institutional support. Strong principal participation and positive teacher attitudes help create Reading First effect sizes sufficiently large to counter-balance the powerful demographic pressures with which Reading First schools, in particular, must contend.

## Conclusions

Are Schools Implementing "adequately"? To interpret the implementation data, we rely on the procedures developed in prior reports that validate the RFII as a satisfactory measure of implementation. The RFII serves as a comparative benchmark for examining implementation by every school in the Reading First program. The RFII of an individual school can be viewed relative to some standard reference point that characterizes the population of schools as a whole. In the first year of implementation, the average RFII was 36. This became the (somewhat arbitrary) cut-point between "High Implementation" schools and "Low Implementation" schools. This distinction was used in conjunction with school achievement measures in other chapters to track the different achievement trend-lines for high implementing and low implementing Reading First schools (see Chapter 4 of the Year 4 Report, Chapter 2 of this report and of the Year 5 Report). To preserve comparability over time, the 36 as a cut-point continues to be used to define the upper boundary of the lower implementing schools. However, based on advice in 2007 from the EAG, the "High Implementation" schools were redefined to be at least one standard deviation above 36 – a new cut-point of 41.4. This has the benefit of sharpening the distinction between high and low implementing schools, but at the cost of leaving out schools that are in the mid-range between 36 and 41.4. In the current report (Chapter 2), that mid-range is defined as "Medium" and is included in the charts and tables.

Because the cut-point of 36 has over the course of the evaluation been used to distinguish high from low implementing schools, it serves as a reasonable definition of the lower bound of "Adequate."[5] By that criterion, the histogram in Figure 3.1 and Table 3.1 above reveal that schools are on average doing an "adequate" job of implementing the Reading First program, since the mean 2009 RFII of 38 is greater than 36 by a little less than half a standard deviation.

Examining the mean RFII over time, it appears that the index has risen modestly and begun to decline. In 2004 and 2005, the mean RFII was 36 while in 2006, 2007 and 2008 it was 39. In 2009 it is down to 38. Though adequate, schools are not improving their level of Reading First implementation and the program shows unmistakable signs of being phased out.

## References

Dane, A. V. & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? Clinical Psychology Review, 18(1), 23-45.

Gresham, F., Gansle, K. A., & Noell, H. (1993, January 1). Treatment Integrity in Applied Behavior Analysis with Children. *Journal of Applied Behavior Analysis, 26*(2), 257.

Leinhardt, G., Zigmond, N., & Cooley, W. (1981). Reading instruction and its effects. *American Educational Research Journal, 18*, 343-361.

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

Power, T., Blom-Hoffman, J., Clarke, A., Riley-Tillman, T., Kelleher, C., & Manz, P. (2005). Reconceptualizing Intervention Integrity: A Partnership-Based Framework for Linking Research with Practice. *Psychology in the Schools, 42*(5), 495-508.

Ruiz-Primo, M. (2006, February 1). A Multi-Method and Multi-Source Approach for Studying Fidelity of Implementation. CSE Report 677. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*, (ERIC Document Reproduction Service No. ED492864) Retrieved October 30, 2007, from ERIC database.

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

[5] Note, however, that this usage of the term "adequate" differs fundamentally from that used in previous reports. In the Year 4 Report and earlier, "adequate" was defined in a manner parallel to "more than adequate" – i.e., as a teacher's propensity to score a school in or above the "adequate" rating scale category for each item. While psychometrically defensible, this definition has proven needlessly confusing and is here replaced with a simpler "cut-point based" definition that is in harmony with how implementation is conceptualized in the achievement section of the evaluation.

# CHAPTER 4: LESSONS LEARNED FROM IMPLEMENTATION OF READING FIRST IN CALIFORNIA

The purpose of this chapter is to provide a cumulative review of the impact of Reading First on California's students and teachers. In this chapter, we draw on previous reports to reflect on the process and outcomes of the Reading First program in the state. The scope and range of Reading First funding in California is unprecedented in the state's educational history. The number of students, teachers, and administrators involved in Reading First schools is larger than any other state. In this chapter, we draw conclusions about how Reading First has impacted student achievement over time. Additionally, we summarize findings regarding implementation. In this chapter, we present lessons learned regarding student achievement and implementation. To that end, we summarize key findings from previous reports.

This chapter yields the following key findings:

- Reading First has led to consistent achievement gains in California. Various metrics demonstrated that Reading First schools realized significant achievement gains in reading compared to non-Reading First schools and a statistical control group across all years of the program. The Reading First Achievement Index steadily rose in Reading First schools.
- Reading First effects extended beyond the K-3 grades. The Year 5 report included analyses of grade 4 achievement, and the Year 6 report examined grade 5, to determine the long-term effects of the program for students whose K-3 teachers had participated in Reading First. Achievement gains were higher for students in these grades compared to non-Reading First schools and a statistical control group. Possible explanations include that schools or districts may have extended program elements such as coaching and professional development into upper grades, and students may have received a strong foundation in beginning reading skills during the K-3 years that enhanced their later success.
- Higher implementation is associated with higher achievement. Reports from Year 3 through Year 7 showed significantly higher achievement for high implementing schools compared to lower implementing schools, as measured by the Reading First Implementation Index and various achievement metrics.
- There are inherent difficulties in comparing Reading First schools with comparison or control schools. Features of the Reading First program were evident in non-Reading First schools due to statewide and district level initiatives. The original comparison group of Reading First-eligible schools and non-Reading First schools were demographically different from the Reading First

schools, making it impossible to definitively measure comparative effects of the program. A hypothetical statistical control group was formulated based on predictive models.

