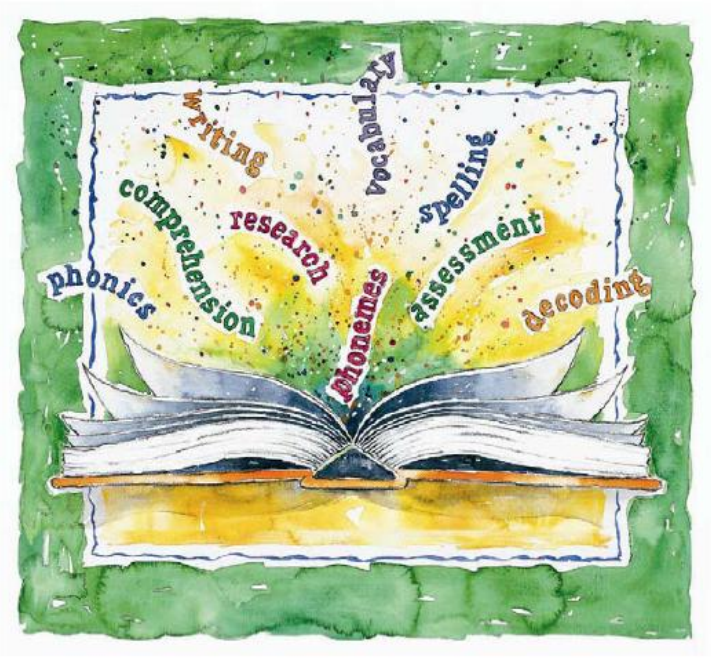
The California Reading First Year 5 Evaluation Report

January 2008



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

For five years California has been participating in the Reading First program, a federal initiative aimed at improving reading instruction in the United States. This report provides an evaluation of Reading First implementation and student reading achievement in California for those five years. The key findings are summarized below, and the remainder of the report contains the detailed analysis to support these conclusions.

Finding #1: Reading First is effective. After controlling for school demographic characteristics, Reading First implementation is a statistically significant predictor of achievement on all achievement metrics, especially those associated with grades 2 and 4. The more faithfully the program is applied, the greater the effect on achievement.

Finding #2: Growth has been significant. The Reading First Achievement Index (RFAI), a composite of K-3 achievement metrics for Reading First schools, has risen an average of 3.4 points per year, equivalent to 17 points over 5 years.

Finding #3: Reading First schools out-perform the control group. Reading First schools out-perform a statistical control group by 1.6 points per year on the RFAI, equivalent to an 8-point advantage over five years.

Finding #4: Reading First schools out-perform non-Reading First schools. While non-Reading First schools have also shown substantial growth since 2002, it is consistently and significantly less than the growth of Reading First schools, similar to the growth of the statistical control group.

Finding #5: The Reading First effect generalizes across students. Reading First effects generalize to all performance levels of the Reading First student population. On the California Standards Test (CST) metrics, the migration of students into "Proficient and Above" is matched by a comparable migration of students out of "Below Basic and Far Below Basic." These migrations are confirmed by average student CST scale score gains, on the order of 20 scale score points over a 5-year period. The rising average scale score shows that the student population, on average, experiences gains.

Finding #6: Reading First improves grade 4 performance. These findings are replicated in grade 4, even though Reading First is a K-3 program. In grade 4, Reading First schools grew 4.1 CST scale score points per year (20.5 scale score points over 5 years), versus 2.4 scale score points per year (12.0 points over 5 years) for the control group, a difference of 8.5 scale score points. Thus the program shows evidence of a sustained effect of Reading First. This is a new finding, not available in previous evaluation reports.

Finding #7: Implementation has been adequate. Most schools in the Reading First program are implementing the program adequately. The average level of implementation has risen throughout the duration of the Reading First program. The average Reading First Implementation Index (RFII) across all schools was 39 in 2006 and 2007, compared to 36 in 2004 and 2005.

Finding #8: Coaches are viewed as essential. Reading or literacy coaches are an integral part of the Reading First program in California. When participants were asked which elements of the Reading First program they would keep if funding were discontinued, coaches and the curricular materials were cited most frequently.

Finding #9: The Reading First effect generalizes to English learners. 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.

Finding #10: English learners in non-waivered classrooms out-perform English learners in waivered classrooms. In 2007 English learners in non-waivered classrooms scored 18 scale score points higher, on average, than English learners in waivered classrooms on the grade 2 CSTs . They scored 8 scale score points higher on the grade 3 CSTs.

Although there is ample room for improvement in program implementation and in the program itself, there are no significant negative findings to report.

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 Act, Reading First promotes the use of scientifically based reading practices in grades K-3. The initiative provides a significant amount of federal funding for improving reading instruction for large proportions of students experiencing academic difficulty and socio-economic disadvantage.

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

- Use of a state-adopted reading program

[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."

- Access to training programs authorized by state legislation and based on research-based reading instruction: Senate Bill (SB) 472 teacher and coach professional development and AB 75 principal professional development
- Access to assessment tools that test students' skills every six to eight weeks
- Hiring of reading coaches, expert teachers who support program implementation

Anecdotal evidence indicates that many non-Reading First schools have voluntarily been adopting some or all of these components over the same 5-year period, giving this evaluation study a relevance that extends beyond the Reading First population.

This report evaluates California's progress in implementation and achievement during the first five 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 four cohorts of Reading First schools and how they compare to non-Reading First schools.

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.

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 an analysis of perceptions of the relative importance of the various Reading First program elements.

Chapter 5 focuses on the use of reading coaches in Reading First schools, how they are perceived and the benefits they provide.

Chapter 6 provides achievement statistics and trend-lines showing the growth of the English learner subpopulation in Reading First schools since 2002.

Chapter 7 compares the grade 2 and grade 3 CST performance of English learners in waivered classrooms with that of English learners in non-waivered classrooms.

Attached 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 2005-2007 (Appendix F)

A Data Example from Grade 2

Our core findings are exemplified in Table ES.1.0 and Figures ES.1.0 – ES.1.2, representing the growth of Reading First schools on various grade 2 achievement metrics since their entry into the program five years ago. Similar charts for the other grades, school cohorts, and achievement metrics, as well as a summary table of gain scores for all Reading First schools, can be found in Chapter 2 of the main evaluation report. Because of their novelty and importance, we report the corresponding results for grade 4 in Table ES.2.0 and Figures ES 2.0 – ES 2.2, following the grade 2 trend-lines.

Table ES.1.0: CST Metric, Years in Program = 5, Grade = 2

Years in Program (YIP): 5 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	259	28	101	N/A	4,053
% Proficient and Above					
2002	15.4	14.8	14.8	15.4	37.8
2007	34.2	36.7	33.0	30.4	52.3
Change Since Starting Year	**18.9**[abc]	**22.0**[abc]	**18.2**[abc]	**15.0**	**14.5**
% Below and Far Below Basic					
2002	54.3	53.8	55.6	54.3	30.5
2007	36.7	33.6	38.7	41.1	23.0
Change Since Starting Year	**-17.6**[abc]	**-20.2**[abc]	**-16.9**[abc]	**-13.2**	**-7.6**
Mean Scale Score					
2002	299.8	299.5	298.5	299.8	333.4
2007	324.7	328.6	322.3	318.8	350.9
Change Since Starting Year	**25.0**[abc]	**29.0**[abc]	**23.8**[abc]	**19.0**	**17.5**

[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.
Note: Numbers reporting change since starting year were rounded and may not appear to be an exact difference between 2002 and 2007 figures.

Table ES.1.0 pertains only to those schools that have been in the program five years (Cohort 1) and it reports only their grade 2 CST scores. Referring to the "All Reading First Schools" column, we note the following. There were 259 Reading First schools in this first cohort, which had data for grade 2. On the

"% Proficient and Above" achievement metric, an average of 15.4 percent of students in these schools scored "Proficient and Above." By 2007, this percentage had increased so that 34.2 percent of students were scoring "Proficient and Above." The size of the gain was 34.2 minus 15.4, or 18.9 percentage points. Note that rounding accounts for any seeming discrepancies in computing the change from 2002 to 2007. The superscripts "abc" tell us this gain was "significantly" larger than the gains of the "statistical control group,"[2] the gain of the non-Reading First schools in California, and that the gain is significantly larger than zero. "Significant" means there is a 95% probability that a gain that large would not have occurred by chance.

Referring to the same column, we see the percent of students scoring "Below Basic or Far Below Basic" in 2002 and in 2007, and the subsequent change. This change is negative because it refers to students moving *out of* the bottom two performance level categories. Then we see the average student scale score (a test score ranging roughly from 250 to 450) in 2002 and in 2007, and the difference between them. Remember that these are students who were in grade 2 in 2002, and that there was another group of students who were in grade 2 in 2007. On average the 2007 students scored an average of 25 scale score points higher than their 2002 predecessors. For context, that is halfway between the "Basic" cut-point (300) and the "Proficient" cut-point (350).

The remaining columns report the same statistics for schools that have been classified as "high implementing" (using the RFII statistics, based on teacher, coach, and principal responses to the Reading First implementation survey) and "low implementing." The "Statistical Control Group" column reports the same statistics for a theoretical group of schools that are similar to the Reading First schools but not implementing the program. The last column reports the same statistics for the remaining 4,053 elementary schools in California that are not in the Reading First program. Note that this population has much higher starting scores than the Reading First schools. Therefore, for display in the trend-line charts, the starting points for "All Non-Reading First Schools" have been adjusted downward to coincide with the starting points of the Reading First groupings.

Comparing the bolded gain scores across the columns, we see that "All Reading First" schools grew faster than the "Statistical Control Group," that "High Implementation" schools grew faster than "Low Implementation" schools, and that they all grew faster than the "Non-Reading First" elementary schools in the rest of the state. All differences are statistically significant.

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

These findings demonstrating the efficacy of Reading First extend and confirm the findings from the Year 4 and Year 3 California Reading First Evaluation Reports.

The trend-lines corresponding to Table ES.1.0 are presented below.

Figure ES.1.0: CST % Proficient & Above, YIP = 5, Grade = 2

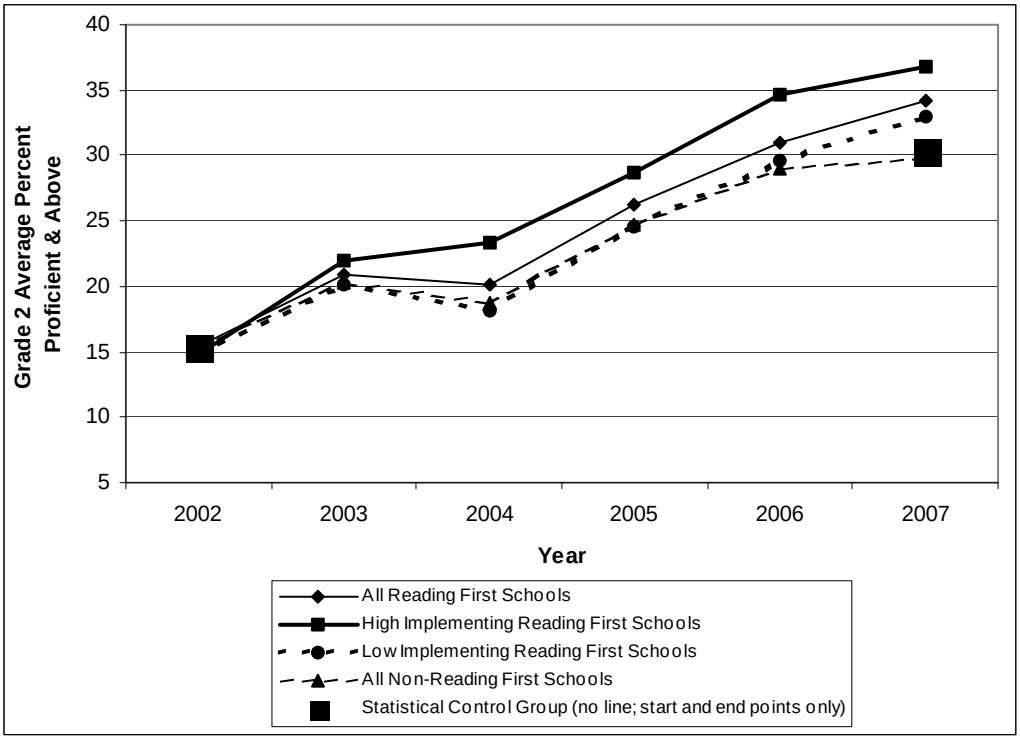


Figure ES.1.1: CST % Below Basic & Far Below Basic, YIP = 5, Grade = 2

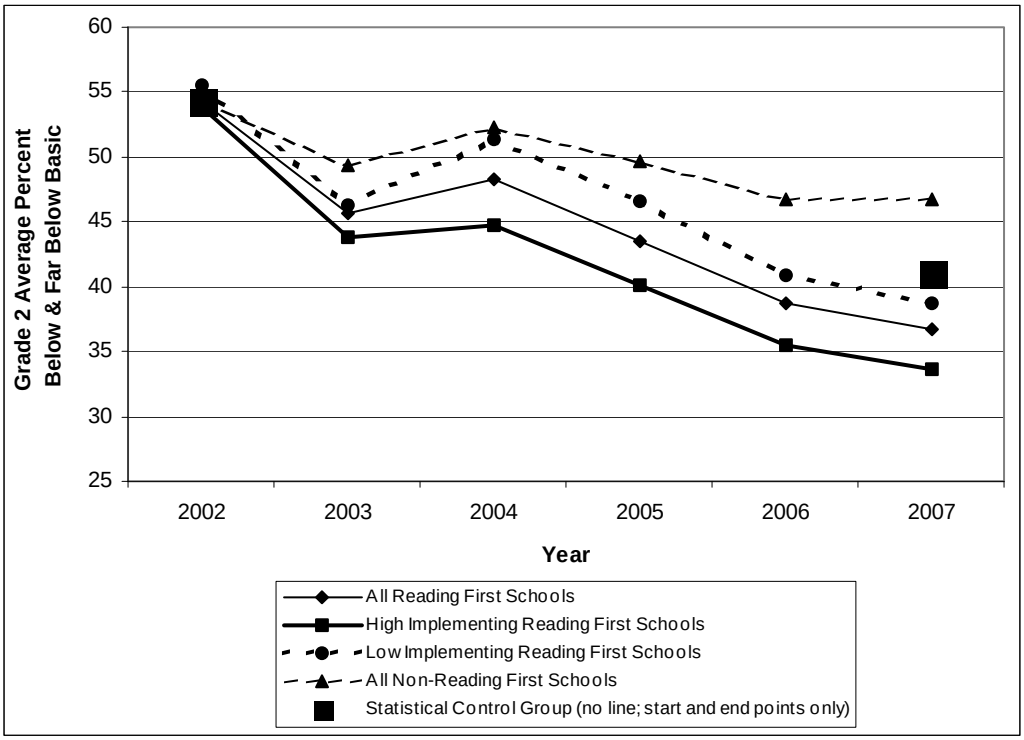
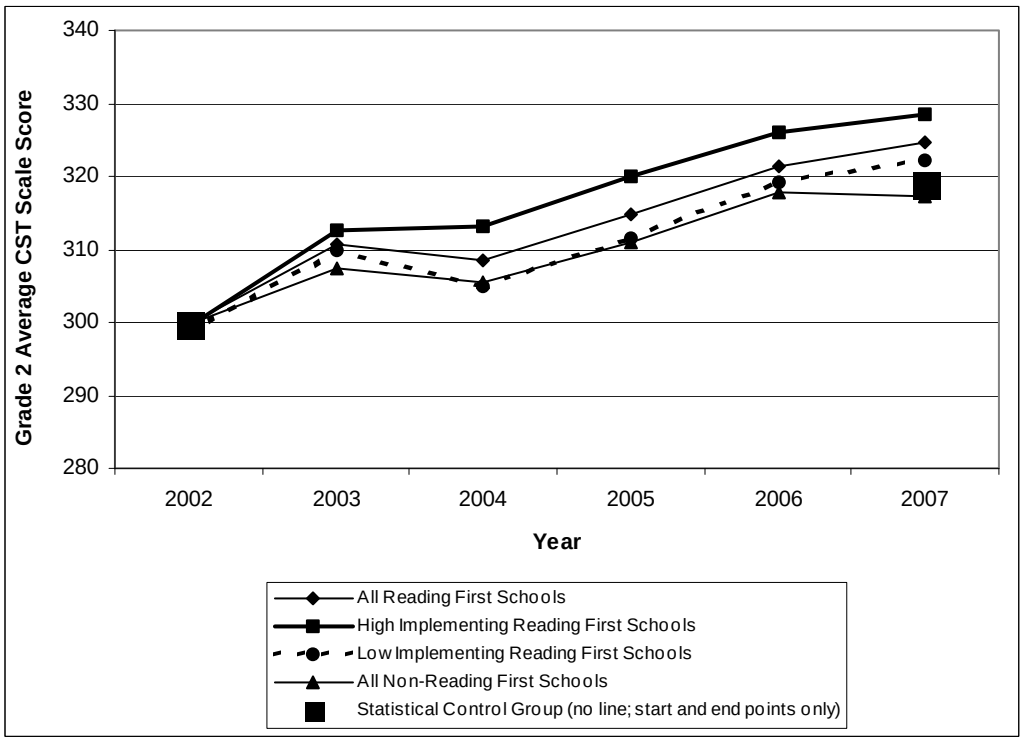


Figure ES.1.2: CST Mean Scale Score, YIP = 5, Grade = 2



A Data Example from Grade 4

This Year 5 Reading First Evaluation Report expands the scope of the evaluation by adding grade 4 CST performance as an achievement outcome. Because Reading First is administered only in grades K-3, the grade 4 results shed light on whether student exposure to Reading First in the earlier grades improves their ability to read in grades 4 and above. Table ES.2.0 and Figures ES.2.0 – ES.2.2 show that it does.

Table ES.2.0: CSTs, YIP = 5, Grade = 4

Years in Program (YIP): 5 Grade: 4	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	255	26	101	N/A	3,992
% Proficient and Above					
2002	15.2	14.2	15.9	15.2	42.1
2007	31.3	34.8	30.0	27.5	55.7
Change Since Starting Year	**16.1**[abc]	**20.6**[abc]	**14.1**[ac]	**12.3**	**13.6**
% Below and Far Below Basic					
2002	47.8	48.6	46.7	47.8	23.2
2007	32.2	28.4	33.4	34.9	16.9
Change Since Starting Year	**-15.6**[abc]	**-20.2**[abc]	**-13.3**[bc]	**-12.9**	**-6.3**
Mean Scale Score					
2002	306.8	305.1	307.7	306.7	340.9
2007	327.3	331.3	325.8	322.1	359.5
Change Since Starting Year	**20.5**[abc]	**26.2**[abc]	**18.1**[ac]	**15.4**	**18.6**

[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.

The number of schools in Table ES.2.0 differs from that in Table ES.1.0 because not all schools have the same grade configurations. As noted above, any seeming discrepancies in computing the change from 2002 to 2007 are the result of rounding.

Figure ES.2.0: CST % Proficient & Above, YIP = 5, Grade = 4

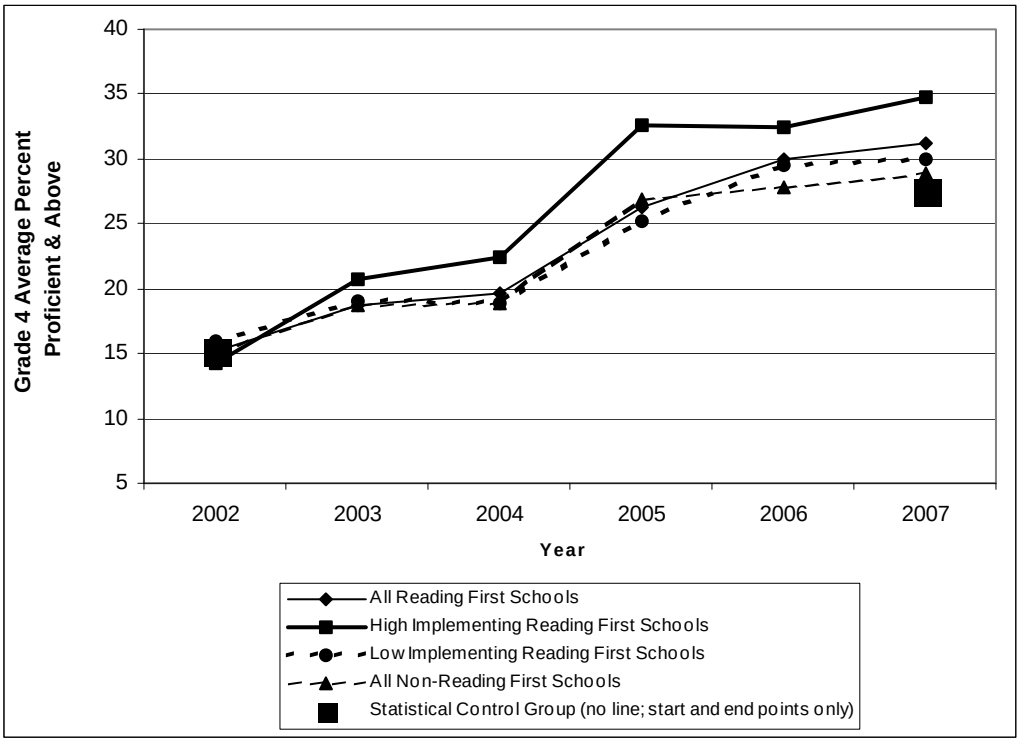


Figure ES.2.1: CST % Below Basic & Far Below Basic, YIP = 5, Grade = 4

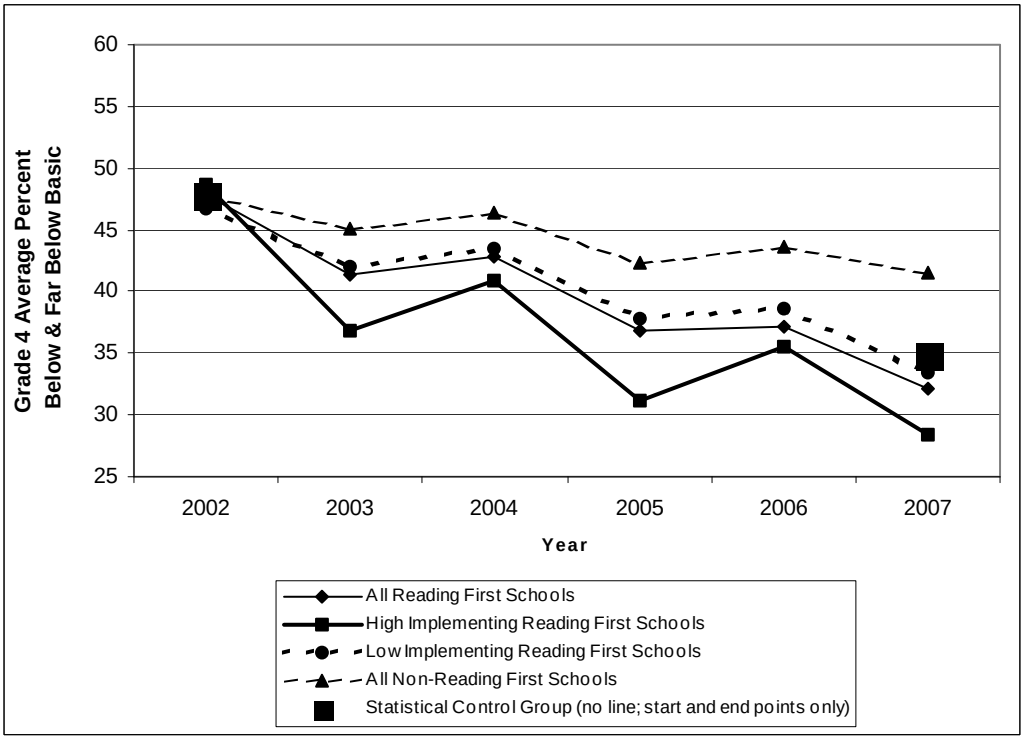
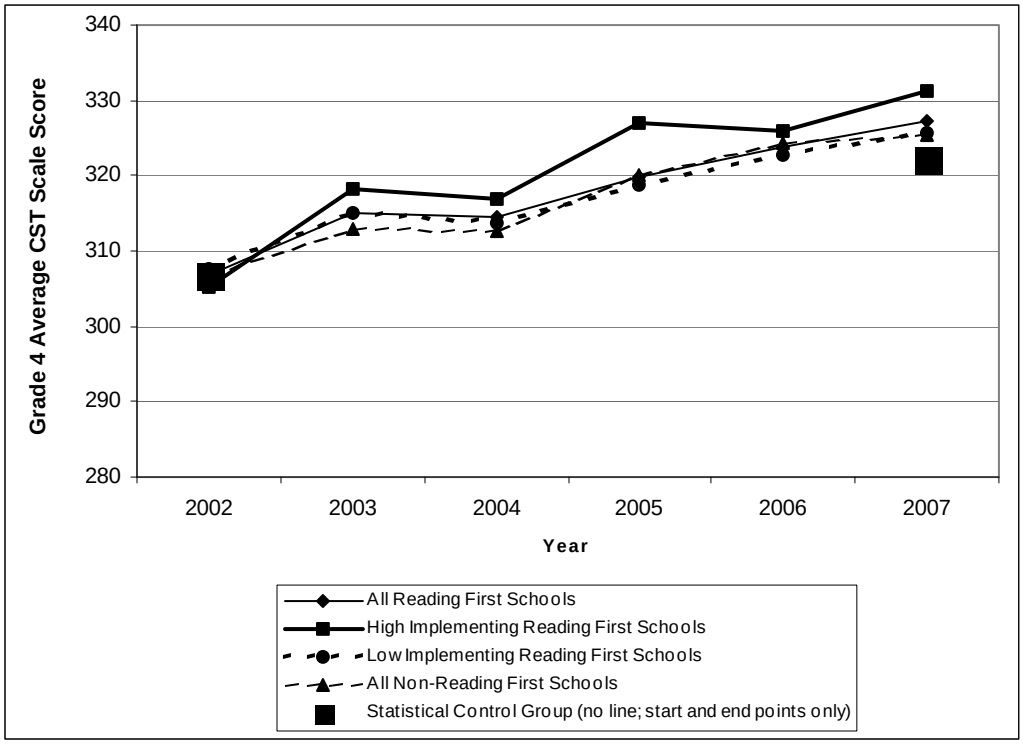


Figure ES.2.2: CST Mean Scale Score, YIP = 5, Grade = 4



Policy Recommendations

The policy recommendations of the California Reading First Year 5 Evaluation Report are unchanged from those of the California Reading First Year 4 Evaluation Report.[3]

- Continue to focus on full implementation of Reading First, especially in schools with high proportions of English learners.
- Support participation in Reading First over multiple years.
- Provide for a statewide data collection effort to facilitate accurate comparisons of student achievement across Reading First and non-Reading First schools, focusing on their use of the components that are required in Reading First. A statewide database of teacher and school data would confirm or refute the hypothesis that the statewide trend toward higher student proficiency is the result of voluntary adoption of program elements required as part of Reading First.
- Continue to support the extensive and focused professional development provided to teachers, coaches, and administrators. Continue to support the coaching model as a means for achieving instructional coherence and implementation of research-based instruction.

[3] *The California Reading First Year 4 Evaluation Report* can be accessed online at: www.eddata.com/resources/publications/.

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 five years of implementation from academic year 2002-03 to 2006-07.

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. It 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 out of bilingual instruction into 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 2-5, 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-embedded 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 (1,371) two Coach Institutes annually for in-depth training and a Leadership Program for selected experts (110) 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 75/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 five 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 (347 schools) has been receiving funding and implementing the program for approximately four and one-half years. LEAs in Cohort 2 (372 schools) were selected for funding in 2003-04. Cohort 3 (146 schools) was added in 2004-05. A small

number of LEAs were added in 2006-07 to make a new cohort, Cohort 4 (21 schools). A total of 886 schools in 120 LEAs are included in this Reading First Year 5 report.

California Reading First Year 5 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 4. This current report represents the Year 5 evaluation report, and will include outcomes from the 2006-07 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?

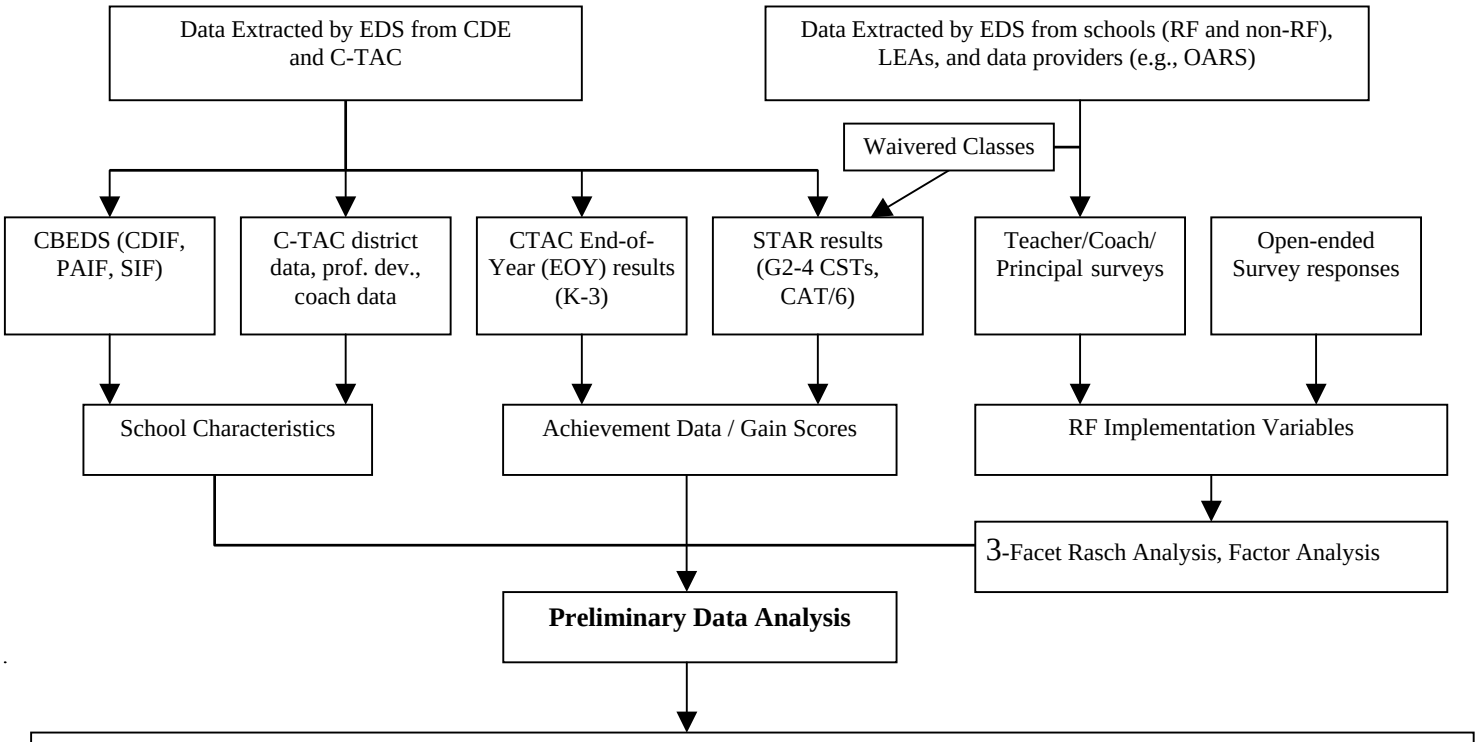
The conceptual framework below provides an overview of the evaluation study design. As described in the conceptual framework, the Reading First data can be organized into three types: a) school and district characteristics; b) achievement data; and c) implementation data. The school and district characteristics are described later in this chapter, with data drawn from state databases, including the California Basic Educational Data System (CBEDS) file, the demographic sections of the California English Language Development Test (CELDT) and STAR files, and the LEA-level database compiled by C-TAC to capture LEA internal evaluation data. The achievement data consist of school-level California Standards Test

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

(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) in a percentile metric, 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 average school performance level. An additional analysis yields the Reading First Achievement Index (RFAI), which combines the STAR and EOY data. To examine implementation, the Many-Facet Rasch models are 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. Note that in the first 3 chapters of the Year 5 report, there will be no qualitative analyses of open-ended survey responses. We refer the reader to this analysis in the Year 4 report and to supplemental chapters to be published as an extension of this report in January 2008.

Figure 1.1: Conceptual Framework – Year 5



Derived Variables of Study

Independent Variables

- Student subgroups: English language learners, special education, socio-economically disadvantaged, ethnic/racial subgroups
- School characteristics: RF or non-RF, number of students, aggregated student characteristics, aggregated teacher characteristics, cohort membership, years in program
- Teacher characteristics: years of experience, credentials, educational level, ethnicity, area of authorization, full/part time
- School-level program implementation variables, subscale measures on 17 implementation dimensions plus RFII (Reading First Implementation Index)

Dependent Variables

- Performance level proportions on the G2-4 CSTs and CAT/6 scores, End-of-Year results, Reading First Achievement Index (RFAI), school gain scores organized by years in program, STAR results for waivered and EL non-waivered classrooms

Data Analysis

- Descriptive statistics (means, standard deviations, graphs, cross-tabs, correlations)
- Rasch IRT Models (to convert survey data into measures)
- Factor Analysis (to group implementation variables)
- Multivariate Regression (to quantify the effect of implementation on achievement, after controlling for other factors)

Reports

(By school, years in program, cohort, student subgroup, and classroom waivered status)

- School achievement over time on various scales
- Implementation over time
- Effect of implementation on achievement over time
- Demographic descriptions of Reading First cohorts
- Reading First program evaluation recommendations

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 who 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 the starting place 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 were present in non-Reading First schools, making it impossible to discern the true impact of the Reading First program. The reader is referred to the Year 4 report for a complete discussion of these difficulties. For this Year 5 report, no data are reported for non-Reading First comparison schools due to inherent difficulties in establishing adequate comparisons. However, analyses are conducted using a statistically derived comparison group, as described in the Year 4 report 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 each subsequent year except for 2005-06, 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.

The following is a summary of the LEA cohorts, the typical YIP for that cohort, and the number of schools (a total of 886 in the 2006-07 academic year) from the cohort included in the current report:

(a) Cohort 1, first funded in 2002-03, with 13 LEAs (347 schools in current report); YIP 5
(b) Cohort 2, first funded in 2003-04, with 60 LEAs (372 schools in current report); YIP 4
(c) Cohort 3, first funded in 2004-05, with 27 LEAs (146 schools in current report); YIP 3
(d) Cohort 4, first funded in 2006-07, with 10 LEAs (21 schools in current report); YIP 1

The demographic data included in this chapter are extracted from the STAR research file published on the CDE website[3]. 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, and the CBEDS 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 as compared to all elementary schools in the state. Table 1.1 displays the percentage of SED students in each cohort of Reading First and in all elementary schools in the state for each year of the program. It is evident that Cohort 1 had the highest percentage of SED students compared to other cohorts, with 92.2% in 2007. Cohorts 2 and 3 in 2007 both had 86.8% SED students. Cohort 4 had the lowest percentage of SED students, 73.4%.

English Learners (ELs)

In 2007, Reading First schools also had higher percentages of ELs than the category of All Elementary Schools. The percentage of ELs in Cohorts 1, 2 and 3 was 53.7%, 54.7% and 58.5% respectively. Cohort 4, with 31.2% ELs, more closely resembled the statewide figure of 29.5%.

Students with Disabilities

In 2007, the percentage of students with disabilities was reported as 8.3% for Cohort 1, 7.6% for Cohort 2, 6.7% for Cohort 3 and 7.9% for Cohort 4. This varies somewhat from the statewide percentage of 10.6%. It is interesting to note that for Cohort 1, the percentage has risen over time while the percentage has dropped slightly for Cohorts 2 and 3 since their participation in Reading First.

Ethnicity Breakdown of Reading First Schools

Table 1.1 shows the percentage of students in each ethnicity category for each cohort, by year 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 1, 2 and 3 had significantly higher percentages of Hispanic students than Cohort 4. Additionally, it is evident that African American students were significantly over-represented in Cohort 1 compared to Cohorts 2, 3, and 4 and the All Elementary Schools category.

[3] The STAR research file used for the 2006-07 data was the version obtained by EDS on September 24, 2007, referred to as "P2."