- Reading First showed consistent achievement gains for English Learners. Participation in Reading First led to significant achievement gains for Reading First schools with large numbers of English learners. Open-ended responses to survey questions indicated that teachers, coaches and principals viewed Reading First as effective for English learners.
- The Reading First program has created sustainable features of reading/language arts instruction. Participants reported positive regard for key elements of the program and the desire to maintain such features as collaborative planning time, a protected time block for reading/language arts, a common curriculum, coaching, professional development, and other features.
- Fidelity of Implementation is critical to achieving a positive impact on instructional practices and student achievement in reading. Monitoring and supporting implementation were factors in ensuring success. Implementation of the Reading First elements increased over time and, overall, the program achieved a fairly high degree of implementation.
- Professional Development was extensive and effective in Reading First. Basic and advanced levels of professional development and 80 hours of follow-up led to increased expertise in the research base of reading and how to fully implement state-adopted curricula.
- The use of research-based curricula and materials was a key feature of Reading First. Through ongoing monitoring and support, participants developed expertise in the adopted curricula and materials and came to view them as having a significant and positive impact on student achievement.
- Coaching has become well integrated into Reading First schools and has supported implementation. The Reading First program has played a significant role in establishing reading (or literacy) coaches in districts throughout the state. Coaches have developed a high level of expertise, are highly valued, and have had a significant impact on improving reading instruction in Reading First schools.

Reading First from a National Perspective

Few national educational initiatives have received as much attention as the Reading First initiative. In April 2008, the U.S. Department of Education Institute of Education Sciences released a national Reading First study, titled *Reading First Impact Study: Interim Report* (Gamse, Bloom, Kemple & Jacob, 2008), which included only 18 study sites, representing 13 states smaller in size than California. This report

caused a stir in reporting no significant impact of the program on reading comprehension achievement compared to non-Reading First schools, though they did find a significant impact on students' decoding skills in first grade. While noting "positive, statistically significant impacts on the five essential components of reading instruction[1] promoted by the program," the study also stated that "on average across the 18 study sites, Reading First did not have statistically significant impacts on reading comprehension test scores in grades 1-3." In contrast, our annual reports of reading achievement in California's Reading First schools showed a consistent and steady gain on various metrics of reading achievement for California's students, including English learners, discussed in depth below. It is noted that the California evaluation has a substantially higher sample size than the national study, that it has been conducted over a longer period of time, and that, by defining its control group in terms of degree of school-level Reading First implementation it avoids contamination of the control group by Reading First treatment elements. We consider it possible, even likely, that the findings of the national study lead to an incorrect conclusion regarding the efficacy of the Reading First program, and that the California evaluation leads to a more correct conclusion, certainly with respect to California, and possibly with respect to the rest of the country as well.

Additional findings from the national impact study focused on implementation, finding that participation in the Reading First program led to increased instructional time focused on the key elements of reading development and improved practices of teachers through professional development. Similarly, our external evaluation of California's Reading First program has supported and expanded on this finding. Our annual reports have included extensive information about the positive impact of Reading First on the nature and quality of classroom instruction, discussed in more depth below. A second national report investigated Reading First implementation. The Reading First Implementation Evaluation Final Report (Moss, Fountain, Boulay, Horst, Rodger & Brown-Lyons, 2008) found that Reading First schools were more likely than non-Reading First schools to devote sufficient time to reading instruction, have the benefit of an on-site reading coach to support instruction, use materials and methods that reflect the scientific evidence base in reading, provide extra support for struggling readers and engage in substantial professional development. This is consistent with implementation findings in California's evaluation reports.

[1] The five essential components of reading instruction are: phonemic awareness, phonics, vocabulary, fluency and comprehension.

## Lessons Learned from Achievement Data

Reading First has led to consistent achievement gains.

Each year of evaluating the Reading First program in California, findings have indicated that students in Reading First schools make significant gains in reading achievement using various achievement metrics. Reading achievement has been measured using the CST results in grades 2 and 3, a norm-referenced standardized test in grade 3, and End of Year (EOY) tests in grades K-3. In addition, we have calculated a Reading First Achievement Index for each participating school using these metrics in a weighted formula. All of these metrics have indicated that California's students have realized achievement gains associated with their schools' participation in the program. Reading growth has been consistent over time. The RFAI, a composite of three achievement metrics has risen steadily over seven years, indicating a steady rise in reading achievement.

Previous reports have compared the achievement results of Reading First schools, non-Reading First schools and a statistical control group. Consistently, using various metrics of reading achievement, Reading First schools have experienced significantly higher reading achievement gains than non-Reading First schools or a statistical control group. There were minor exceptions in some years with some cohorts of students, but overall, achievement growth has been higher in Reading First schools. For example, the Year 3 report indicated that in Grade 3 of Cohort 1, there was no growth on CST scores for Reading First schools, but the comparison groups also showed no growth, which may have been an anomaly associated with that particular test and year. Yet, achievement gains were significant and present for each year of reporting.

To determine an overall effect of the Reading First program on achievement, the Year 6 report included a meta-analysis of a series of multiple regressions to calculate an effect size controlling for numerous demographic variables and starting point for achievement from 2003 to 2008. This report showed a standardized Beta coefficient of .082 with a standard error of .004. This translates to an effect that is 16 standard errors higher than zero (no effect) where two standard errors above zero would indicate a statistically significant effect with 95% confidence. In other words, there is little doubt that the Reading First program has had a significant and positive impact on achievement for Reading First schools.

Reading First effects extended beyond the K-3 grades.

When students were educated in Reading First schools during their K-3 years, their achievement in grades 4 and 5 continued to show significant and positive effects for the CST, particularly in high implementing schools. Officially, Reading First program elements were focused on grades K-3, but two factors may have influenced the grades 4-5 effects. First, schools may have extended program elements such as the

research-based curriculum, professional development and coaching to the upper grades. Also, students may have received a strong foundation of reading skills in grades K-3 that enhanced their reading success in the upper grades.

Higher implementation is associated with higher achievement.

Over several years, analyses consistently showed a significant and positive relationship between implementation and achievement. This finding held true for K-3 students in general, but also particularly for English learners.

There are inherent difficulties in comparing Reading First schools with comparison or control schools.

From Years 1 through 3, the California evaluators attempted to make meaningful use of comparison groups by which to gauge the effect of Reading First. Like evaluations in other states, there were inherent problems with this process, and California's evaluation reports moved by Year 4 to using a statistical control group for comparison. In Years 1 and 2 of the evaluation, Reading First schools were compared to two groups of non-Reading First schools. One group included Reading First Eligible schools, meaning that they met low achievement and socio-economically disadvantaged (SED) criteria for Reading First but had not yet applied for or been accepted into Reading First. The second group consisted of demographically matched non-Reading First schools, that also were not eligible to apply for Reading First funding, selected through cluster methodology and randomly selected from a list of schools that most closely matched the Reading First schools for SED and English Learner (EL) student population percentages. The Year 3 Report showed that Reading First Eligible schools were not demographically representative of the schools that had already entered the program, specifically, that they had smaller numbers of EL students, a factor that could significantly skew results. The Year 3 report used the demographically matched schools as a comparison group. Though these schools were similar to the Reading First schools with regard to English Learners, the Reading First Eligible school group still had much lower percentages of EL students than did the Reading First schools. The Year 4 report, then, provided a rationale and detailed methodology for creating a statistical control group using statistical methods to hold constant the effect of demographic characteristics and implementation factors. The statistical control group is interpreted as a hypothetical set of schools that would have the same starting point in achievement and implementation, demographic characteristics and years in the program as if they were not in the Reading First program at all. Their performance on various Reading First indicators, then, is hypothetical, based on predictive regression models.

Besides the difficulty of finding similar non-participating schools for comparison, an added problem was the widespread effect of Reading Firs-like elements in non-Reading First schools. The Reading First program coincided with state and district reading initiatives, making it impossible to find similar schools

that did not implement elements such as the use of state-adopted curricula, participation in professional development, use of reading coaches, and ongoing assessment to guide instruction.

Despite these difficulties, the evaluation of the Reading First program over time has demonstrated significantly higher achievement results for schools in the less-than-perfect comparison groups and the hypothetical statistical control group.

Reading First showed consistent achievement gains for English Learners.

The number of ELs in California's K-3 classrooms has risen steadily during the past decade. Due to the pervasive difficulties in learning to read for ELs, there has been great concern about how to effectively teach reading to EL students in California. In each annual Reading First evaluation report, achievement data were disaggregated by subgroups to determine the effectiveness of Reading First for English learners, socio-economically disadvantaged students and students with disabilities. The Year 6 report presented achievement data for the EL subgroup over time. Despite complications, such as the inability to determine what proportion of students may be reclassified as English proficient in a given year, this report showed significant and steady gains over time in Reading First schools compared to non-Reading First schools. Fidelity of implementation seemed to be an important factor: ELs in high implementing schools experienced significantly higher growth than ELs in low implementing schools and the student population as a whole.

In the Year 5 report, teachers, coaches and principals were invited to respond to an open-ended question, "In what ways has your school's participation in Reading First impacted the learning of English learners in your school? Explain your response." Using qualitative research methodology, the responses were analyzed to capture the nature of participants' perceptions as well as the strength. The strongest categories of responses showed that instruction and reading outcomes improved for ELs. This comment from a coach captures the theme, "We have found that English learners benefit from research-based instructional practice. ELs are improving in their academic achievement because teachers fully implement the instructional program and provide opportunities for scaffolds." A teacher commented, "Our English learners have improved and made great progress in reading as a result of the excellent reading program."

## Lessons Learned Regarding Implementation

To examine the process of implementation, we revisit the Reading First survey and include item analysis from Years 3 through 7 survey results to compile lessons learned over the course of implementation. Additionally, we summarize findings from topical inquiries included in previous reports using narrative responses to open-ended survey questions. These open-ended questions focused on the importance of coaching and the assessments used in Reading First schools, the impact of Reading First on English

learners and Spanish-language instruction, how special education teachers and students were included in Reading First, and the sustainability of the program. We have organized this section on implementation to discuss lessons learned regarding the sustainability of the program, fidelity of implementation, professional development, curriculum, coaching, and assessment.

The Reading First program has created sustainable features of reading/language arts instruction.

*"We would want to keep as much of the program in place as possible because a comprehensive language arts program that is well articulated and universally implemented is needed to enable all students to achieve." (Reading Coach reflection)*

As the Reading First program comes to an end, what practices or elements to sustain becomes an important issue The Years 6 and 7 surveys included a question regarding sustainability. Table 4.1 displays the results. Note that percentages do not total 100% because respondents were able to select multiple elements. In this table, we see similarities across respondent groups regarding priorities for sustainability: a protected Reading/Language Arts time block, collaborative planning, professional development, assessment and data analysis, a reading coach, and small group or universal access time. The percentages are relatively lower overall for teachers than for coaches and principals for all elements.

**Table 4.1: Percentages of Teachers', Coaches', and Principals' Responses Regarding Sustainability of Elements of Reading First From Year 6 and Year 7 Reports**

| **Item I5 for Teachers; I7 for Coaches and Principals: If elements of your Reading First program had to be cut for funding or other reasons, which elements of the program would you most strongly support keeping in place? Check all that apply.** | **Teachers %** | | **Coaches %** | | **Principals %** | |
|---|---|---|---|---|---|---|
| | **2007-2008** | **2008-2009** | **2007-2008** | **2008-2009** | **2007-2008** | **2008-2009** |
| a. Structured Teacher Planning Time | 53 | 53 | 69 | 71 | 68 | 69 |
| b. Reading/Language Arts Time Block | 65 | 66 | 85 | 83 | 78 | 80 |
| c. Collaboration/Lesson Studies | 42 | 42 | 73 | 72 | 70 | 70 |
| d. Substitute Days/Release Time | 33 | 31 | 45 | 38 | 49 | 42 |
| e. Curriculum/Materials for waivered classrooms | 19 | 20 | 17 | 19 | 16 | 15 |
| f. Pacing Plan or Guide | 44 | 44 | 68 | 69 | 66 | 65 |
| g. Instructional Strategies | 56 | 54 | 78 | 80 | 71 | 73 |
| h. Professional Development | 48 | 42 | 80 | 77 | 76 | 71 |
| i. English learner handbook or support guide | 35 | 32 | 50 | 51 | 54 | 47 |
| j. Assessment and Data Analysis | 45 | 45 | 82 | 82 | 78 | 74 |
| k. Your school's reading coach | 49 | 44 | 83 | 80 | 81 | 77 |
| l. Curriculum/Materials, for non-waivered classrooms | 24 | 24 | 30 | 29 | 22 | 23 |
| m. Supplementary Materials | 43 | 42 | 26 | 28 | 26 | 28 |
| n. Small Group Instruction/Universal Access | 58 | 63 | 81 | 84 | 69 | 76 |

The Year 5 report included an open-ended question regarding sustainability. The findings of the Year 5 report were similar. Curriculum, coaching collaborative planning, assessment and professional development were the key components identified by participants as important to sustain.