Table 1.1: Student Demographic Data, 2003 to 2007

	Reading First Schools													All Elementary Schools[1]				
	Cohort 1					Cohort 2				Cohort 3			Cohort 4					
	2003	2004	2005	2006	2007	2004	2005	2006	2007	2005	2006	2007	2007	2003	2004	2005	2006	2007
Number of Schools	329	329	325	329	336	343	353	370	370	136	143	144	19	5823	5919	5977	5983	6057
SED (%)	91.3	92.7	89.1	91.4	92.2	82.7	86.7	83.5	86.8	85.1	85.8	86.8	73.4	51.0	51.6	53.3	53.4	54.0
EL (%)	58.5	58.6	58.8	56.5	53.7	53.0	55.5	54.9	54.7	57.5	57.2	58.5	31.2	27.2	28.2	29.3	29.3	29.5
Students with Disabilities (%)	7.5	8.4	8.5	8.4	8.3	8.0	7.7	7.9	7.6	7.1	7.7	6.7	7.9	9.8	11.0	11.1	10.8	10.6
African American (%)	17.2	16.6	15.4	14.1	13.7	8.8	8.2	8.0	7.8	6.6	6.4	6.3	14.1	7.8	7.8	7.6	7.8	7.7
American Indian (%)	0.3	0.3	0.3	0.3	0.3	1.0	0.9	0.8	0.7	0.8	0.6	0.8	7.7	1.3	1.3	1.3	1.4	1.4
Asian (%)	4.0	3.6	3.8	3.2	3.0	4.6	4.1	3.9	3.9	1.1	1.0	1.0	1.7	7.3	7.3	7.5	7.3	7.3
Filipino (%)	1.0	0.9	1.0	1.0	1.0	1.7	1.6	1.5	1.5	1.3	1.0	0.9	4.6	2.2	2.2	2.3	2.4	2.4
Hispanic (%)	71.5	73.3	74.4	76.5	77.2	72.0	74.1	75.2	76.1	77.1	78.6	79.7	50.5	40.2	41.5	42.6	43.6	44.1
Pacific Islander (%)	0.5	0.5	0.4	0.4	0.4	0.8	0.8	0.8	0.7	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.7	0.7
White (%)	3.7	3.4	3.2	2.8	2.7	9.6	8.8	8.0	7.5	11.2	10.3	9.4	19.3	36.5	35.2	33.9	33.0	32.3

[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 Reading First LEA cohorts and for all cohorts combined. In this table, it is evident that most of the LEAs in Cohort 1 were designated as large or mid-sized cities, while Cohort 2 included primarily large, mid-size and both large and mid-size fringe categories. Cohort 3 included mainly mid-size cities, urban fringe of large and mid-size cities and rural designations. Cohort 4 had the highest percentage of rural LEAs.

Table 1.2: Urban-Rural Distribution for Reading First Districts 2007

	Cohort 1		Cohort 2		Cohort 3		Cohort 4		All Cohorts	
District Location	N of Districts	Percent of Districts[1]	N of Districts	Percent of Districts	N of Districts	Percent of Districts	N of Districts	Percent of Districts	N of Districts	Percent of Districts
Large City	6	46.2	10	16.7	3	8.1	1	10.0	20	16. 7
Mid-size City	4	30.8	11	18.3	7	18.9	3	30.0	25	20.8
Urban Fringe of Large City	1	7.7	17	28.3	10	27.0	0	0.0	28	23.3
Urban Fringe of Mid-size City	1	7.7	14	23.3	9	24.3	1	10.0	25	20.8
Small Town	0	0.0	1	1.7	1	2.7	1	10.0	3	2.5
Rural	1	7.7	7	11.7	7	18.9	4	40.0	19	15.8
Total	13	100.0	60	100.0	37	100.0	10	100.0	120[2]	100.0

[1] The percent of the districts in that cohort in a particular type of location.
[2]There are 118 LEAs that are school districts and two that are independent charter schools.
Data source: National Center for Education Statistics (NCES)

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. Comparing cohorts, the teachers in Cohort 1 had somewhat lower percentages of advanced degrees than teachers in the other cohorts while Cohort 1 also had a higher proportion of teachers with bachelors degrees only. Examining the percent of teachers who were fully credentialed in each cohort, it is interesting to examine the changes over time in the percentages of fully credentialed teachers at Reading First schools. Cohort 1 had the greatest gain, moving from 77.8% to 95.55% in five years.

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 shows that Cohort 1 Reading First schools had lower Weighted Teacher Qualification indices (2.01 to 2.24) than the other cohorts (ranging from 2.24 to 2.36) and the non-Reading First schools.

Table 1.3: Elementary Teacher Credentials and Experience 2003 – 2007

	Reading First Schools													All Elementary Schools[2]				
	Cohort 1					Cohort 2				Cohort 3			Cohort 4					
	2003	2004	2005	2006	2007	2004	2005	2006	2007	2005	2006	2007	2007	2003	2004	2005	2006	2007
Number of Schools	329	329	325	329	347	359	344	370	372	135	143	146	20	5647	5694	5720	5837	5917
PhDs (%)	.55	.74	.75	.78	.71	.66	.66	.65	.70	.59	.51	.62	1.75	.90	.80	.80	.71	.69
Masters plus 30 or more semester units (%)	9.39	11.69	12.62	14.20	16.03	13.73	13.57	13.43	14.42	16.28	14.31	15.86	14.06	14.00	14.50	14.30	14.09	15.52
Masters (%)	10.95	11.80	12.23	12.43	13.02	16.86	18.53	20.01	20.18	16.63	13.55	13.77	17.40	15.50	16.90	18.10	18.97	19.32
Bachelors plus 30 or more semester units (%)	41.25	44.22	45.26	45.67	47.30	49.36	49.58	48.98	49.96	47.05	53.04	53.55	52.09	51.30	51.70	50.60	49.77	50.27
Total Advanced Degrees (%)	**62.14**	**68.45**	**70.87**	**73.07**	**77.06**	**80.61**	**82.35**	**83.06**	**85.25**	**80.56**	**81.41**	**83.81**	**85.30**	**81.70**	**83.90**	**83.80**	**83.54**	**85.80**
Bachelors (%)	35.06	30.92	28.38	26.37	22.74	19.30	16.89	16.82	14.67	19.33	18.34	16.11	14.55	16.40	15.80	15.90	16.22	14.05
Less than Bachelors (%)	.74	.58	.81	.53	.09	.10	.82	.08	.06	.22	.17	.08	.15	.20	.20	.40	.20	.10
Total Bachelors or less (%)	**35.80**	**31.50**	**29.19**	**26.90**	**22.83**	**19.40**	**17.71**	**16.90**	**14.74**	**19.55**	**18.52**	**16.19**	**14.69**	**16.60**	**16.00**	**16.30**	**16.42**	**14.15**
Weighted Teacher Qualification[1]	2.01	2.05	2.10	2.15	2.24	2.26	2.29	2.31	2.36	2.31	2.24	2.31	2.36	2.20	2.30	2.32	2.32	2.38
Fully Credentialed Teachers (%)	77.80	82.12	91.29	94.85	95.55	93.73	96.00	97.22	97.57	92.05	93.77	95.35	98.03	90.90	93.70	95.80	96.55	97.20
Average years teaching	10.75	10.91	11.27	11.47	11.57	11.25	11.60	11.87	11.96	11.40	11.88	11.95	13.06	12.70	12.80	12.80	12.89	13.00

[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:

- For this Year 5 report, no data are reported for comparison schools due to inherent difficulties in establishing adequate comparisons; however, a statistically derived comparison group is used in the achievement analyses in chapter 2.
- 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.
- Cohort 1 had the highest percentage of socio-economic disadvantage (SED) students at 92.2% in 2007 demographic files. Other cohorts ranged from 73.4% to 86.8%. The figure for All Elementary Schools was 54.0%.
- Reading First schools had higher percentages of ELs than the figure for All Elementary Schools (29.5%). Percentages of ELs in cohorts ranged from 31.2% to 58.5%.
- Reading First schools had higher percentages of Hispanic students and lower percentages of White students than the All Elementary Schools category.
- Cohorts 1, 2 and 3 had significantly higher percentages of Hispanic students than Cohort 4. Additionally, African American students were significantly over-represented in Cohort 1 compared to the other cohorts and the All Elementary Schools category.
- 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 mid-size cities, urban fringe of large and mid-size cities and rural designations. Cohort 4 had the highest percentage of 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 2006-07 CBEDS data, Cohort 1 Reading First schools had lower weighted teacher qualification indices than the other cohorts and the All Elementary Schools category.
- In 2007, 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 addresses a *new* question: To what extent does participation in the K-3 Reading First program improve student achievement in grade 4? The key findings in this chapter are:

- After controlling for school demographic characteristics, Reading First implementation is a statistically significant predictor of achievement on all achievement metrics, especially those associated with grades 2 and 4, at the 95% confidence level. The more faithfully the program is applied, the greater the effect on achievement.
- The Reading First Achievement Index (RFAI), a composite of K-3 achievement metrics for Reading First schools, has risen an average of 3.4 points per year, equivalent to 17 points over 5 years.
- Reading First schools out-perform a statistical control group by 1.6 points per year on the RFAI, equivalent to an 8-point advantage over 5 years.
- Since 2002, Reading First schools have shown significantly more growth than either non-Reading First schools or the statistical control group.
- 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 by a comparable 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 20 scale score points over a 5-year period.
- These findings are replicated in grade 4. Reading First schools grew 4.1 CST scale score points per year (20.5 scale score points over 5 years) in grade 4, versus 2.4 scale score points per year (12.0 points over 5 years) for the control group, a difference of 8.5 scale score points. Thus the program effect is sustained in grades to which the program is not administered.

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 California Achievement Test (CAT/6) – and the Reading First End-of-Year (EOY) curriculum-embedded assessments. As of this report, grade 4 CST results are included to assess the sustained effects of Reading First. Achievement is compared in four ways:

1. between years (gain scores)
2. between Reading First and non-Reading First schools
3. between Reading First schools and a statistical control group
4. between high implementation and low implementation Reading First schools[1]

The objective of this evaluation is to determine whether or not, and to what degree, the Reading First program is effective. What is meant by "effective"? According to the federal guidelines for Reading First, the program is effective to the degree it ensures "that 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.

1. 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 grade 2 and grade 4
2. Compare Reading First schools to comparable non-Reading First schools
3. Compare Reading First schools to a "statistical control group" by using statistical methods to profile how a school that is similar to Reading First schools would perform without access to the program
4. Compare high implementation Reading First schools to low implementation Reading First schools

The first approach looks at the absolute size of the achievement gains of Reading First schools from just before they started, and 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 Reading First 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 that are explained in Chapter 2 of that report.[2] Given the constraints of the study, it is not possible to identify non-Reading First schools that are not to some

[1] A detailed discussion of Reading First program implementation as embodied in the Reading First Implementation Index (RFII), an implementation statistic computed using responses to surveys administered to teachers, coaches, and principals in every Reading First school, is deferred to Chapter 3 of this report.
[2] See Chapter 2 of The California Reading First Year 4 Evaluation Report, available at: *www.eddata.com/resources/publications/.*

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 Year 4 and Year 5 Reports uses multiple regression to calculate the achievement gains that would be expected of schools that are similar to Reading First schools but that 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 RFII is based on a calculated relationship between implementation and achievement, which is used to extrapolate the performance of "non-implementing" schools, even though these do not exist *per se*.

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.

Based on these four approaches, Reading First will 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, and 4.
2. Reading First schools show higher achievement gains than all non-Reading First schools for grades 2, 3, and 4.
3. Reading First schools show higher achievement gains than what would be predicted from a statistical control group for grades 2, 3, and 4.
4. High Implementing Reading First schools show higher achievement gains than Low Implementing Reading First schools for grades 2, 3, and 4.

Measures of Achievement Gains

School progress or growth, also called achievement gains, is measured using the CSTs, the CAT/6 Mean Percentile Ranks (called here "MeanPR"), 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, and 4. The inclusion of grade 4 commences with the Year 5 Report because students in grade 4 can be expected to have experienced Reading First since kindergarten. Within ELA, we study the percentage of students per school that fall within each of the two following performance categories, which are a simplification 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.[3] 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. Using scale scores (with equal intervals) to measure growth reduces anomalies due to statistical artifacts caused by unequal intervals between values. Mean scale scores have not been provided in previous evaluation reports due to the NCLB emphasis on percentage of students in a performance category and to a desire to use reporting metrics with which the public and legislators are likely to be familiar. We introduce them in the Year 5 Report to address possible misinterpretation that growth is 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 confusion may have led to the perception that the rest of the students who do not change categories somehow have not grown, and that Reading First has not affected them. 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 score reported in the tables of this chapter is the 2007 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

[3] The "Basic" category has been 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.

averaged across a specified population of schools to produce the tabular statistics presented in this chapter.

CAT/6 MeanPR. As of the spring 2005 administration of the California STAR assessment, the CAT/6 component was discontinued in all elementary grades except for grade 3, so only grade 3 CAT/6 Reading, Language Arts, and Spelling data are used in this study. The "MeanPR" of a school is the average of the National Percentile Rank (NPR) scores of each of its students. The National Percentile Rank tells what percentage of students nationwide is expected to score below the student with a given NPR. An NPR of 45 would mean that the student is likely to score better than 45% of the national student population who take the tests. The MeanPR gain score for each school is its MeanPR in 2007 minus its MeanPR in the year immediately *preceding* its first year of Reading First implementation. The CAT/6 gain scores reported in the tables of this chapter are an average of these MeanPR gain scores across a specified sample of schools. Note that they are interpreted as a change in national percentile ranking, not as a change in the percentage of students meeting some benchmark or performance standard.

End-of-Year (EOY) Test. As the name suggests, the EOY is a curriculum-based test administered by all Reading First schools to students in grades K-3 at the end of the academic year. The kindergarten EOY test consists of eight subtests: Consonants, Lower Case Letters, Phonics, Rhyming, Syllables, Upper Case Letters, Vowels, and Consonant-Vowel-Consonant. The EOY tests for grades 1, 2 and 3 consist of 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 (in which instruction is conducted in Spanish). 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). The gain score for that grade is its 2007 EOY score minus its EOY score at the end of the *first* year of Reading First funding (not the year previous), which for schools in the program 4 or 5 years is 2004. For schools in the program 3 years, it is 2005.

Reading First Achievement Index (RFAI). The RFAI is a weighted combination of school-level percentages of students meeting various performance levels and benchmarks drawn from the CSTs, the CAT/6 Mean PR, and the EOY, with the heaviest weights placed on the CSTs. Refer to Appendix E for a detailed explanation of how the RFAI is computed. The RFAI was first computed in 2004. As of this study YIP 5 has four years of RFAI data (2004, 2005, 2006, 2007), as compared to six years of data for the CSTs. That is because the RFAI was not available in 2002 or 2003. Like the CST, each school RFAI can be interpreted as a percentage of students meeting a set of combined benchmarks and performance

levels. Because the RFAI is not based on a single benchmark or performance level, it is not interpretable as a single percentage. The RFAI gain score for each school is its 2007 RFAI minus its RFAI at the end of its *first* year of Reading First implementation.

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

Starting with the Year 4 report, for analyses of achievement schools have been grouped by Years in Program (YIP) rather than LEA funding cohort. As explained in the Year 4 Report, 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 was deemed necessary to group schools according to the actual year in which they started implementing the program rather than by the funding cohort of their LEA.

It is often found in educational research that intervention program effects often 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. YIP 5 is notably more urban than YIP 4 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 2007, we focus on just those Reading First schools that have been in the program for 5 years (the longest), 4 years, and 3 years (YIPs 5, 4, and 3). We have omitted schools in YIPs 1 and 2. They have relatively few schools (92 combined) and it has been established in previous reports that implementation tends to be relatively weak for many schools until the second or third year of the program.

Because the various achievement metrics did not all become available at the same time, the baselines for the achievement metrics vary. The CST metrics take 2002 (the year previous to implementation) as their baseline, whereas the EOY and RFAI take 2004 as their baseline. Each achievement gain takes the earliest year for which that achievement metric was available for that YIP. There is an additional complication relating to the baseline year for the Spanish version of the EOY test, which only became available in 2005. For more details about the relationship between the Reading First YIPs and the various achievement metrics, see Chapter 4 of the Year 4 Report.

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 in detail below. Nonetheless, we continue to report on 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.[4]

Comparison of High Implementation 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, 2005, 2006, and 2007 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. Now we define a high implementation school as 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.[5] This change has the effect of introducing a more stringent definition of high implementation, and also of leaving out the schools between 36.0 and 41.4 from the high and low groups. (They continue to be represented in the "All Reading First schools" category.) Therefore, the number of high implementation schools in 2007 is not comparable to that in 2006 or earlier.

[4] 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.

[5] 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.

It may be wondered why we used the "average yearly RFII" in defining the high and low groups. In trying to explain total gain in achievement since a school started in the program, since each year's implementation contributes to the total achievement gain score for a school, we need to take into account each year's implementation (RFII). Therefore, we sum the RFIIs across all the years the school has been in the program and divide by the number of years to come up with an average yearly RFII.

Nonetheless, acting on advice of the EAG, when reporting an individual RFII for each school we average its preliminary RFII (computed from the 2007 surveys) and its 2006 RFII on the theory that a rolling 2-year average is more stable and reliable than the RFII computed from a single year's worth of data.

Calculating Achievement for the Statistical Control Group

As discussed in chapters 2, 3 and 4 of the Year 4 Report, the statistical control group is defined using regression models to calculate the 2007 achievement score that a school which is similar to the Reading First schools (the same demographic and starting characteristics as the Reading First YIP under consideration) *would* obtain if it were *not* implementing the program. For reasons described in Chapter 3 of the Year 4 Report, we chose an RFII of 25 to signify a school that is not implementing the program. Thus, 25 is entered into the regression equation to calculate an expected 2007 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.) The detailed procedure for computing the statistical control group achievement statistics is described in Chapter 4 of the Year 4 Report.

Achievement Results

The following pages present a series of tables and trend-line charts showing starting scores, ending (2007) scores, and gains on each of 12 achievement metrics. They are the heart of the Year 5 Report and the basis of our finding that Reading First is an effective program. Table 2.1 summarizes the gains of *all* Reading First schools taken as a whole, not broken out by Years in Program (YIP). Presenting gains of schools that have been in the program differing lengths of time, this table compares them using an "average yearly achievement gain" metric. This metric differs from the metric in the YIP-specific tables, which report *total* achievement gain since the starting year.

Table 2.2 reports total RFAI gains broken out for YIPs 3, 4, and 5.

Tables 2.3, 2.4, and 2.5, with accompanying trend-line charts, show total gains on the CST and CAT/6 metrics for YIP 5, Grades 2, 3, and 4. Similar tables and charts for YIPs 3 and 4 are available in Appendix D.

Before presenting the achievement results, we touch on two points that may prove useful in interpreting the data in the tables:

1. Interpreting Significance Tests. The statistics in the achievement tables provided in this chapter are sometimes 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 Elementary Schools" group. The "c" means the new group average is significantly higher than where it started from, 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.