Fidelity of Implementation is critical to achieving a positive impact on instructional practices and student achievement in reading.

*"Reading First has been a great part of professional development that has built teachers' practices through planning and collaboration. Grade level meetings, lesson study, and focused classroom observations have led to consistent and effective delivery of instruction in all grades (K-5$^{th}$)." (Reading Coach reflection)*

The assurances that districts had to comply with included provisions for ensuring full implementation of the Reading First guidelines and adopted reading/language arts curricula. Fidelity of implementation is defined as "the degree to which an intervention [or program] is implemented as planned" (Noell , Gresham &Gansle, 1993). Previous reports have investigated adherence to the program elements and found that high implementation was significantly related to higher student achievement. Previous reports have also included extensive analysis of fidelity of implementation with respect to various program elements. In summary, implementation of the Reading First elements increased over time and, overall, the program achieved a fairly high degree of implementation. Level of implementation was significantly correlated with level of achievement.

There were several aspects of the survey that showed how program elements supported implementation. Findings over time are included here.

Findings from the survey illustrate the importance of the required time allotment for reading/language arts instruction. The Reading First program and state reading/language arts framework require a minimum of 150 minutes per day of reading/language arts instruction in grades 1 – 3 and 60 minutes in kindergarten. Table 4.2 shows the amount of time reported by teachers spent in teaching their adopted curriculum from 2004 forward. The time allocation results are fairly stable over time. Comments from open-ended questions reinforced the importance of an uninterrupted time block, though teachers often felt a time pressure. They felt that they did not always have enough time, even in their protected time block, to complete the program every day.

**Table 4.2: Percentages of Teachers Responses Regarding Time Allocation**

| **Item F1: On average over the last four instructional weeks, how many minutes per day have you spent teaching the district's adopted reading/language arts program?** | **2004-05 %** | **2005-06 %** | **2006-07 %** | **2007-08 %** | **2008-09 %** |
|---|---|---|---|---|---|
| **Kindergarten Teachers** | | | | | |
| a. Less than 20 minutes | 0 | 0 | 0 | 0 | 0 |
| b. 20-39 minutes | 1 | 1 | 1 | 1 | 1 |
| c. 40-59 minutes | 4 | 3 | 3 | 2 | 4 |
| d. 60-79 minutes | 4 | 4 | 13 | 12 | 16 |
| e. 80-99 minutes | 6 | 8 | 21 | 19 | 22 |
| f. 100-119 minutes | 13 | 12 | 12 | 12 | 10 |
| g. 120-139 minutes | 13 | 13 | 21 | 21 | 18 |
| h. 140-159 minutes | 15 | 17 | 7 | 7 | 6 |
| i. 160-179 minutes | 19 | 20 | 5 | 5 | 3 |
| j. 180 minutes or more | 25 | 22 | 17 | 19 | 19 |
| **Grades 1-3 Teachers** | | | | | |
| a. Less than 20 minutes | 0 | 0 | 0 | 0 | 0 |
| b. 20-39 minutes | 0 | 0 | 0 | 0 | 1 |
| c. 40-59 minutes | 1 | 1 | 1 | 1 | 1 |
| d. 60-79 minutes | 3 | 2 | 3 | 3 | 4 |
| e. 80-99 minutes | 5 | 4 | 5 | 5 | 6 |
| f. 100-119 minutes | 5 | 5 | 5 | 5 | 6 |
| g. 120-139 minutes | 12 | 12 | 18 | 19 | 20 |
| h. 140-159 minutes | 19 | 18 | 19 | 19 | 19 |
| i. 160-179 minutes | 19 | 20 | 11 | 11 | 9 |
| j. 180 minutes or more | 36 | 36 | 36 | 35 | 34 |

Note: This table excludes teachers of split grade combination classes and teachers who did not specify a grade.

Numerous data sources over time have pointed to the importance of having regularly scheduled and sufficient opportunities for teachers to plan lessons and review data collaboratively and with their reading coaches. Reading First required schools to engage in collaborative planning meetings twice monthly with a focus on analyzing student data, understanding the curriculum materials, improving instructional strategies, and assisting struggling readers. Table 4.3 presents findings from a question asked of teachers (Question D2), coaches (Question E2) and principals (Question E2) regarding how often the school provided time for teachers to plan collaboratively. Data are displayed for Years 3 – 7. Findings for each respondent group were consistent across years, though fewer teachers reported two or more times per month than coaches and principals. A majority of survey respondents reported two or more meetings per month. However, despite the view that these meetings were important, a relatively high number of teachers over time reported fewer than two planning meetings per month. This may be an indication that it is difficult for schools to allocate teacher meeting time for planning due to competing demands on limited out-of-classroom time for teachers.

**Table 4.3: Percentages of Teachers, Coaches, and Principals Regarding Collaborative Planning Time**

| **Item D2, Teachers, Item E2, Coaches/Principals: How often does the school leadership provide time for teachers to plan collaboratively?** | **2004-2005 %** | **2005-2006 %** | **2006-2007 %** | **2007-2008 %** | **2008-2009 %** |
|---|---|---|---|---|---|
| **Teachers** | | | | | |
| a. Hardly ever | 20 | 17 | 18 | 18 | 14 |
| b. Monthly | 28 | 28 | 28 | 25 | 23 |
| c. Twice monthly | 22 | 22 | 22 | 22 | 28 |
| d. Weekly | 28 | 30 | 31 | 33 | 33 |
| e. Daily | 1 | 1 | 1 | 1 | 1 |
| **Coaches** | | | | | |
| a. Hardly ever | 8 | 5 | 4 | 4 | 2 |
| b. Monthly | 23 | 20 | 21 | 19 | 15 |
| c. Twice monthly | 34 | 35 | 36 | 34 | 38 |
| d. Weekly | 34 | 36 | 37 | 42 | 42 |
| e. Daily | 1 | 1 | 0 | 1 | 1 |
| **Principals** | | | | | |
| a. Hardly ever | 1 | 1 | 1 | 1 | 0 |
| b. Monthly | 16 | 16 | 15 | 13 | 13 |
| c. Twice monthly | 37 | 34 | 36 | 35 | 38 |
| d. Weekly | 44 | 45 | 45 | 48 | 45 |
| e. Daily | 1 | 1 | 1 | 2 | 2 |

The Reading First program has required districts to develop pacing plans or guides for ensuring consistency across classrooms in terms of content covered and to ensure that students move through the grade-level standards and aligned curriculum. Pacing plans provide guidelines for what lessons should be taught in time periods spaced throughout an academic year. If teachers adhere to the pacing guidelines, they should cover the entire year's curriculum. The survey asked participants whether they had a pacing schedule and how closely they adhered to it. Table 4.4 presents results from teachers (Question D1), coaches (E1) and principals (E1) on this question. Nearly all participants reported that they have a pacing schedule. A smaller proportion of teachers than coaches and principals reported that their pacing schedule provides detailed guidance about what lessons to teach on a daily or weekly basis. These percentages are consistent with previous reports.

**Table 4.4: Percentages of Teachers, Coaches, and Principals Regarding Pacing Plans**

| **Year 7 Survey, Item D1, Teachers; Item E1, Coaches/ Principals: Does your school have a pacing schedule?** | **Teachers %** | **Coaches %** | **Principals %** |
|---|---|---|---|
| a. My school does not have a pacing schedule | 2 | 0 | 0 |
| b. My school has a pacing schedule based only on the assessment schedule | 30 | 21 | 16 |
| c. My school has a pacing schedule that identifies lessons on a daily or weekly schedule and when to give assessments | 68 | 78 | 82 |

Additional information about pacing schedules was provided by teachers (Question F4). Table 4.5 shows that teachers who reported precise adherence to the pacing schedule significantly increased from Year 3 to Year 4, and then increased slightly for Years 5 and 6. This suggests that, as teachers developed expertise and depth of knowledge through long-term participation in the Reading First program, they were more able to fully implement the pacing schedule.

**Table 4.5: Teacher Survey Results for Adherence to Pacing Schedule**

| Item F4: To what degree do you follow your school's pacing schedule for reading/language arts? | 2004-2005 % | 2005-2006 % | 2006-2007 % | 2007-2008 % | 2008-2009 % |
|---|---|---|---|---|---|
| a. Our school does not have a pacing schedule | 2 | 1 | 0 | 0 | 1 |
| b. I do not follow the existing pacing schedule | 2 | 1 | 1 | 0 | 1 |
| c. I keep in mind where I want to be and aim for that | 8 | 6 | 5 | 4 | 5 |
| d. I follow the pacing schedule approximately | 38 | 27 | 24 | 25 | 23 |
| e. I follow the pacing schedule quite precisely | 49 | 64 | 69 | 69 | 70 |

Improving the capacity of school leadership to support an effective reading program has been an element of maintaining fidelity. Teachers responded to two questions regarding the role of the school administrator in program implementation, displayed in Table 4.6. Question D11 indicates that over 80% of teachers have consistently reported adequate or more than adequate support from their principals. Question D12, asked only from Years 4-6, indicates that 88-90% of teachers reported that they were required by the school principal to fully implement their adopted reading/language arts program.

**Table 4.6: Teacher Survey Results Regarding School Leadership**

| **D11: In general, what level of support are you getting from your principal related to your teaching of the adopted reading/language arts program?** | **2004-2005 %** | **2005-2006 %** | **2006-2007 %** | **2007-2008 %** | **2008-2009 %** |
|---|---|---|---|---|---|
| a. Little or no support | 19 | 18 | 17 | 17 | 16 |
| b. Adequate support | 55 | 53 | 55 | 55 | 55 |
| c. More than adequate support | 24 | 27 | 27 | 27 | 28 |
| D12: Does your school leadership require K-3 teachers to fully implement the adopted reading/language arts program? | | | | | |
| a. Full implementation is required | | 90 | 89 | 88 | 83 |
| b. Some variation from full implementation is permitted | | 8 | 10 | 11 | 15 |

The Reading First program brought an increased focus on assessment to California's schools. The curriculum-embedded assessments used in Reading First were used on an ongoing basis to monitor student learning and guide the collaborative data analysis and planning meetings. With this requirement of Reading First, common reading assessments have become part of the fabric of reading/language arts instruction. The assessments facilitated professional dialogue and allowed teachers to determine students' specific learning needs, monitor progress and adjust instruction.

The Year 5 survey included an open-ended question focusing on the assessments, *"What is your opinion of the 6-8 Weeks Skills Assessments (from SCOE)? How are they helpful? How could they be improved?"* The qualitative analysis yielded several important findings as well as suggestions for improving the system. Respondents indicated that their schools used the assessments to monitor student progress and guide instruction, in line with the purpose for which they were designed. Additionally, the assessments were useful for pinpointing specific needs of students and helping teachers to form small groups for supplemental instruction. Some of the suggestions for improvement focused on getting a better fit in aligning with the skills taught at particular time points and with the state standards, improving the test format so that students could better understand it, modifying the pacing or timing of the assessments and examining the difficulty level of the tests. This comment from a teacher illustrates the importance of the data for instruction, *"I find them very useful when I plan. The data help me guide my instruction. It also serves as a tool to show the parents the various areas where their children are doing well or need more help."*

Professional Development was extensive and effective in Reading First.