2. Rounding Errors. Sometimes we report a gain score that does not appear to equal the difference between the starting score and the ending score for a given metric that may be off a decimal value. The explanation is that the reported starting and ending scores have been rounded to one decimal place, whereas the reported difference or gain was 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.

Summary Gains (Table 2.1)

Table 2.1, reports average yearly gains for all Reading First and non-Reading First schools across all the YIPs (Years in Program) for each achievement metric. As such, it summarizes all the primary findings of the Year 5 report and answers the question, "What has been the effect of Reading First on all schools currently in the program?" Because it combines all five YIPs in one set of statistics, it does not report starting scores and ending scores since these naturally differ for each YIP. For the same reason, it is not accompanied by a trend-line chart.

Table 2.1: Summary Gains, All YIPs Combined, All Grades, Mean Yearly Gain

All YIPs Combined **All Grades** **Mean Yearly Gain (Average Change Per Year)**	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)	
Grade 2, CSTs	(N=831)	(N=137)	(N=295)	(N=N/A)	(N=4053)
% Proficient and Above	3.8[abc]	4.3[abc]	3.5[abc]	2.8	2.9
% Below and Far Below Basic	-3.3[abc]	-3.9[abc]	-3.0[abc]	-2.2	-1.5
Scale Score Metric	4.5[abc]	5.1[abc]	4.1[abc]	3.1	3.5
Grade 3, CSTs	(N=832)	(N=138)	(N=296)	(N=N/A)	(N=4048)
% Proficient and Above	1.6[abc]	1.8[bc]	1.4[bc]	1.4	0.2
% Below and Far Below Basic	-2.8[abc]	-2.9[abc]	-2.7[bc]	-2.3	-0.8
Scale Score Metric	2.9[bc]	3.1[bc]	2.7[bc]	2.6	0.5
Grade 3, CAT/6, Mean Percentile Rank	(N=832)	(N=138)	(N=296)	(N=N/A)	(N=4045)
Reading, Mean PR Metric	1.0[abc]	1.2[abc]	1.0[abc]	0.7	0.1
Language, Mean PR Metric	1.2[abc]	1.3[bc]	1.1[bc]	1.0	0.5
Spelling, Mean PR Metric	2.4[abc]	2.8[abc]	2.2[abc]	1.7	1.0
Grade 4, CSTs	(N=255)[1]	(N=26)	(N=101)	(N=N/A)	(N=3992)
% Proficient and Above	3.2[abc]	4.1[abc]	2.8[bc]	2.0	2.7
% Below and Far Below Basic	-3.1[bc]	-4.0[abc]	-2.7[bc]	-2.3	-1.3
Scale Score Metric	4.1[abc]	5.2[abc]	3.6[bc]	2.4	3.7
Reading First Achievement Index	(N=826)	(N=135)	(N=293)	(N=N/A)	-
RFAI Metric	3.4[abc]	3.6[abc]	3.4[abc]	1.8	-

[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.

[1] The grade 4 sample includes only YIP 5 schools, hence the much smaller N.

These statistics report the average difference between a school's starting score, in the year previous to entry into Reading First (except for the RFAI, which started in 2004 and is relative to the first implementation year), and its ending year in 2007, divided by the number of years it has been in the program. Thus it is the average growth per year on a variety of metrics. Because these statistics reflect *average* yearly gains rather than *total* gains, they are smaller than the statistics reported in Tables 2.2 – 2.5. Multiply by 5 to get a 5-year expected gain.

The story is consistent. Growth is substantial in grades 2 and 4, more modest in grade 3. Reading First schools grow faster than the statistical control group and the other elementary schools in the state. High implementing schools grow faster than low implementing schools.

Note that the number of schools in each grade is not necessarily the same. This reflects the fact that not all schools teach the same grades or have complete data. The grade 4 number of schools reflects the fact

that grade 4 data were collected only for YIP 5 schools. The N's of the high and low implementing schools do not necessarily add up to the N of all implementing schools because many schools have RFII statistics higher than 36.0 and less than 41.4 and don't fall in either the "low" or "high" category.

RFAI Gains (Table 2.2)

Table 2.2 reports starting points, ending points, and total RFAI gains for YIP 3, 4, and 5 schools, starting with 2004 (the first year the RFAI was computed) or 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 2.2: RFAI Gains, YIPs 3, 4 and 5

	Reading First Schools			
	All Reading First Schools	High Implementation Schools	Low Implementation Schools	Statistical Control Group
Year in Program: 5				
Number of Schools	261	28	102	N/A
2004	36.4	38.1	35.4	36.4
2007	45.4	48.5	44.4	44.9
RFAI Gain	9.0[c]	10.4 [c]	8.9[c]	8.5
Year in Program: 4				
Number of Schools	371	75	119	N/A
2004	34.5	35.7	33.2	34.4
2007	44.9	46.9	42.7	42.3
RFAI Gain	10.4[ac]	11.2 [ac]	9.5 [ac]	8.0
Year in Program: 3				
Number of Schools	151	26	57	N/A
2005	34.4	37.0	31.0	34.4
2007	42.8	44.6	40.2	40.7
RFAI Gain	8.4[ac]	7.5 [c]	9.3 [ac]	6.4

[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 in general support the hypothesis that Reading First schools are growing, that they grow more quickly than the statistical control group, and that high implementing schools grow faster than low implementing schools (though not all differences are statistically significant). The schools in YIP 3 offer an exception to the pattern, however. High implementing schools show a smaller gain than low implementing schools. In considering this exception and the differences that are not statistically significant, it is worth bearing in mind that the high implementing schools had a substantially higher starting RFAI (37.0) than the low implementing schools, which might have depressed their growth. It is

also worth bearing in mind that the RFAI statistic is 45% composed of data from grade 3 which, as is discussed later in this chapter, has a more complex relationship with implementation and years in program than grade 2 does.

CST Results for Grade 2 (Table 2.3 and Figures 2.3a – 2.3c)

Table 2.3 reports the starting and ending grade 2 CST scores of students in schools that have been in the program five years.

Table 2.3: CST Metric, YIP = 5, Grade = 2

Years in Program: 5 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	259	28	101	N/A	4,053
% Proficient and Above					
2002	15.4	14.8	14.8	15.4	37.8
2007	34.2	36.7	33.0	30.4	52.3
Change Since Starting Year	**18.9**[abc]	**22.0**[abc]	**18.2**[abc]	**15.0**	**14.5**
% Below and Far Below Basic					
2002	54.3	53.8	55.6	54.3	30.5
2007	36.7	33.6	38.7	41.1	23.0
Change Since Starting Year	**-17.6**[abc]	**-20.2**[abc]	**-16.9**[abc]	**-13.2**	**-7.6**
Mean Scale Score					
2002	299.8	299.5	298.5	299.8	333.4
2007	324.7	328.6	322.3	318.8	350.9
Change Since Starting Year	**25.0**[abc]	**29.0**[abc]	**23.8**[abc]	**19.0**	**17.5**

[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 gains in "% Proficient and Above" have risen from 15.7 percentage points in the Year 4 Report to 18.9 percentage points in Year 5, continuing a strong growth trend, although the gain is somewhat less from 2006 to 2007 than it was in previous years, or five per year. The growth rate is equivalent to a gain of 25 scale score points on the grade 2 CST over five years. Consider that the scale score difference between "Basic" (which starts at 300) and "Proficient" (which starts at 350) is 50 scale score points and that the CSTs range from approximately 200 to 500. If Reading First schools continue their current growth trajectory, they will have moved one whole performance level in 10 years, from 300 to 350, from the average student scoring "Basic" to the average student scoring "Proficient." This trajectory is more remarkable when one remembers that this gain is at the school level, 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 could interpret this rate of growth to mean that Reading First schools are now entering each new cohort of students 25 scale score points further up the scale in three years than they were, with similar cohorts, five years ago. Relative to the ordinarily slow pace of school improvement, and in light of the fact that this average comprises more than 30,000 students in YIP 5 alone, the pace of change is considerable.

We see that the rest of the state's elementary schools have also shown significant growth, but they lag behind Reading First schools by 7.5 scale score points. We also see that lower performing students are moving out of the bottom performance levels at the same rate that mid-range students are moving into the top two performance levels, a pattern not reproduced in non-Reading First schools. Students in non-Reading First schools exit the lower categories at almost half the rate that students enter the top categories. This is a key and important difference between Reading First and non-Reading First schools, one that holds up even in light of the fact that the two groups of schools are not ordinarily comparable.

We also see that gains for high implementing schools are 4 scale score points higher than for Reading First schools as a whole, more than 5 scale score points higher than for low implementing schools. This demonstrates that fidelity of implementation makes a measurable improvement in achievement.

Figures 2.3a, 2.3b, and 2.3c make the same points graphically by showing how school groups with differing degrees of participation in the program (i.e., implementation) experience different rates of growth.

Note that 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.

Figure 2.3a: CST % Proficient & Above, YIP = 5, Grade = 2

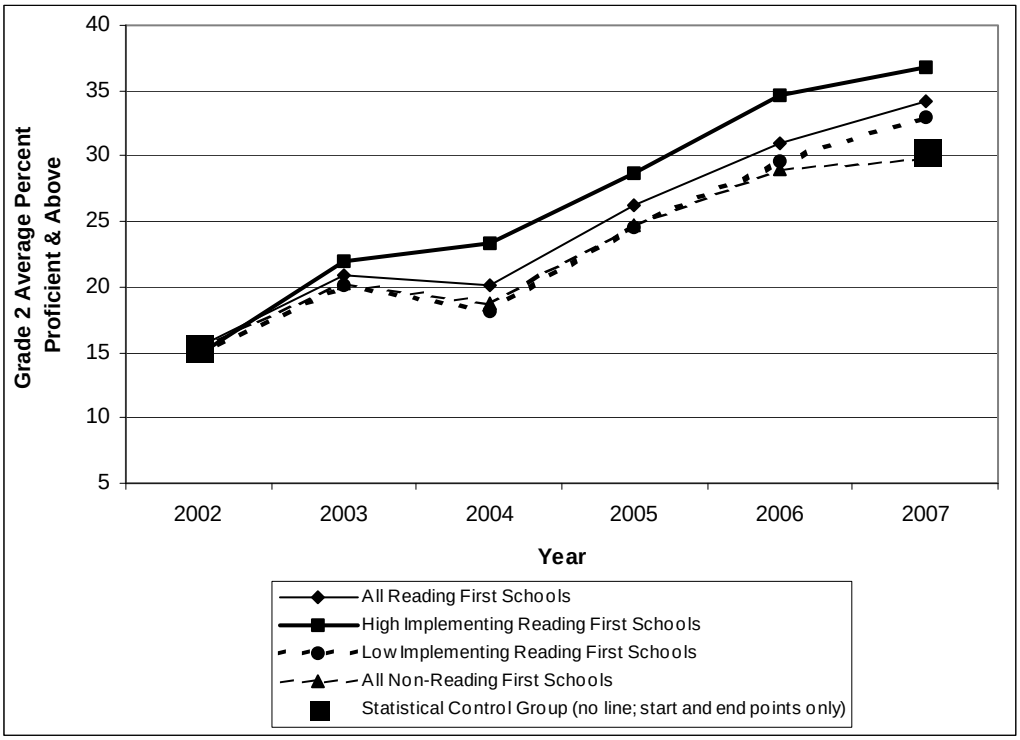


Figure 2.3b: CST % Below and Far Below Basic, YIP = 5, Grade = 2

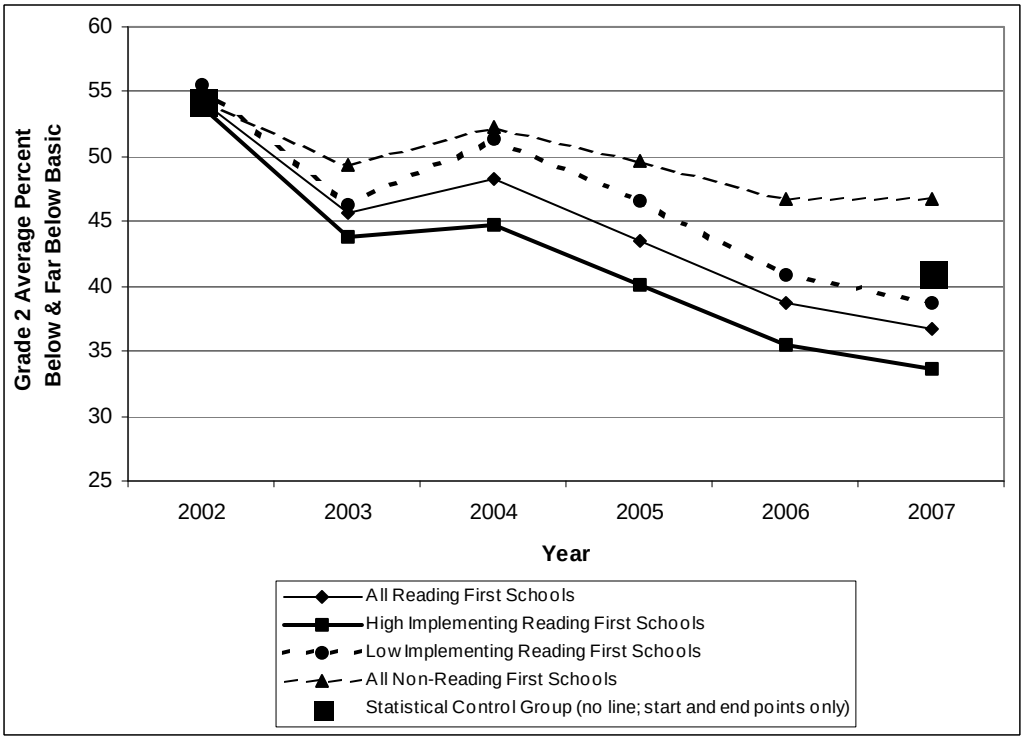
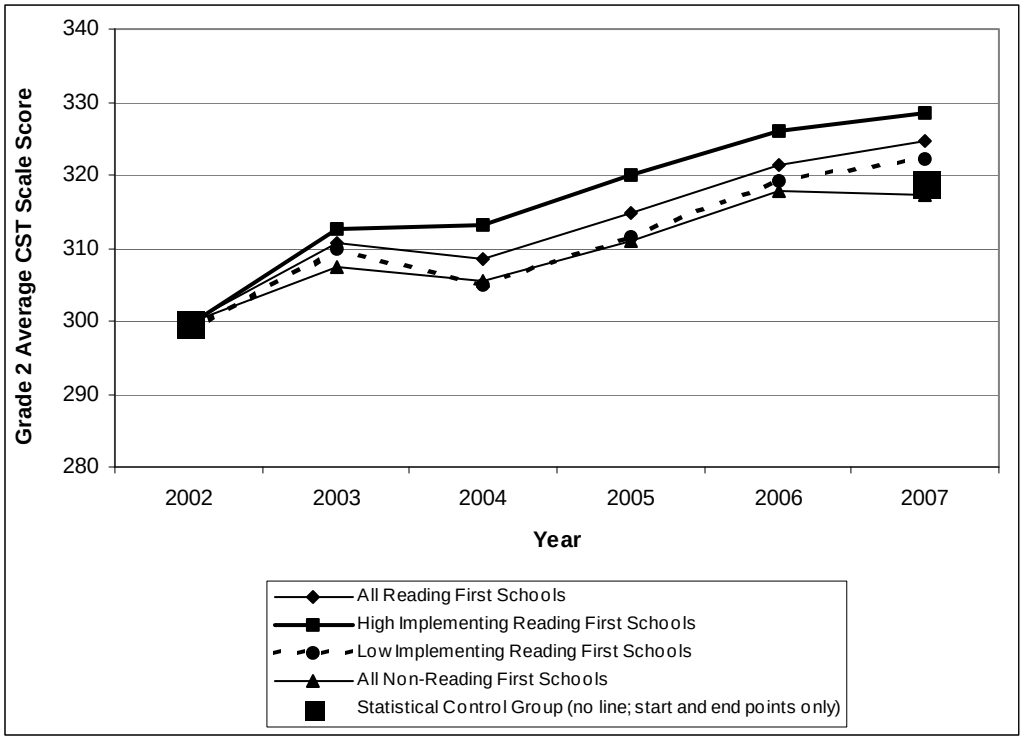


Figure 2.3c: CST Mean Scale Score, YIP = 5, Grade = 2



In addition to the patterns discussed above, we see that growth on the grade 2 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. This "fan" pattern strongly supports the finding of program efficacy and rules out the hypothesis that different growth rates are an artifact of different starting points. We also see that the statistical control group and the non-Reading First population have similar growth rates, supporting our contention that the statistical control group is a reasonable proxy for comparable non-Reading First schools.

Notice that the growth from 2006 to 2007 is flatter than for previous years. If this flattening continues in 2008 it may be evidence of a "plateau" effect, about which we have hypothesized in previous reports.

CST and CAT/6 Results for Grade 3 (Table 2.4 and Figures 2.4a – 2.4f)

Table 2.4 reports gains, starting scores, and ending scores for grade 3. In addition to CST scores, grade 3 offers CAT/6 scores for three subject areas: Reading, Language Arts, and Spelling. Grade 3 is unique in this regard, and the extra information proves critical in interpreting the grade 3 results.

Table 2.4: CST and CAT/6 Metrics, YIP = 5, Grade = 3

Years in Program: 5 **Grade: 3**	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	259	28	101	N/A	4,048
% Proficient and Above					
2002	14.8	13.2	14.7	14.8	40.1
2007	20.8	25.8	20.0	20.0	41.0
Change Since Starting Year	**6.0**[bc]	**12.6**[abc]	**5.3**[bc]	**5.2**	**1.0**
% Below and Far Below Basic					
2002	57.9	58.4	57.9	57.9	31.2
2007	45.0	40.3	46.6	46.3	27.4
Change Since Starting Year	**-12.9**[bc]	**-18.1**[abc]	**-11.3**[bc]	**-11.6**	**-3.8**
Mean Scale Score					
2002	294.5	293.7	294.3	294.5	333.9
2007	307.4	313.8	305.8	306.7	336.1
Change Since Starting Year	**12.9**[bc]	**20.2**[abc]	**11.4**[bc]	**12.2**	**2.3**
CAT/6 Mean Percentile Rank Reading					
2002	22.5	22.8	22.1	22.5	45.8
2007	27.4	30.5	26.4	26.7	46.3
Change Since Starting Year	**4.9**[bc]	**7.6**[abc]	**4.3**[bc]	**4.1**	**0.5**
CAT/6 Mean Percentile Rank Language					
2002	24.8	24.4	24.9	24.8	44.6
2007	30.4	33.4	29.6	29.9	47.2
Change Since Starting Year	**5.6**[bc]	**9.0**[abc]	**4.6**[bc]	**5.1**	**2.6**
CAT/6 Mean Percentile Rank Spelling					
2002	36.5	35.3	36.3	36.5	52.2
2007	49.6	52.4	47.6	46.6	57.4
Change Since Starting Year	**13.1**[abc]	**17.1**[abc]	**11.2**[bc]	**10.1**	**5.2**

[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.

Gains in grade 3 "% Proficient and Above" are less impressive than those for grade 2, one-third as much. Movement out of the bottom categories is substantial and lags grade 2 by only 5 percentage points. As with grade 2, the Reading First schools strongly out-perform non-Reading First schools in moving students out of the lower performance levels. The mean scale score gain is half that of grade 2. Gains relative to the Statistical Control Group are small, in most cases not significant. On the other hand, the differences between high- and low-implementing schools are much larger than for grade 2. Figures 2.4a – 2.4f reveal that grade 3 has a complexity not shared by the other grades.

Figure 2.4a: CST % Proficient & Above, YIP = 5, Grade = 3

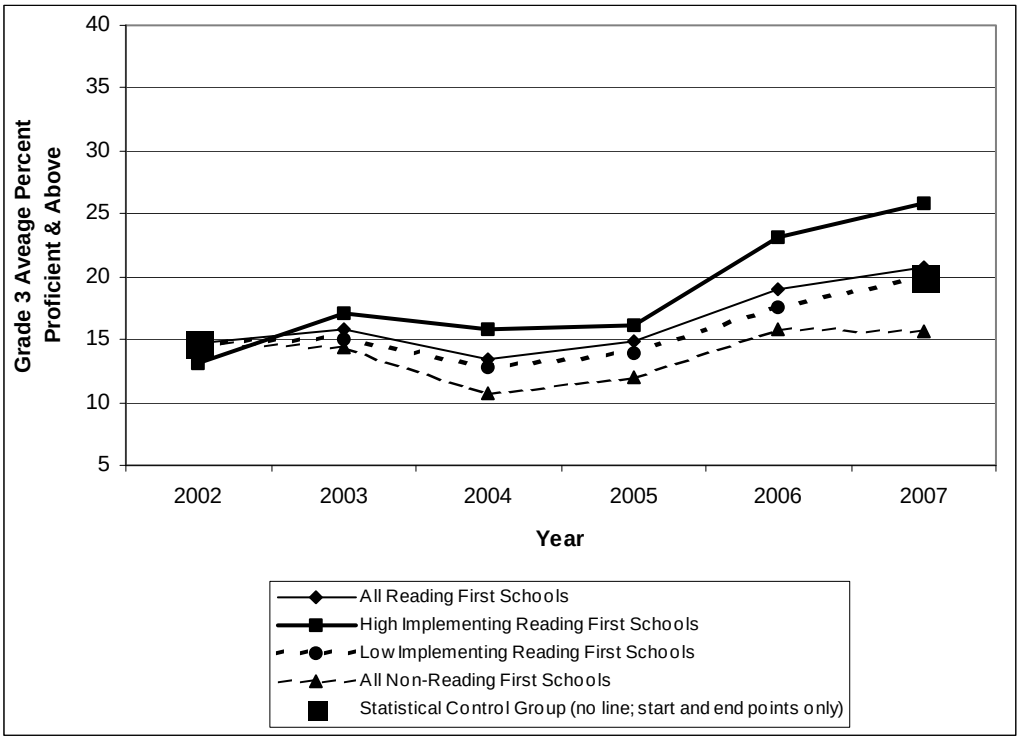


Figure 2.4b: CST % Below and Far Below Basic, YIP = 5, Grade = 3

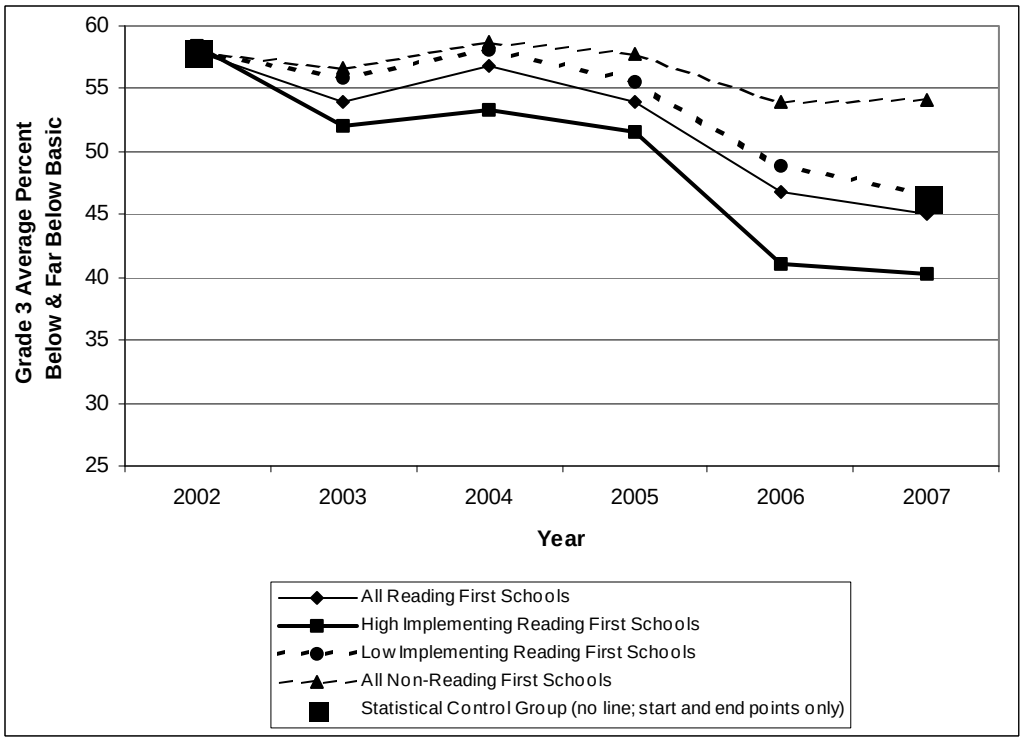


Figure 2.4c: CST Mean Scale Score Per Student, YIP = 5, Grade = 3

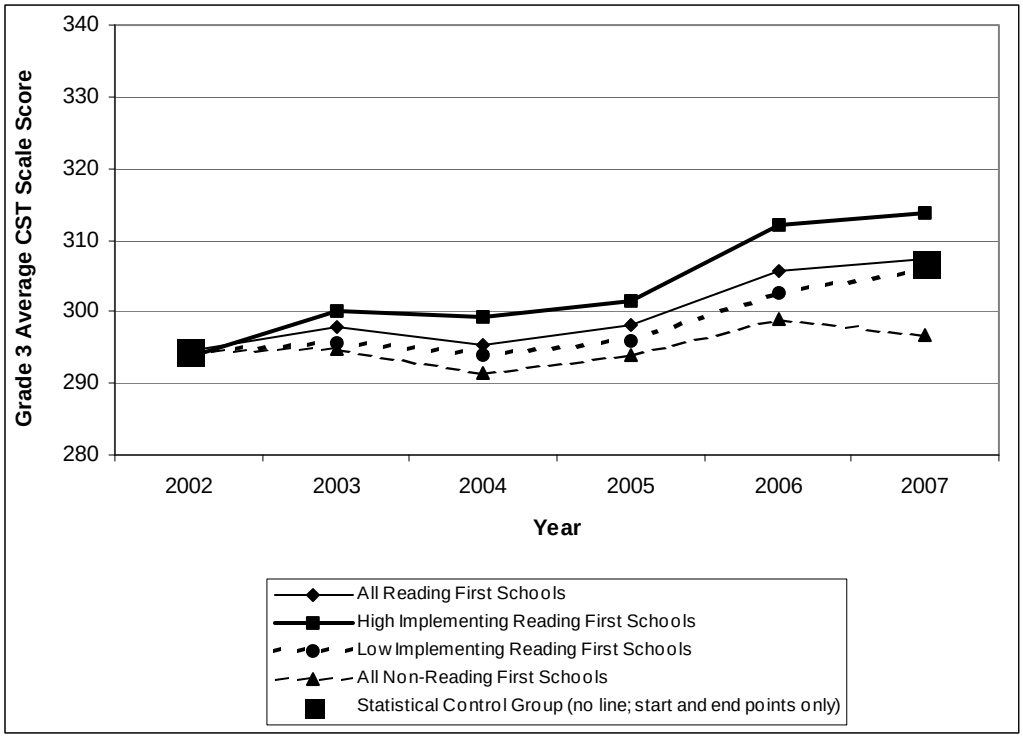


Figure 2.4d: CAT/6 Reading, Mean Percentile Rank, YIP = 5, Grade = 3

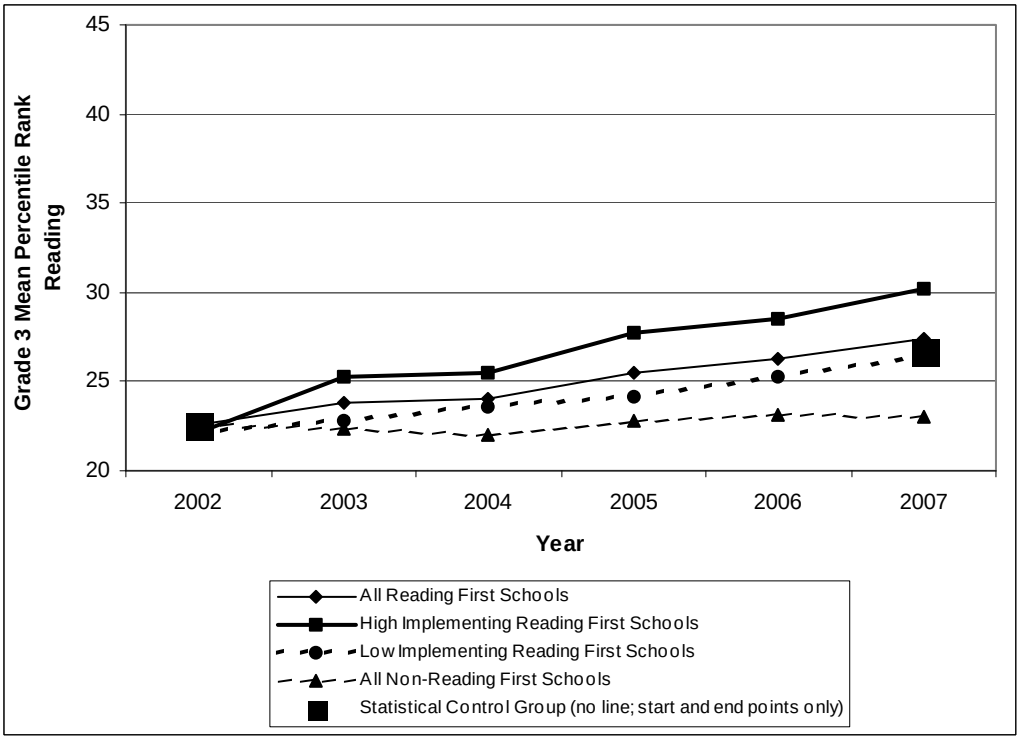


Figure 2.4e: CAT/6 Language, Mean Percentile Rank, YIP = 5, Grade = 3

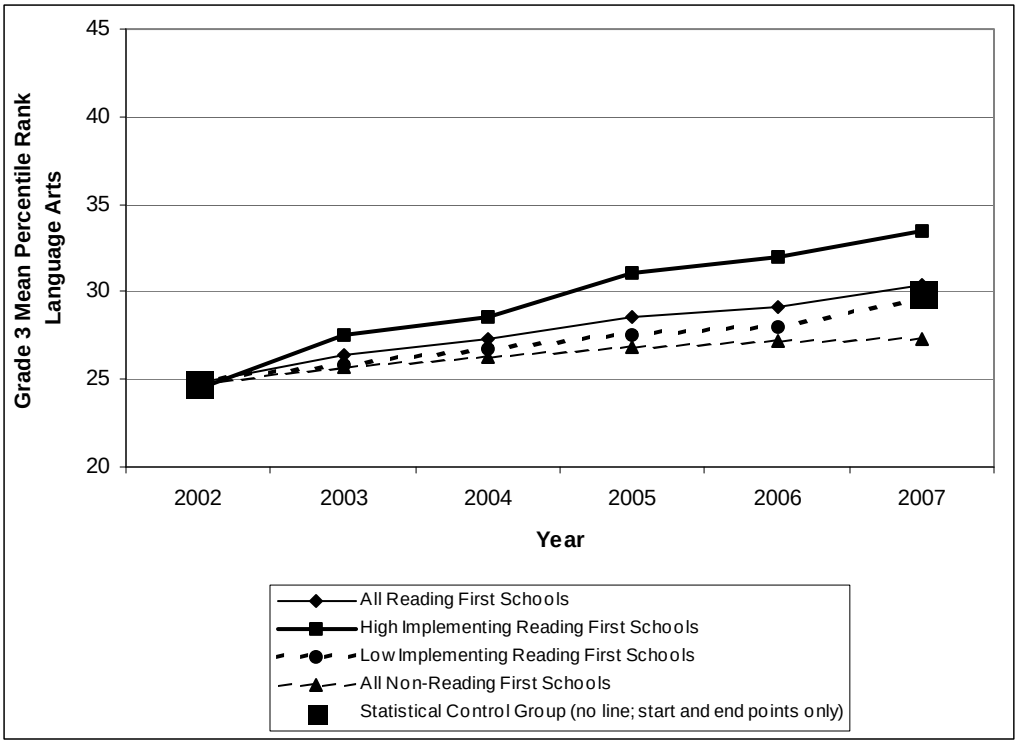
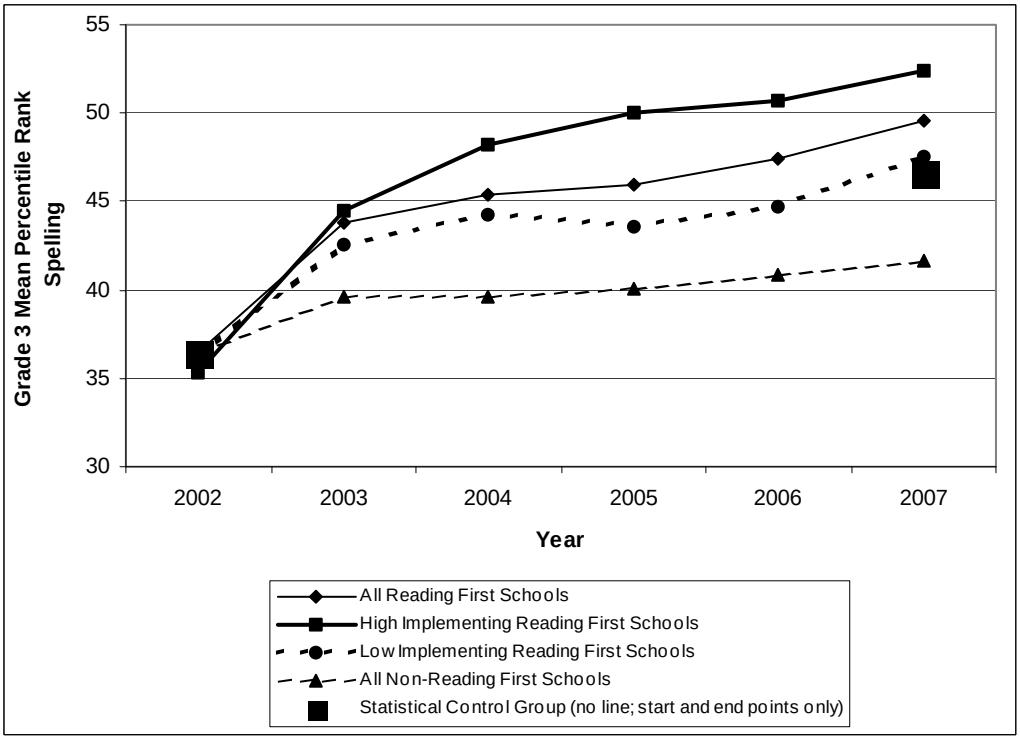


Figure 2.4f: CAT/6 Spelling, Mean Percentile Rank, YIP = 5, Grade = 3



Figures 2.4a – 2.4f reveal a number of important patterns that are not readily apparent in the statistics of Table 2.4. The most obvious, noted in preceding reports, is that the grade 3 CST scores dip substantially in 2004, creating a "U" shape. We see that even though the CST trends for Reading First schools is somewhat flat relative to grade 2 (Figures 2.4a – 2.4c), they are substantially more positive than those for the non-Reading First schools. After 2004, the trends are steadily positive, with a bit of flattening from 2006 to 2007.

A second pattern is that the CAT/6 trend-lines are qualitatively different than those for the CSTs. There is no "U" shape, just a steady positive trend ranging from slight in the cases of Reading and Language to large in the case of Spelling. In combination with the grade 2 and grade 4 results (below), this cautions us not to place too much weight on the shape and relative direction of the grade 3 CST trend-lines.

As regards the CAT/6 trend-lines, Spelling has a substantially higher starting point than Reading and Language and its trend lines range from 40 to 50 on the Mean Percentile Rank metric. This puts its trend-lines around the lower inflection point of the nationally normed CAT/6 population, where a given amount of ability growth is likely to show the largest changes in the percentile metric. The Reading and Language trend-lines are lower in the distribution where the same amount of ability growth will cause a smaller change in percentiles. This warns us that the absolute size of the trends in the CAT/6 metric may be in part an artifact of their position on the distribution.

A fourth pattern is that the statistical control group tends to show much higher gains, positive or negative, than the non-Reading First schools. In other words, for grade 3 the control group does not seem to behave like "comparable non-Reading First schools." It is behaving more like the low-implementing Reading First schools and is only marginally lower than the trend-line for all Reading First schools.

This highlights a fifth pattern. While the relative proximity of the low-implementing schools, all Reading First schools, and the statistical control group would seem to indicate a weak statistical relationship between implementation and achievement on the grade 3 metric, we see that the high implementing schools show *dramatically* higher gains than all the other schools. Thus, there does seem to be a strong implementation effect for grade 3, but only above a certain threshold of implementation, presumably around an RFII of 41. Schools below this threshold tend to show much more modest growth.

The sixth pattern is that while the "All Reading First" trend-lines may be modest relative to the high implementing schools, the trend-lines for non-Reading First schools show little or no growth on all the grade 3 achievement metrics. They do not seem to have improved much at all over the same period.