*"Training is crucial to ensuring that our instruction does not become stagnant but that it continues to evolve and grow." (Teacher reflection)*

The Reading First program has provided extensive professional development for teachers, coaches and principals. Teachers acquired basic knowledge about their adopted curriculum in their first year in the program and then in subsequent years, teachers received advanced training either through the Reading Implementation Centers or their district. Administrators and coaches also received extensive professional development over their years in the program. Table 4.7 below shows teachers' responses over time to a question regarding the quality of professional development received. During the last two years of the program, fewer teachers attended professional development, explaining the increase in the numbers who did not respond. Teacher perceptions over time indicated that the quality of professional development was strong, with the majority of teachers reporting that it prepared them "adequately" or "very well."

**Table 4.7: Teacher Survey Results for Quality of Professional Development**

| **Item B4. How well did the reading Professional Development Institute training prepare you to teach the district's adopted reading/language arts program?** | **2004-2005 %** | **2005-2006 %** | **2006-2007 %** | **2007-2008 %** | **2008-2009 %** |
|---|---|---|---|---|---|
| a. Not applicable | 3 | 2 | 2 | 3 | 3 |
| b. It did not prepare me well | 12 | 10 | 9 | 7 | 7 |
| c. It prepared me adequately | 58 | 55 | 52 | 43 | 41 |
| d. It prepared me very well | 15 | 16 | 16 | 15 | 14 |
| Did not respond to this item | 12 | 17 | 21 | 32 | 35 |

The open-ended questions from the Years 5 and 6 reports provide additional information about the quality and appropriateness of the Reading First sponsored professional development. Participants were asked, *"In your opinion, what aspects of Reading First-funded coaching do you view as most valuable and why?"* The responses from teachers, coaches and principals illustrated the importance of coaches in the professional development process. One strong finding was the importance of the demonstration of specific lessons using the reading curricula. In the lesson demonstration process, coaches and teachers discuss the lesson elements, followed by the coach demonstration, then perhaps side-by-side teaching or an observation-feedback cycle. Respondents felt strongly that this was a very effective method of

providing professional development. According to one teacher, *"Seeing the actual lesson plan presented helps me visualize what I need to do and helps me understand how the lesson should be carried out."*

Professional development focusing specifically on effective, research-based strategies for teaching English learners was another highlight of the Reading First professional development. Teachers, coaches and principals expressed that it was helpful in delivering the instruction and that it linked directly with improved achievement outcomes for EL students. *"It has provided teachers with much-needed professional development opportunities that have increased their knowledge of research-based practices that benefit all students, including English language learners (a coach)."*

Professional development did not always reach special education teachers or benefit students with disabilities. As reported in the Year 6 report, almost half of special education teachers in Reading First schools reported that they had participated in very little professional development in reading/language arts; 46% reported either zero hours or 1 to 5 hours. On the survey, only 52% of special education teachers reported that the Reading First professional development prepared them to teach the adopted program "adequately" or "very well." Only 37% participated in the full 80 hours of follow-up. In the Year 6 report, respondents were invited to write in responses to this question, *"What impact, if any, has your school's involvement in Reading First had on special education teachers and students with disabilities?"* The majority of responses fell into the "Don't Know/ Not Sure" of the impact or "No Impact" categories. Yet, a substantial number also indicated a positive impact. One special education teacher wrote, *"I feel that my students have really benefited from Reading First. I feel that they are truly prepared for the next grade level,"* and another wrote, *"I have been given many new strategies and ideas. I enjoy having a reading coach be there for me."* It appears that teachers who did participate found some benefit.

<u>The use of research-based curricula and materials was a key feature of Reading First.</u>

*"As long as we have the same reading program, I would implement it the same way, with or without Reading First." (Teacher reflection)*

*"[I would want to keep] the reading program and its tools. They have proven to keep us all on the same page." (Teacher reflection)*

The implementation of a state-adopted Reading/Language Arts curriculum was one of the assurances that districts had to adopt to participate. A major goal of the Reading First program was to achieve full implementation of the curriculum through professional development and ongoing support and monitoring. A large section of the survey focused on whether participants had received appropriate materials, used them and found them to be effective. These questions were very specific and asked questions about all the components of the state adopted curricula, including the Spanish language materials. For details, the

reader is referred to Section C of the teacher survey, and Section D of the coach and principal surveys in the appendix. One item in particular provides insight into trends in curriculum use over time. Item F3 from the teacher survey, asks, "What percentage of your total reading/language arts instruction relies on materials from your district's adopted program?" Table 4.8 displays the results of this question asked in Years 3 – 7. Consistently over time, approximately 80% of teachers have reported that 80% to 100% of their instruction relies on their adopted curriculum materials.

**Table 4.8: Teacher Survey Results for Curriculum Use**

| **Item F3: What percentage of your total reading/language arts instruction relies on materials from your district's adopted program?** | **2004-2005 %** | **2005-2006 %** | **2006-2007 %** | **2007-2008 %** | **2008-2009 %** |
|---|---|---|---|---|---|
| a. 0% - 19% | 0 | 0 | 0 | 0 | 0 |
| b. 20% - 39% | 1 | 1 | 1 | 1 | 1 |
| c. 40% - 59% | 5 | 4 | 4 | 4 | 5 |
| d. 60% - 79% | 15 | 14 | 13 | 15 | 12 |
| e. 80% - 100% | 77 | 79 | 80 | 78 | 80 |

Note: Rounding of percentages and items left blank on individual surveys result in less than 100% reported here.

Teachers, coaches and principals were asked to rate the overall effectiveness of their district's adopted reading/language arts program in item I1 in each year of the survey. Results from this item over time are included in Table 4.9 below. Principals' responses were the most positive and teachers were the least, yet most responses fell in the "good" or "excellent" categories.