Thus, despite considerable statistical complexity, we find that Reading First efficacy is supported by the grade 3 achievement trend-lines.

CST Results for Grade 4 (Table 2.5 and Figures 2.5a – 2.5c)

Table 2.5 reports the CST results for grade 4 which have been collected only for YIP 5 schools. (The grade 2 and grade 3 results for YIPs 3 and 4 are reported in Appendix D.) Table 2.5 and its accompanying trend-lines demonstrate that Reading First is having a sustained effect that supports the student population as they move into the upper grades. This may prove to be the most telling of the Reading First effects since it supports the hypothesis that students in Reading First classrooms are learning skills that generalize beyond the course content and test material of the first few grades in elementary school.

Table 2.5: CSTs, YIP = 5, Grade = 4

Years in Program: 5 Grade: 4	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	255	26	101	N/A	3,992
% Proficient and Above					
2002	15.2	14.2	15.9	15.2	42.1
2007	31.3	34.8	30.0	27.5	55.7
Change Since Starting Year	**16.1[abc]**	**20.6[abc]**	**14.1[ac]**	**12.3**	**13.6**
% Below and Far Below Basic					
2002	47.8	48.6	46.7	47.8	23.2
2007	32.2	28.4	33.4	34.9	16.9
Change Since Starting Year	**-15.6[abc]**	**-20.2[abc]**	**-13.3[bc]**	**-12.9**	**-6.3**
Mean Scale Score					
2002	306.8	305.1	307.7	306.7	340.9
2007	327.3	331.3	325.8	322.1	359.5
Change Since Starting Year	**20.5[abc]**	**26.2[abc]**	**18.1[ac]**	**15.4**	**18.6**

[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.

Grade 4 sharply reinforces the growth picture presented by the grade 2 trend-lines. Reading First schools grow significantly faster than the control group. They grow marginally faster than the non-Reading First schools on the "% Proficient and Above" and "Mean Scale Score" metrics, but dramatically faster on the "% Below and Far Below Basic" metric. Movement out of the bottom two categories matches movement into the top categories, unlike non-Reading First schools. The average scale score growth is 20 points over five years, not far shy of the 25 points seen in grade 2. What makes this table remarkable is that Reading First is only administered in grades K-3. There is no grade 4 Reading First program. Yet the CST scores are almost what one would expect if Reading First extended to grade 4. This demonstrates

that Reading First students have been able to carry with them the skills and habits 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 slightly smaller and much less uniform across the population as can be seen in Figures 2.5a – 2.5c.

Figure 2.5a: CST % Proficient & Above, YIP = 5, Grade = 4

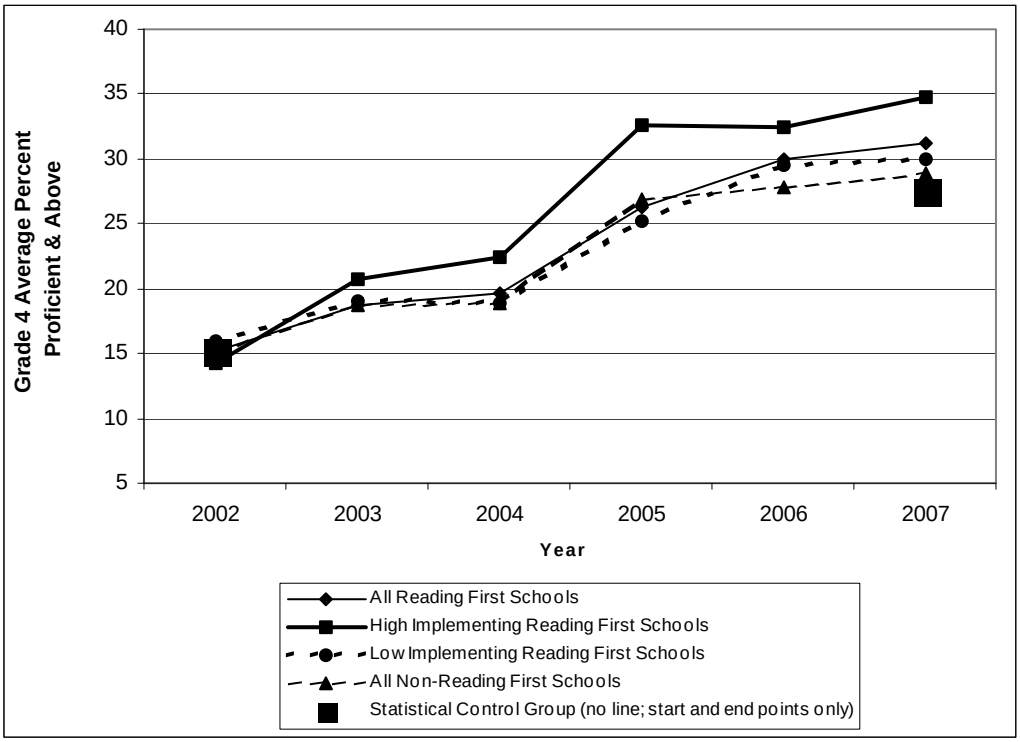


Figure 2.5b: CST % Below and Far Below Basic, YIP = 5, Grade = 4

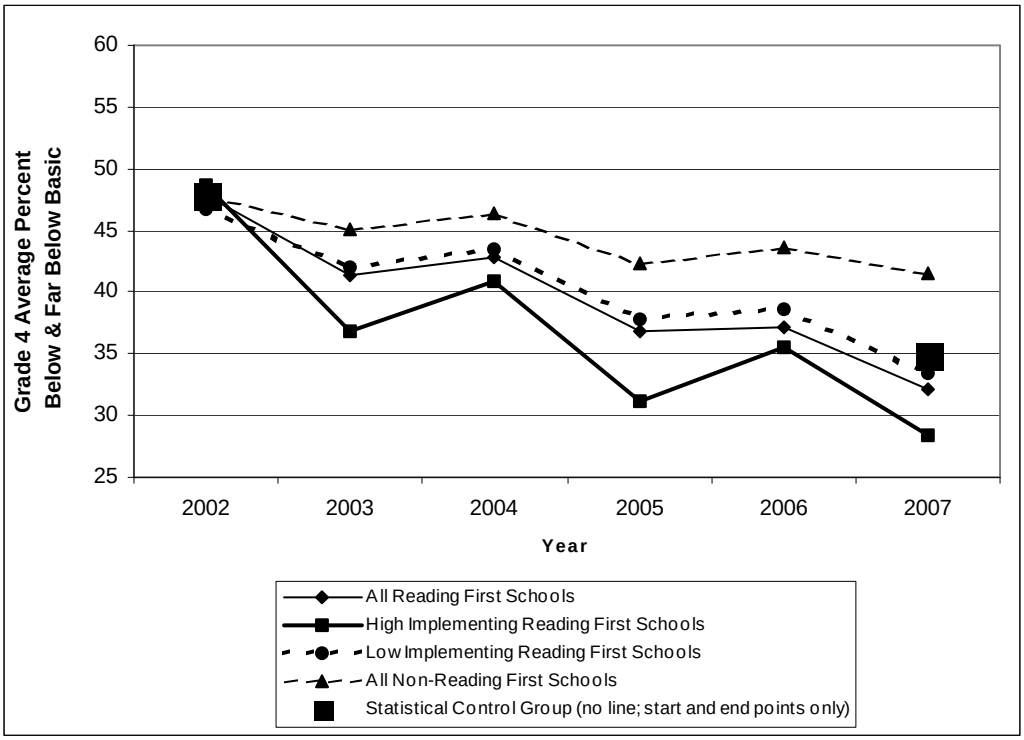
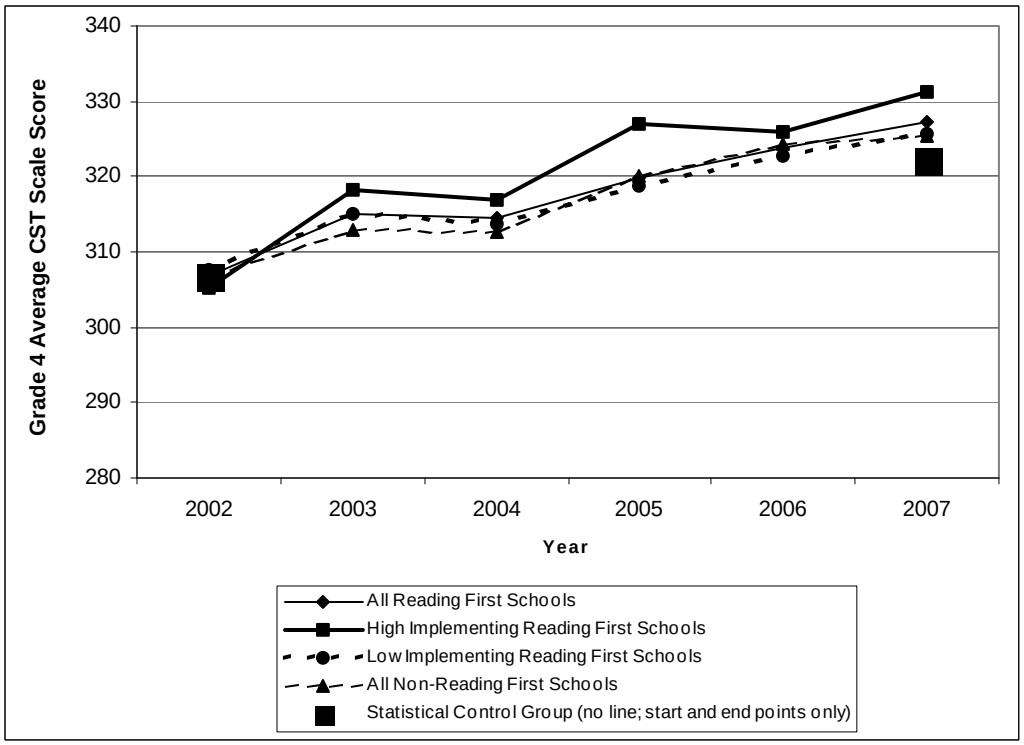


Figure 2.5c: CST Mean Scale Score, YIP = 5, Grade = 4



Figures 2.5a (% Proficient and Above) and 2.5c (Mean Scale Score) suggest that the All Reading First, Low Implementing, and non-Reading First trend-lines, as well as the statistical control group, do not grow in ways that are particularly different. They all show robust growth. Unlike for grade 3, the significance statistics for grade 4 (Table 2.5) best capture the differential effect of Reading First.

However Figure 2.5b, the effect of Reading First on movement out of the lower categories, reveals that Reading First schools far outpace non-Reading First schools in the lower performance levels. Low-performing students in non-Reading First schools run a real risk of becoming mired in the Below Basic and Far Below Basic performance levels, unable to get beyond elementary reading tasks even as their peers surge ahead. By comparison, low-performing students in Reading First schools appear to enjoy a decisive advantage and have the tools to keep up with their peers.

In addition, we see in grade 4 a repetition of the pattern that was so evident in grade 3, that high implementing schools "break out" from the rest of the schools and produce distinct and impressive trend-lines. While the grade 4 high implementing trend-lines tend to be somewhat jagged, that is probably an artifact of the relatively small number of schools (26 out of 255) in this group. The smaller the sample, the less stable the trend-line. The "break out" effect suggests, again, that there is some threshold of implementation above which schools experience a qualitatively higher level of achievement and sustainability.

The grade 4 effect strongly supports the strategy of focusing on 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 central to the national Reading First initiative (e.g., Foorman & Torgesen, 2001; National Reading Panel, 2000; Snow, Burns & Griffin, 1998).

Regression Effect Sizes

We have mentioned that in order to calculate gains for the statistical control group we perform a regression analysis for every achievement variable. To summarize the results of these regressions, we present regression tables that show the effect of Reading First implementation on the RFAI and grade 4 CST "% Proficient and Above" metrics. In both cases, the dependent variable to be predicted is the 2007 score for the relevant achievement variable, i.e., the 2007 RFAI and the 2007 grade 4 percent Proficient and above. We select the RFAI as an outcome variable because it embodies data from the grade 2 and 3 CSTs, the CAT/6, and K-3 EOY data. We select the grade 4 percent Proficient and above as an outcome variable because it measures the degree to which participation in the K-3 Reading First program influences performance in subsequent grades. Also, it is the only achievement variable not contained in the RFAI.

Regression analysis involves identifying a number of "predictor" variables that contribute information regarding the "dependent" variable. In this case, we found that, in addition to "Average Yearly RFII" and "Years in Program," the percentage of Socio-Economically Disadvantaged (SED) students in a school was a significant predictor of achievement. The percentage of English learners proved not, in general, to be a significant predictor of the RFAI and grade 4 % Proficient and above dependent variables and was not included.

In a separate regression we multiply Average Yearly RFII by Years in Program to create a composite variable that reflects total Reading First implementation over the years in the program. This effect is presented with the others in Tables 2.6 and 2.7 and is highlighted with bold type. The footnotes refer to both tables below.

Table 2.6: Effect Size of Variables Predicting Percent of Students Proficient & Above on Grade 4 CSTs in 2007 ($R^2 = 0.22$)[1]

Predictor Variable (Predicting 2007 Grade 4 % Proficient & Above)	Standardized Beta Coefficient Effect (standard deviation units)[2]	t-test (t > 1.96 implies significance with 95% confidence)[3]	Probability the Effect is by Chance[4]
Starting CST Gr. 4 % Proficient & Above	0.40	11.0	0.0000
Number of Years in Program	0.18	5.4	0.0000
Average Yearly RFII	0.14	4.3	0.0000
Yearly RFII * Years in Program[5]	**0.21**	6.6	0.0000
Percent of SEDs in School	-0.12	-3.5	0.0006

Table 2.7: Effect Size of Variables Predicting the 2007 RFAI ($R^2 = 0.51$)[1]

Predictor Variable (Predicting the 2007 RFAI)	Standardized Beta Coefficient Effect (standard deviation units)[2]	t-test (t > 1.96 implies significance with 95% confidence)[3]	Probability the Effect is by Chance[4]
Starting RFAI	0.68	27.4	0.0000
Number of Years in Program	0.21	8.7	0.0000
Average Yearly RFII	0.07	2.9	0.0045
Yearly RFII * Years in Program[5]	**0.22**	9.2	0.0000
Percent of SEDs in School	-0.06	-2.5	0.0126

[1]The R^2 statistic reports the percentage of variance that is explained by the model.
[2]The "Standardized Beta Coefficient" shows how many standard deviations the CST "% Proficient & Above" increases for every one standard deviation increase of that predictor variable.
[3]The "t-test" shows how many times larger the effect is than what would be predicted by chance.
[4]The "Probability" column uses the t-statistic to compute the probability that the observed effect occurred by chance.
[5]The "Yearly RFII * Years in Program" predictor variable is the product of a school's "Average Yearly RFII" and its "Number of Years in Program" (equal to the sum of its RFII statistics over time). To avoid collinearity, its effect size was computed in a separate regression run in which "Average Yearly RFII" and "Years in Program" were removed.

How to Interpret the Regression Tables

The predictor variables we are interested in are "Number of Years in Program" and "Average Yearly RFII" and the variable that is obtained by multiplying them together, "Yearly RFII*Years in Program." The latter can be thought of as a school's total amount of Reading First implementation over time. The "Percent of SEDs in School" is not a variable of primary interest, but its role here is to remove confounding influences that socio-economic status might have on the implementation effect. The role of the "Starting" variable is to remove the effect of the school's achievement starting point so that we can treat all schools in the sample as if they started at the same performance level.

The two right columns – the t-test and the probability -- answer the question: How likely is it that we would have encountered the observed effect size in this row by chance? We see that the t-test statistics are all above 1.96 and the probabilities (which are calculated from the t-statistics) are all well below 0.05. That means it is very unlikely that we would have obtained these effect sizes by chance.

That Reading First implementation is a significant predictor of achievement gain supports the claim of efficacy. We have demonstrated that Reading First implementation matters and that it is not an artifact of SED or Starting Point or percent of English Learners (not shown here because it is not a significant predictor). Though we are naturally drawn to examine the effect size, called here the "standardized beta coefficient," it is very difficult to interpret and there are no accepted industry standards on how to decide whether a given standardized beta coefficient is "good" or not. To examine "effect size," it is better to examine the relative gains reported in the tables earlier in this chapter, especially relative to the statistical control group gains that were computed, in fact, from regression tables just like these.

That said, the accepted way to interpret a standardized beta coefficient is as the change in standard deviation units that is expected in the dependent variable given a change of one standard deviation in the predictor variable. We see that the "Yearly RFII*Years in Program" beta coefficient equals 0.21 and 0.22 for the Year 4 Proficient & Above and RFAI dependent variables. That means for every one standard deviation increase on the total implementation scale we can expect achievement to increase about one fifth of a standard deviation on the achievement scale. Calculating the standard deviations of the variables and looking at typical total implementation levels, we see that this means that total implementation will generally increase both the RFAI and grade 4 scores around 6 to 8 points in a 5-year period for an average implementing school. And that is what we see in the trend-lines – generally a 6 to 8 point difference between the statistical control group and the average Reading First school after 5 years, depending on the achievement metric.

Measurement Error Lowers Effect Size

While a 0.21 or 0.22 effect size is reasonably large for a study of this kind, it strongly *understates* the "true" effect of Reading First implementation on achievement, both in the regression tables and in the trend-line charts and tables.

The standardized beta coefficient and the gain score differences between high and low implementing schools and between all Reading First schools and the statistical control group – all of these differences assume that our RFII implementation measures and achievement measures are *perfectly* precise. Obviously they are not. Achievement tests have a wide margin of error when measuring student ability (reduced when aggregated to the school level). Most important, the RFII has a very large margin of error, which arises from a variety of sources:

- Ambiguity in the survey questions
- Biases caused by teachers overstating or understating their school's level of implementation
- Uncertainty caused by teachers not understanding the questions or encountering questions that do not apply to them
- Schools that show high achievement gains but report low Reading First implementation because of the use of effective non-Reading First programs, initiatives, and other causal factors
- Schools that report high Reading First implementation but show low achievement gains due to circumstances out of their control or perhaps to biased reporting.

As measurement error increases, the measured or *observed* effect becomes smaller in accordance with the statistical law known as "regression to the mean." This will happen even in cases where there is a nearly perfect causal effect. For instance, even the observed relationship between physical exercise and muscle tone can approach zero if the instruments used to measure physical exercise and muscle tone have a high degree of measurement error.

Therefore, it is important to remember that the effect sizes reported in the Reading First evaluation are on the conservative side, as they are for most evaluation studies. There is no obvious or widely accepted way to correct for measurement error with these kinds of variables.

Is the implementation effect as strong in the Year 5 Report as it was in the Year 4 Report?

The answer is yes, for the most part. The Year 4 "Yearly RFII*Years in Program" effect was 0.17. The corresponding Year 5 statistic is 0.22. However, the Average Yearly RFII effect, taken in isolation,

dropped from 0.09 to 0.07. As regards CST % Proficient & Above, even though the Year 4 report used the grade 2 results and the Year 5 report uses grade 4 results, we find that the effect sizes for all the implementation variables are virtually identical. Thus, grade 2 and grade 4 substantially confirm each other. The grade 3 CSTs have an anomalous relationship with all the remaining achievement variables, including the grade 3 CAT/6.

Why is the Average Yearly RFII coefficient different between the RFAI and Grade 4?

The Average Yearly RFII effect is 0.14 for grade 4 Proficient and Above but 0.07 for the RFAI – half as much. This is a consequence of the 45% weight of grade 3 achievement metrics in the calculation of the RFAI. Whatever anomalies exist in grade 3 are to some extent inherited by the RFAI. However, because the non-grade 3 achievement metrics "straighten out" grade 3's U-shaped CST trend-line, the Years in Program effect is larger on the RFAI than for grade 4, and the two implementation variables together are sufficient to yield an overall 0.22 "Mean RFII * Years in Program" effect for the RFAI.

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

When Average Yearly RFII is multiplied by Years in Program, we see that the total implementation effect on the RFAI is 0.22 (consistent with, or higher than, effects reported in the Year 3 and Year 4 reports), whereas the effect of Average Yearly RFII alone is smaller (0.07). We argue that the 0.22 effect size is probably the more valid estimate of the total Reading First effect, but this requires us to assume that the Years in Program effect is sensitive primarily to the school's implementation of Reading First and that it is independent of non-Reading First effects on achievement.

This is a strong assumption but not unreasonable. Reading First schools agree, as a condition of funding, not to implement competing programs or initiatives that are not aligned with Reading First. This has an important theoretical implication. Because we see strong achievement gains in schools with high RFIIs – in fact gains that are stronger than for lower implementing schools – and because higher RFIIs imply that such schools are implementing Reading First more *exclusively*, we can conclude that most of the Years in Program effect that we observe in high implementing Reading First schools is caused by Reading First and *not* by non-Reading First reading programs or non-Reading First pedagogical practices that would be precluded by the program.

This observation leads to a vitally important question. If the achievement gains experienced by Reading First schools over a five-year period are primarily a Reading First effect, as we suggest, why do non-Reading First schools also show substantial gains (though not as large) over the same period of time?

Assuming the gains are real, and not an artifact of the tests, we see two possibilities:

Non-Reading First schools have, over the same period, begun implementing non-Reading First educational strategies that happen to be effective; or

Non-Reading First schools have been implementing some or all of the same program elements that make Reading First effective.

A review of state educational initiatives supports the second possibility. The state, in January of 2002, adopted two reading curricula for K-8 schools to use. These are the same Houghton-Mifflin and Open Court reading programs required in Reading First. Schools that adopt these programs have access to SB 472 teacher professional development, AB 75 principal professional development, and the 6-8 week skills assessments. In addition, many LEAs and schools have opted to hire reading coaches at their own expense. Such non-Reading First schools become virtually indistinguishable from Reading First schools in terms of educational practices in the classroom. The main difference is that the non-Reading First schools must use other funding sources to hire reading coaches and provide professional development.

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

Conclusions

The conclusions in the Year 5 Report reinforce and extend those of the Year 4 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, and 4.
2. Reading First schools show higher achievement gains than non-Reading First schools for grades 2, 3, and 4.
3. Reading First schools show higher achievement gains than what would be predicted from a statistical control group for grades 2, 3, and 4.
4. High implementing Reading First schools show higher achievement gains than low implementing Reading First schools for grades 2, 3, and 4.

The Year 5 Report finds that the answer continues to be “yes” to all four questions with a small number of exceptions in particular instances. It extends these findings especially to grade 4. This confirms the findings of the previous grades and supports the hypothesis that students who have progressed through Reading First programs in grades K-3 are better prepared for higher grades than students who have not.

We conclude this chapter by restating from the Year 4 Report an important idea discussed in this chapter. 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 such schools since 2002 is the result of Reading First. Because the rest of the state K-3 schools have shown similar, though less dramatic, upward trends over the same time period, it is possible that the statewide trend in non-Reading First schools is being driven by the same program elements that are driving the Reading First gains. Anecdotal evidence suggests this to be the case. If subsequent research should bear out this hypothesis, it would validate efforts to make such program elements available to all California schools, not just those in Reading First.

References

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 of Reading First

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 2007. Because participation in the evaluation process is part of the commitment that local education agencies (LEAs) make when they apply for funding, the response rate on the surveys has been high. In 2007, a total of 17,261 usable surveys were received from teachers, 1,028 from reading coaches, and 1,073 from principals, totaling 19,362 and yielding a response rate of 91%.[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).
- Most schools in the Reading First program are implementing the program adequately.
- The average level of implementation has risen throughout the duration of the Reading First program. The average (RFII) across all schools was 39 in 2006 and 2007, compared to 36 in 2004 and 2005.
- Schools that have been in Reading First for two or more years have higher average implementation than newcomers.
- The RFII can be interpreted as a (theoretical) percentage of times that teachers rate their schools "more than adequate" on relevant survey questions. Using the distribution of school RFII measures, it is possible to state how many schools in the state meet the "more than adequate" standard from the point of view of teachers on selected dimensions.

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

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. Rather, 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 skimps on 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 2007 there were 886 Reading First schools in California. To measure implementation in each school, the external evaluator would ideally send trained auditors to observe each Reading First classroom over an extended period of time. While this would not be practical for the complete population of schools, it could in theory be done with a representative sample of schools (absent legal restrictions). However, the State has specifically requested 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.

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 would be a strong incentive for some school personnel 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 four years of the survey use.[2] To summarize, the above issues are addressed as follows:

[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.

6. Schools are measured relative to each other rather than against an absolute standard.
7. 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.
8. Questions are worded so that their "correct" answers are not immediately obvious, increasing the chance that respondents select truthful answers.
9. 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.).
10. 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 are based entirely on achievement data.
11. Equating methods are used to equate responses across respondent groups and across program years.
12. The potential ambiguity at the question level is addressed by using statistical methods to group items' coherent dimensions that seem to 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 well 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. 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 as a means for measuring implementation at the school level.

[3] Information on "significant progress" as available at: http://www.cde.ca.gov/pd/ca/rl/rdfst06achievedef.asp.

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 each round of revisions, efforts were made to retain enough "old questions" to link the different survey administrations together.

Anecdotal information received from teachers and coaches indicates that it took 20 to 30 minutes to complete the survey.

Calculating the Reading First Implementation Index (RFII)

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 program called Facets, subsets of questions across the three surveys are used to generate measures on 17-19 dimensions.[4]

Three of these dimensions are used to calculate each school's RFII. They are: School Implementation Overall (SIO), Overall Reading First Understanding (OUND), and Teacher/Coach Professional Development (TCPD).

[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.

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).

Based on advice from the EAG, starting in 2007, the RFII of a school in a given year is averaged with its RFII from the preceding year. It is hoped that this will make the RFII more robust to changes in the sample of teachers in each school who take the survey each year while allowing it to be reflective of the school's recent implementation history. For this report, when we refer to the 2007 RFII, it is actually the average of the 2006 and 2007 RFIIs for each school.

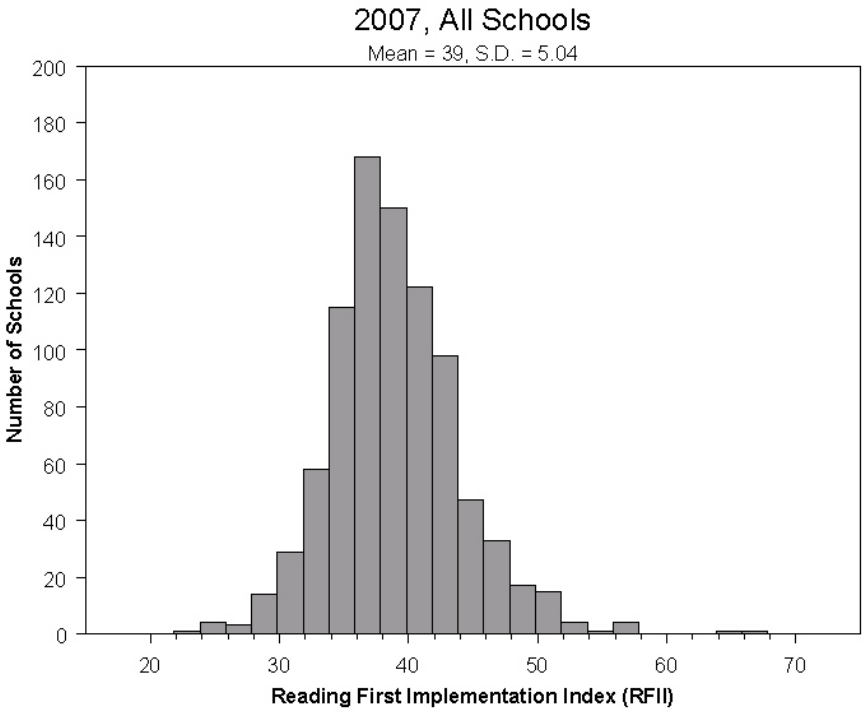
Implementation Results

Distribution and Interpretation of the RFII

Figure 3.1 shows how the RFII was distributed across all Reading First schools in 2007. The mean 2007 RFII was 39; the standard deviation around the mean was 5. This can be practically interpreted as follows: Reading First *teachers* on average found their schools to be "more than adequate" 39% of the time (i.e., on 39% of the relevant items). 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. Interpreting it as a percentage of items scored "more than adequate" makes it easier to understand, however.

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: All Schools – 2007 Reading First Implementation Index (RFII), Distribution of Schools



Degree of implementation is likely to develop over time. Schools with more years in the program (Years In Program, or YIP) may have different implementation profiles than schools newer to the program. The histograms in Figures 3.2 through 3.6 show the distribution of the RFII for schools participating in Reading First for different lengths of time. Measures on the far right tail of the distributions of the figures (above 55) should be viewed with caution; such schools tend to show unusually high inter-respondent agreement, possibly suggesting coached answers or collaborative group completion. In the histograms, some patterns are evident. First, there is a consistent pattern of the distribution across YIPs and all are similar to the All Schools graph in Figure 3.1. Schools in the program only one or two years (YIP 1 and YIP 2) have modes slightly lower than schools with more years in the program. YIP 2 schools show a bimodal pattern, which repeats a bimodal distribution observed in the Year 3 Report for schools that were at that time (2005) in YIP 2. (In 2006, the distribution of those schools, then in YIP 3, coalesced to a single-mode distribution, presumably because the lower mode "caught up" with the rest of the cohort.) It appears that bimodal distributions may be a recurring characteristic of schools in the second year of implementation; they divide into "fast adopters" and "slow adopters." YIP 1 schools, a smaller number of schools altogether, seem to require additional time to build their level of implementation. YIPs 3, 4, and 5 seem most closely to resemble the All Schools distribution.

Figure 3.2: YIP = 5 –2007 Reading First Implementation Index (RFII), Distribution of Schools

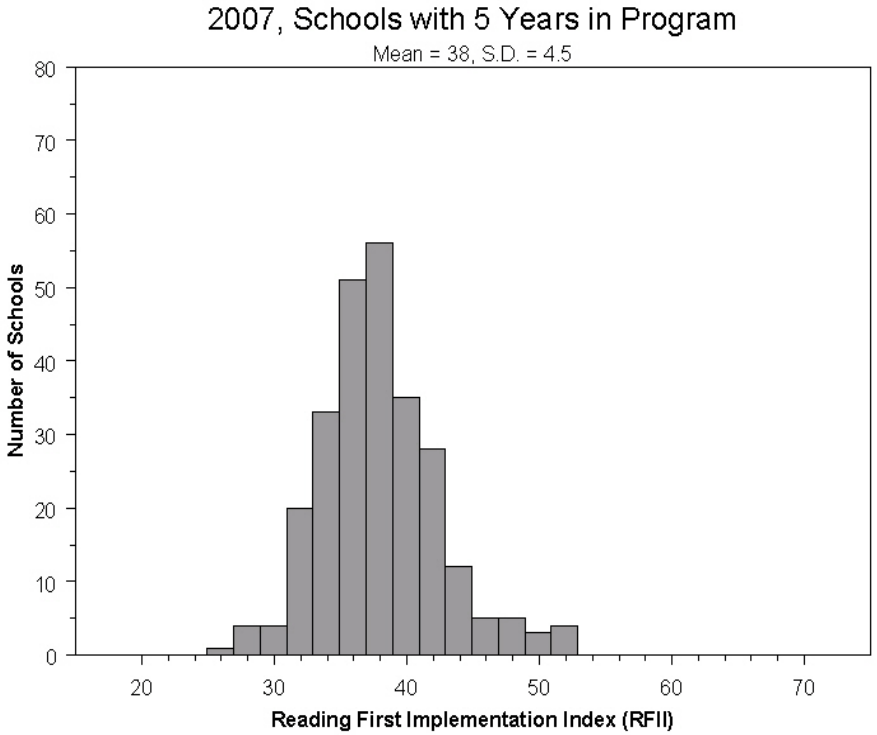


Figure 3.3: YIP = 4 –2007 Reading First Implementation Index (RFII), Distribution of Schools

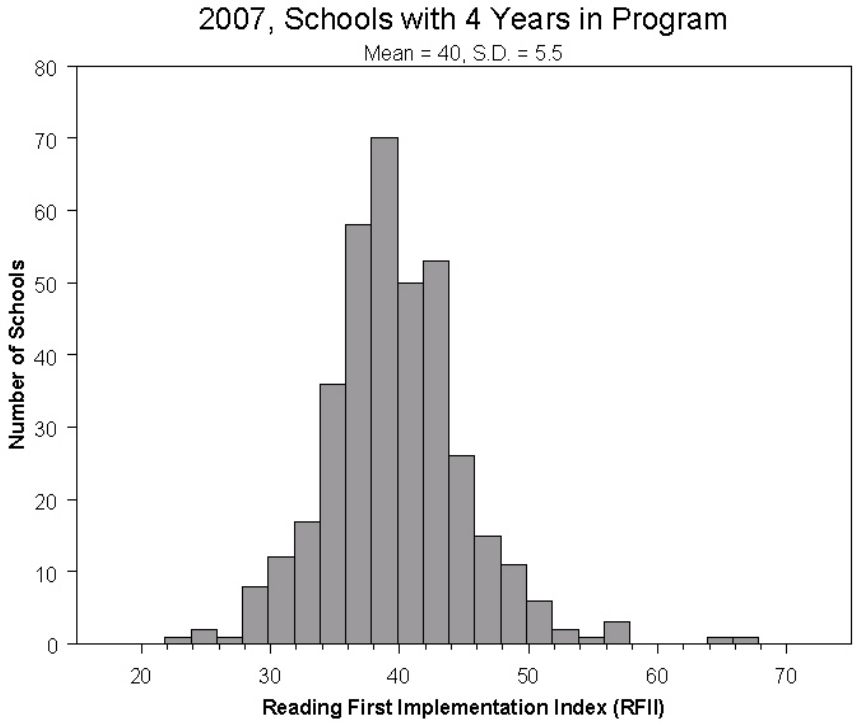


Figure 3.4: YIP = 3 –2007 Reading First Implementation Index (RFII), Distribution of Schools

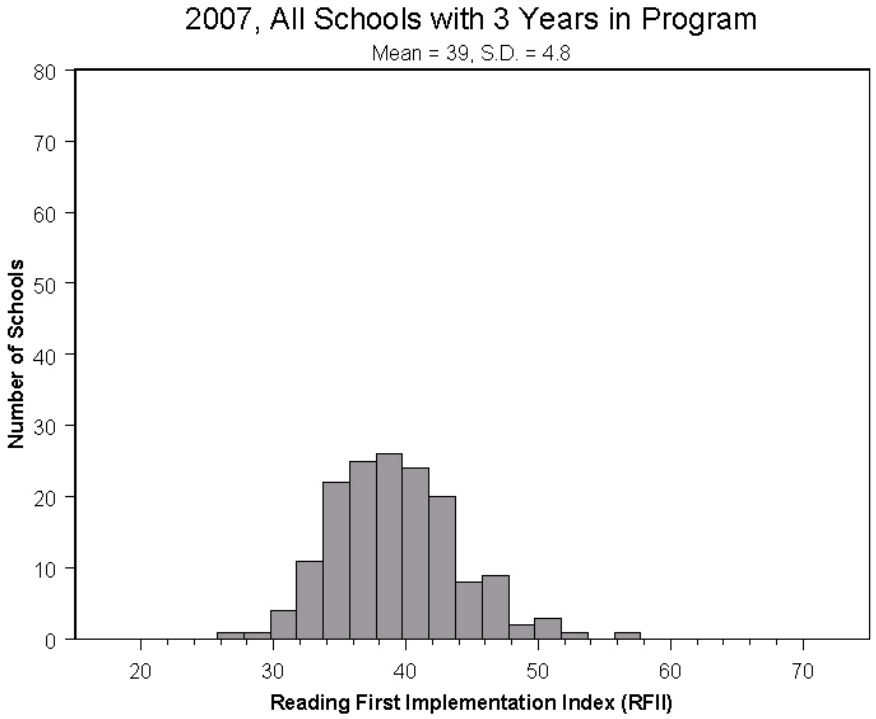


Figure 3.5: YIP = 2 –2007 Reading First Implementation Index (RFII), Distribution of Schools