**Table 4.9: Percentages of Teachers', Coaches', and Principals' Responses Regarding Curriculum Effectiveness**

| Item I1: Overall, how would you rate the effectiveness of your district's adopted reading/language arts program in your school? | 2004-2005 % | 2005-2006 % | 2006-2007 % | 2007-2008 % | 2008-2009 % |
|---|---|---|---|---|---|
| **Teachers** | | | | | |
| a. Poor | 3 | 3 | 3 | 3 | 2 |
| b. Fair | 20 | 18 | 19 | 20 | 19 |
| c. Good | 55 | 55 | 56 | 56 | 57 |
| d. Excellent | 20 | 21 | 21 | 20 | 21 |
| **Coaches** | | | | | |
| a. Poor | 0 | 1 | 0 | 1 | 0 |
| b. Fair | 8 | 11 | 13 | 13 | 11 |
| c. Good | 58 | 54 | 59 | 60 | 60 |
| d. Excellent | 31 | 30 | 26 | 25 | 28 |
| **Principals** | | | | | |
| a. Poor | 0 | 1 | 0 | 0 | 1 |
| b. Fair | 6 | 6 | 8 | 9 | 8 |
| c. Good | 55 | 58 | 61 | 59 | 59 |
| d. Excellent | 36 | 32 | 29 | 29 | 31 |

Coaching has become well integrated into Reading First schools and has supported implementation.

*"A Reading First coach is the hub of the wheel- supporting, guiding, and coordinating the school's efforts toward full implementation and data-driven instruction (coach self-reflection)"*

The Reading First program has played a significant role in establishing reading (or literacy) coaches in districts throughout the state. Coaches are out-of-the-classroom teachers with expertise in research-based instructional strategies, state-adopted reading curricula, the reading developmental process and strategies for supporting teachers. The Year 5 report included a separate chapter examining the role of reading

coaches in the Reading First program. Here, we highlight a few key survey items related to the use of reading coaches over time. Then, we summarize key findings from the open-ended question focusing on coaching in Year 5, *"In your opinion, what aspects of the Reading First-funded coaching do you view as most valuable or beneficial and why?"* (The coaches' version of the question was worded slightly differently). The reader is referred to the Year 5 report for an in-depth discussion of the Reading First coaching model (www.eddata.com). The survey data and additional Year 5 open-ended question focusing on coaching indicated that coaches are highly valued and have become integrated into the fabric of reading/language arts instruction throughout the state.

The coaching model developed with support from Reading First funding is an example of how the state has built capacity as a result of the program. Through Reading First, the state has built a highly qualified coaching force and many of the coaches who served in the Reading First program developed sufficient expertise to serve as instructors for professional development, acquired Reading Specialist certification, and moved on to administrative roles.

Table 4.10 shows survey responses regarding the level of support provided by coaches to teachers regarding the implementation of the curriculum. Results were consistent across years.

**Table 4.10: Percentages of Teachers', Coaches', and Principals' Responses Regarding the Coach as a Resource**

| **Teacher Survey (E2): How helpful is your coach in answering questions about how to teach the program?** | **2004-2005 %** | **2005-2006 %** | **2006-2007 %** | **2007-2008 %** | **2008-2009 %** |
|---|---|---|---|---|---|
| Not Applicable. My school does not have a reading coach (only available for Year 3) | 0 | | | | |
| a. The coach often doesn't know more than I do about how to teach the program | 8 | 7 | 7 | 8 | 8 |
| b. The coach gives general answers to questions | 24 | 24 | 24 | 24 | 24 |
| c. The coach gives specific, detailed answers that teachers can use | 64 | 66 | 66 | 65 | 64 |

| **Coach Survey (F5): How helpful do you feel you are in answering teacher questions about how to teach the program?** | **2004-2005 %** | **2005-2006 %** | **2006-2007 %** | **2007-2008 %** | **2008-2009 %** |
|---|---|---|---|---|---|
| a. I often don't know more than the teachers about how to teach the program | 0 | 1 | 1 | 0 | 1 |
| b. I am able to give general answers to questions | 15 | 10 | 12 | 9 | 10 |
| c. I give specific, detailed answers that teachers can use | 83 | 87 | 86 | 90 | 88 |
| **Principal Survey (F5): How helpful is your coach in answering questions about how to teach the program?** | | | | | |
| Not Applicable. My school does not have a reading coach (only available for Year 3) | 1 | | | | |
| a. The coach often doesn't know more than I do about how to teach the program | 1 | 0 | 0 | 0 | 0 |
| b. The coach gives general answers to questions | 10 | 7 | 8 | 7 | 8 |
| c. The coach gives specific, detailed answers that teachers can use | 86 | 88 | 87 | 88 | 85 |

A reading coach typically facilitates grade-level teacher meetings focusing on data analysis, instructional strategies, struggling students, and overall program implementation, a role that may have fallen on the site administrator in the past. Table 4.11 shows results from Years 3 through 7 for a question that asked about the coach's role in these meetings. Findings were consistent across years. Teachers reported at a higher rate than coaches and principals that the coach was not involved in these meetings, but overall coaches seemed to be facilitating the meetings and maintaining focus on instructional needs.

**Table 4.11: Percentages of Teachers', Coaches', and Principals' Responses Regarding the Coach as a Facilitator**

| **Teachers: Does the coach facilitate regular grade-level teacher meetings related to your district's adopted reading/language arts program?** | **2004-2005 %** | **2005-2006 %** | **2006-2007 %** | **2007-2008 %** | **2008-2009 %** |
|---|---|---|---|---|---|
| Not applicable, my school does not have a reading coach. (only available for year 3) | 1 | | | | |
| a. The coach is not involved with the grade-level meetings | 23 | 21 | 23 | 25 | 35 |
| b. The coach helps facilitate the meetings regularly | 45 | 45 | 46 | 44 | 37 |
| c. In addition to facilitating meetings, the coach keeps them focused instructional needs of teachers | 28 | 31 | 29 | 28 | 23 |
| **Coaches: Do you facilitate regular grade-level teacher meetings related to your district's adopted reading/language arts program?** | | | | | |
| a. I am not involved with the grade-level meetings | 12 | 11 | 12 | 15 | 20 |
| b. I facilitate the meetings regularly | 38 | 35 | 37 | 34 | 36 |
| c. In addition to facilitating meetings, I keep them focused on instructional needs of teachers | 46 | 49 | 49 | 48 | 41 |
| **Principal: Does the coach facilitate regular grade-level teacher meetings related to your district's adopted reading/language arts program?** | | | | | |
| Not applicable, our school does not have a reading coach. (only available for year 3) | 1 | | | | |
| a. The coach is not involved with the grade-level meetings | 8 | 7 | 7 | 7 | 11 |
| b. The coach helps facilitate the meetings regularly | 34 | 34 | 33 | 33 | 31 |
| c. In addition to facilitating meetings, the coach keeps the focus on instructional needs of teachers | 55 | 53 | 56 | 56 | 50 |

From the open-ended question, we learned that the coaches provided extensive support to teachers and their implementation of the curriculum. Demonstration lessons were highly valued as was the importance of having a resource to turn to with questions or programmatic needs. A teacher commented, *"The most beneficial aspect of having a coach is that we have someone to support and guide, and help us with anything we need to achieve our goals in teaching reading* (teacher comment)." Along the same line, a principal stated, *"Coaches have been a great asset to our school and provided invaluable information and assistance. The assistance to teachers is immediate and hands-on. It is a win-win situation for everyone, especially our students."*

## Conclusions

Over seven years of implementation, California's Reading First initiative has transformed reading instruction in thousands of classrooms. Reading First funding was focused on improving reading outcomes for students in socio-economically disadvantaged areas and in schools with chronic low achievement. The cumulative reports of reading outcomes in California's Reading First schools show that Reading First has accomplished that goal. Reading achievement has risen steadily in Reading First schools according to various achievement metrics used in the evaluation reports in comparison to comparison groups and a statistical control group. There has been a steady migration of students out of the Below Basic and Far Below Basic achievement groups and into the Proficient and Advanced groups in Reading First schools. These findings also hold true for the subgroup of English learners and beyond the K-3 grades. This evaluation has yielded a school level Reading First Achievement Index, or RFAI. Over time, the RFAI has steadily risen and has proved useful as a measure of significant progress for schools participating in Reading First. Though California has not yet achieved the goal of ensuring that every student reaches proficiency by the end of grade 3, the results of this evaluation show a substantial step toward that goal. It is our recommendation that the state make every effort to sustain this trend through continued vigilance in improving reading instruction in the early grades.

It is impossible to understand the scope of the impact of Reading First without examining implementation in relation to achievement. This seven-year evaluation process has resulted in the ability to examine implementation at macro and micro levels. The Reading First survey data has yielded a Reading First Implementation Index, a measure of each school's level of implementation. A consistent finding of the past five years of reporting has shown a strong and positive correlation between implementation and achievement. There is no doubt that achievement rises when implementation of the program is strong.

This chapter also examined specific elements of implementation. The findings demonstrate that the Reading First program has led to the development of a well-integrated structure and process of providing reading/language arts instruction in California. The program elements outlined in the Reading First

assurances are integral parts to a whole that is more than the sum of its parts. The use of state-adopted curricula, professional development, coaching, ongoing data analysis and collaboration, leadership support, protected time blocks, and other program elements together form an integrated reading program that has had a strong impact on reading achievement in the state. It is important for state leadership and policy makers to consider the importance of sustaining these program elements as interconnected and essential ingredients of an effective reading/language arts program.

In July 2008, the California Reading First program received its final federal grant. Have the findings of the California Reading First Evaluation therefore ceased to be relevant to public policy discussions? We do not believe so. While this evaluation is strictly applicable only to the domain of California schools participating in Reading First, we believe that the findings accumulated over seven years of research generalize beyond that domain. Reading First is characterized by key components that include the use of research-based curricula, ongoing targeted professional development, ongoing data analysis and collaboration, coaching support, leadership support, protected time blocks, and strong program coherence. Given the strong effects in Reading First schools, we believe the following scenarios are plausible for schools that adopt strategies in line with the Reading First paradigm:

- Non-Reading First elementary schools (i.e., those that were not eligible for program funds) would on average move a substantially higher percentage of students out of the Below and Far Below Basic performance levels.
- Teachers who apply the principles of Reading First in the upper grades (grades 4 – 12) would on average show substantially improved scores for their low-performing students.
- A substantially larger percentage of English Learners in all grades in classrooms that adopt Reading First principles would score “Proficient” on the CSTs and would therefore be eligible for reclassification in many districts.
- It is possible that Reading First assurances would be similarly effective if adapted to elementary school mathematics instruction and to other skill-based and habit-based content areas.

Because this evaluation has been focused on low-performing schools and low-performing students, we cannot say whether the Reading First “paradigm” would be as effective with Proficient and Advanced students in high-performing schools. It is possible that other programs would be more effective. Nonetheless, the fact remains that the elements that made Reading First a successful program are largely generic and can be applied to other grades, to other content areas, and to other student populations. For that reason, we believe that Reading First-like program elements deserve serious consideration in State and Federal educational policy discussions, regardless of domain.

## References

Gamse, B.C., Bloom, H.S., Kemple, J.J., Jacob, R.T., (2008). Reading First Impact Study: Interim Report (NCEE 20084016). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Moss, M., Fountain, A. R., Boulay, B., Horst, M., Rodger, C. & Brown Lyons, M. (2008). Reading First Implementation Evaluation Final Report. Cambridge, MA: Abt Associates. Available at: http://www.ed.gov/rschstat/eval/other/readingfirst-final/index.html.

Noell, G., Gresham, F., & Gansle, K. (2002, March). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education*, *11*(1), 51-67.