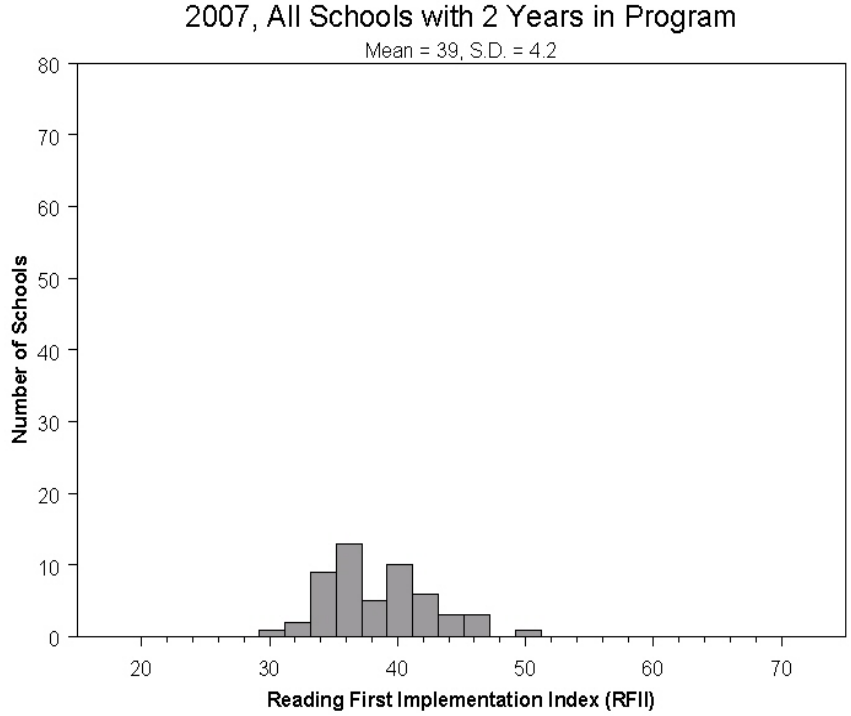
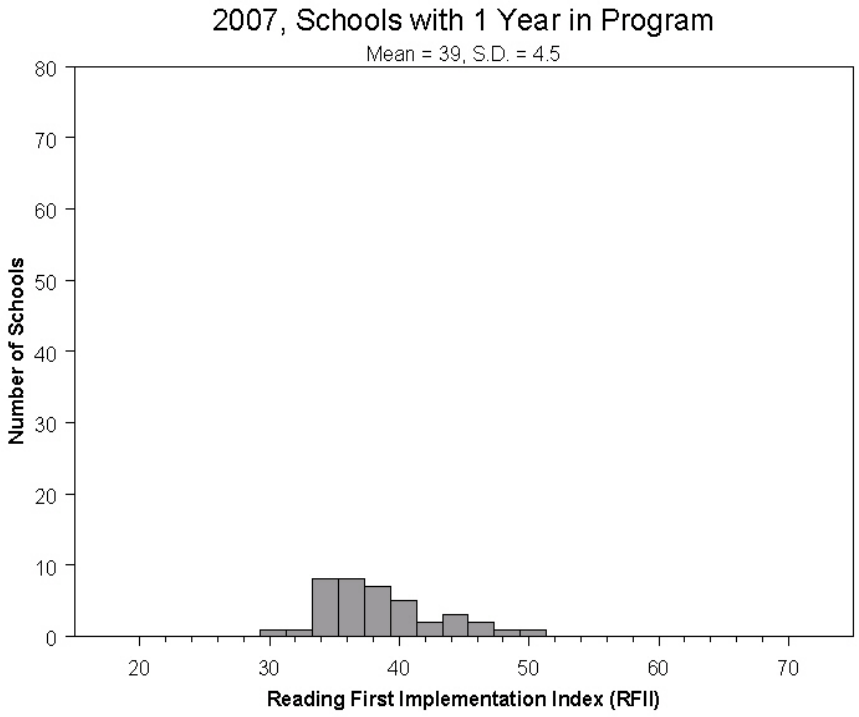


Figure 3.6: YIP = 1 –2007 Reading First Implementation Index (RFII), Distribution of Schools[5]



Dimensions of Implementation

Table 3.1 shows the dimensions derived in the RFII calculation process, their means and standard deviations for each year from 2004 to 2007, and the mean and standard deviation of the RFII for each of those years. The 2007 RFII had a mean of 39 and a standard deviation of 5. In the bottom row, we see the mean school RFIIs for 2004, 2005, 2006 and 2007 for all schools in the Reading First population, with standard deviations (labeled "Plus or Minus"). A list of 18 dimensions and the number of items that comprised each dimension are also reported along with their means and standard deviations for each year. 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.

For this report, we introduce a new column in Table 3.1 that shows the degree to which each dimension correlates with the 2007 RFII. (Correlations range from -1.00 to +1.00, where 0.00 means there is no relationship at all.) Thus, if a school wishes to increase its RFAI most expeditiously, it should focus on those dimensions which: a) are low relative to the state average; and b) have a high correlation with the

[5] Note that YIP 1 schools only have one year in the program and, therefore, averaging the 2006 and 2007 RFII was not possible. Only the 2007 RFII was calculated.

RFII. 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 (2007) = 885, Mean and "Plus or Minus" for Each Dimension, 2004-2007[1, 2, 3, 4]

			% of the time teachers rated their school "More than Adequate"								Correlation with
			2004		2005		2006		2007		
	Dimension	# Items, 2007	Mean	Plus or Minus	Mean	Plus or Minus	Mean	Plus or Minus	Mean	Plus or Minus	2007 RFII
1	Teacher Professional Development	9	38	*14*	36	*10*	35	*10*	34	*9*	0.48
2	Coach Professional Development	7	58	*22*	56	*21*	48	*22*	33	*21*	0.17
3	Principal Professional Development	3	48	*30*	46	*30*	57	*30*	56	*30*	0.16
4	**Teacher Coach Professional Development**	**11**	**40**	***16***	**37**	***11***	**35**	***10***	**34**	***10***	**0.45**
5	Evaluation of Professional Development	5	11	*6*	14	*7*	15	*9*	15	*8*	0.48
6	Implementation, Assurances	11	44	*18*	48	*16*	46	*16*	45	*17*	0.55
7	School Implementation, Materials	175	36	*10*	37	*9*	41	*11*	41	*10*	0.53
8	School Implementation, Instruction	28	34	*6*	36	*6*	40	*6*	40	*6*	0.79
9	**School Implementation Overall**	**210**	**39**	***7***	**40**	***6***	**43**	***7***	**43**	***7***	**0.96**
10	Coaching Implementation	32	46	*16*	48	*14*	50	*12*	49	*12*	0.68
11	Teacher Implementation	33	48	*5*	50	*5*	54	*5*	54	*5*	0.62
12	Teacher RF Understanding	17	27	*6*	29	*5*	30	*6*	30	*5*	0.34
13	Coach RF Understanding	17	36	*15*	39	*14*	38	*8*	39	*8*	0.42
14	Principal RF Understanding	17	17	*9*	19	*10*	20	*6*	20	*6*	0.40
15	**Overall RF Understanding**	**17**	**23**	***5***	**25**	***5***	**26**	***5***	**26**	***5***	**0.34**
16	Teacher RF Evaluation	4	14	*7*	14	*7*	16	*9*	15	*8*	0.58
17	Coach RF Evaluation	6	20	*18*	19	*18*	24	*21*	23	*20*	0.32
18	Principal RF Evaluation	6	23	*24*	24	*24*	23	*20*	20	*19*	0.24
19	**RF Implementation Index (RFII)**	**238**	**36**	***6***	**36**	***5***	**39**	***6***	**39**	***5***	**1.00**

[1] Dimensions 4, 9, and 15 are in bold because they are weighted contributors to Dimension 19, the RFII. The 2007 statistics are across 885 schools from the point of view of teachers for dimensions 1, 4, 5, 7, 8, 9, 10, 11, 12, 15, 16, and 19. Dimensions 2, 13, and 17 are from the point of view of coaches. Dimensions 3, 14, and 18 are from the point of view of principals. Dimension 6 is from the point of view of coaches and principals together. The 2004 statistics are across 628 schools; the 2005 statistics across 808 schools; and the 2006 statistics are across 856 schools.

[2] The statistics in the right column report the 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.

[3] The N-count of schools in this table does not exactly match those from all Reading First schools as reported in Chapters 1 and 2. The schools reported here are those whose teachers, coaches, and principals returned surveys.

[4] The phrase "Plus or Minus" refers to the average distance from the mean of all the measures that went into the mean. This is also known as the Standard Deviation.

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 cut-point – somewhat arbitrary – 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 and Chapter 2 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 have been 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.

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."[6] By that criterion, the histograms and Table 3.1 above reveal that schools are on average doing an "adequate" job of implementing the Reading First program, since the mean 2007 RFII of 39 is greater than 36 by half a standard deviation.

Examining the mean RFII over time, it appears that the index has risen. In 2004 and 2005, the mean RFII was 36 while in 2006 and 2007, it was 39. It stands to reason that program integrity would increase over time and the rise in the RFII statistic supports that conclusion.

[6] 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.

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.

Chapter 4: Importance of Program Elements

The purpose of this chapter is to examine the perceptions of Reading First participants regarding the implementation and sustainability of program elements. Chapter 3 reported the use of the Reading First Implementation Index (RFII) to gauge a school's level of implementation. The RFII provides a measure of overall implementation and is helpful in examining differences between high and low implementing schools with regard to student achievement. Here, we explore implementation issues further to better understand the nature and extent of implementation in Reading First schools. In this chapter, we use the dimensions of implementation derived from the Reading First surveys to determine differences between teachers, coaches and principals regarding specific aspects of implementation, and to examine the relative importance of aspects of the Reading First program. In addition, we examine the responses of participants to an open-ended question that provided teachers, coaches and principals opportunities to express their views of the sustainability of aspects of the Reading First program. Additionally, we examine the differences between high and low implementing schools in terms of the participants' perceptions of sustainable program elements.

This chapter yields the following key findings:

- Through Reading First, the state has developed a high level of expertise in personnel regarding research-based reading instruction, particularly in the cohorts of reading coaches who have participated. Coaches have participated in up to four years of advanced training and many have become expert trainers for the Reading First program.
- Overall, participants view the Reading First program as effective.
- Though implementation of the Reading First program is fairly strong in the areas of professional development institutes, curriculum use, adherence to a pacing schedule, collaborative teacher meetings, time allocation (kindergarten) and leadership support, these areas could still be improved.
- Implementation may need to be strengthened further in the areas of follow-up professional development, time allocation (grades 1-3), and assessment use.
- Coaches and principals report higher levels of adherence to program requirements than do teachers.
- Curriculum/materials, Reading First coaching, and collaborative teacher meetings were viewed as high in importance and worthy of sustaining upon program completion.

Why Implementation Matters

In Chapter 2, we show that higher implementing schools show higher rates of growth for students in terms of reading achievement. This finding replicates the significant and positive relationship between implementation and achievement demonstrated in this evaluation for Years 3 and 4. This is an important finding because implementation is an elusive construct. It is difficult to measure and it is not always possible to show direct correlations between implementation and program outcomes (Dane & Schneider, 1998; Ruiz-Primo, 2006). Additionally, it is difficult to determine how much implementation is needed or if certain elements of an educational program matter more than others (Noell, Gresham & Gansle, 2002).

Power, et al. (2005) suggest that measuring integrity of implementation involves examining the content, or how much of the program is implemented, and the process, including the quality of delivery and the participants' responsiveness. In this chapter, we examine these factors by analyzing responses of the direct participants in the Reading First program – teachers, coaches and principals. We use the Reading First survey results to explore participants' perceptions of the importance and sustainability of program elements.

Data Sources

For this chapter of the evaluation report, we examine items from the survey related to Reading First program components: professional development, curriculum use, time allocation, collaborative teacher meetings, leadership, and assessment. We examine the views of the three respondent groups regarding the importance and quality of these program components.

To gather additional information about the sustainability of program elements, teachers, coaches and principals had the opportunity to write in responses to an open-ended question, *"If Reading First funding was no longer available, what elements of the program would you want to keep in place at your school and why*?" The responses were compiled by respondent group in a text file and used in a qualitative analysis, described later in this chapter.

Perceptions of Program Elements

Professional Development

Extended and deep professional development is a cornerstone of the Reading First program nationwide. Deep knowledge of the science of reading instruction and research-based instructional strategies is key to improving the quality of reading instruction. The Reading First program requires that basic and advanced professional development be provided for program participants (See Chapter 1 for description). The Reading First survey asked participants about their participation in professional development and the extent to which it was helpful. Appendices A, B and C include the responses of participants by survey

items. Approximately 80% of teachers reported attending a Reading Professional Development Institute during Year 5 (See item B1 from Teacher Survey). Approximately 70% of coaches reported attending an institute (item B1 from Coach Survey) and all but 3% of principals reported attending principals' training at some point during the Reading First program (item B1 from Principal Survey). The figures for coaches may be an underestimate of actual attendance because the C-TAC reports that over half of the Reading First coaches served as institute instructors and may not have thought it accurate to report that they "attended" since they were more than participants. Additionally, the percentages reported in the table are calculated using the total number of survey responses and do not take into account missing responses for any single survey item.

It is interesting to note from the teacher surveys that 56% of the teachers and 33% of coaches participated in the Reading Professional Development Institutes on their own time while 21% of teachers and 34% of coaches attended during the instructional day (see item B2, Teacher and Coach Surveys). The institutes were considered effective in preparing teachers to teach their adopted reading curriculum; 52% of the teachers reported (item B4, Teacher Survey) that the institutes prepared them "adequately" and 16% reported "very well." The coach survey (item B4, Coach Survey) indicated that 37% felt "adequately" prepared and 28% were prepared "very well." Again, missing responses may have skewed the percentages somewhat due to the use of the total number of surveys in calculating the percentages.

Follow-up professional development occurred during 80 hours of sessions provided at the district or school level. However, only 58% of teachers reported participating in the full 80 hours, and 64% participated in 40 hours or more (item B5, Teacher Survey). Though 58% reporting full participation is somewhat low, it is important to note that approximately 21% of teachers did not respond to this item and this may not reflect actual participation. Teachers were asked to indicate how well the follow-up professional development prepared them if they had participated in at least 39 hours (item B6, Teacher Survey). Of those who responded, 40% said it had prepared them "adequately" and 18% "very well." Approximately 58% of coaches participated in at least three days of follow-up training and 72% of principals participated in all or part of the 40-hour principal follow-up.

Curriculum Use

A significant portion of the survey included questions regarding 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. One item in particular, F3 from the teacher survey, demonstrates the extent of curriculum use; Table 4.1 illustrates teachers' reported use of their adopted curriculum.

Table 4.1: Teacher Survey Results for Curriculum Use

What percentage of your total reading/language arts instruction relies on materials from your district's adopted program?	N	% of total
a. 0% - 19%	54	0
b. 20% - 39%	165	1
c. 40% - 59%	699	4
d. 60% - 79%	2,318	13
e. 80% - 100%	13,884	80

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

Time Allocation

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. Here we see 96% of kindergarten teachers reporting the use of 60 minutes or more, a high level of compliance. For grades 1-3, we see only 66% reporting from 140 to 180 or more minutes, and 78% reporting 120 minutes or greater. For grades 1-3, time allocation is an area in need of strengthening.

Table 4.2: Teacher Survey Results for Time Allocation

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?	Kindergarten Teachers		Grades 1-3 Teachers	
	N	%	N	%
a. Less than 20 minutes	9	0	17	0
b. 20-39 minutes	23	1	38	0
c. 40-59 minutes	111	3	98	1
d. 60-79 minutes	517	13	370	3
e. 80-99 minutes	829	21	641	5
f. 100-119 minutes	475	12	696	5
g. 120-139 minutes	822	21	2413	18
h. 140-159 minutes	298	7	2546	19
i. 160-179 minutes	182	5	1452	11
j. 180 minutes or more	681	17	4747	36

Note: This table excludes 164 teachers of Kindergarten/Grade 1 split grade combination classes and 6 teachers who did not specify a grade.

Collaborative Teacher Meetings

The Reading First program promotes twice-monthly teacher planning meetings that focus on analyzing student data from ongoing assessments, understanding the curriculum materials, improving instructional strategies, and assisting struggling readers. Teachers reported a variety of topics, with the majority focusing on instructional strategies, assessment results, intervention strategies, and teaching the adopted program. 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. Though there is agreement across raters, it appears that more teachers perceive a lack of time to plan collaboratively than do coaches or principals. Additionally, a higher percentage of principals perceived that there were weekly opportunities for teachers to plan together than coaches or teachers.

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

How often does the school leadership provide time for teachers to plan collaboratively?	**Teachers** **%**	**Coaches** **%**	**Principals** **%**
a. Hardly ever	18	4	1
b. Monthly	28	21	15
c. Twice monthly	22	36	36
d. Weekly	31	37	45
e. Daily	1	0	1

Pacing of Instruction

The Reading First program requires districts to develop pacing plans or guides for moving students 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 the guides are followed, teachers should cover the entire year's curriculum. Participants were asked about whether they had a pacing schedule, how closely they adhered to it, and the support provided by coaches to reinforce it. Table 4.4 presents results from teachers (Question D1), coaches (E1) and principals (E1) on relevant pacing questions. The qualitative analysis included later in this report elaborates on participants' perceptions of the use of pacing plans.

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

Does your school have a pacing schedule?	Teachers %	Coaches %	Principals %
My school does not have a pacing schedule	2	1	0
My school has a pacing schedule based only on the assessment schedule	30	18	14
My school has a pacing schedule that identifies lessons on a daily or weekly schedule and when to give assessments	67	80	85

Additional information about pacing schedules was provided by teachers (Question F4). Only 1% said they do not follow the pacing schedule, 5% said they "keep in mind" where they should be in the schedule, 24% said they follow it approximately and 69% said they follow it closely. Coach responses to the same question about the teachers in their school were very similar. Coaches were asked about their roles in reinforcing the pacing schedule (F8). Only 4% said they do not check on the pacing, 29% said they "occasionally" check, and 65% said they notice when teachers fall behind and help them to catch up.

Leadership

Improving the ability of the school leadership to support research-based reading instruction is a key element of the Reading First program. Teachers, coaches and principals were asked a similar question about the level of support provided by the school principal (Questions D11, E13, E12 respectively). Table 4.5 lists the results by group. In this table, we see a discrepancy in the perceptions across respondent groups, with teachers reporting a much smaller percentage in the "more than adequate" support category than coaches or principals. Principals viewed their level of support more positively than did coaches and teachers. Teachers and coaches had much higher percentages in the "little or no support" category than did principals. This is consistent with findings reported in Chapter 3 indicating that principals and coaches tended to rate higher than teachers in general.

Table 4.5: Percentages of Teachers, Coaches, and Principals Regarding Level of Administrator Support

In general, what level of support are you getting from your principal related to your adopted reading/language arts program? **(For principals – What level of support do you provide the teachers and coaches?)**	**Teachers** **%**	**Coaches** **%**	**Principals** **%**
little or no support	17	11	1
adequate support	55	32	27
more than adequate support	27	56	68

Assessment

Reading First schools are required to use the 6-8 Week Skills Assessments to monitor students' progress and adjust instruction. The assessment results are reported to schools and districts, and only the End of Year tests are reported to state personnel. They are also the focus of collaborative planning discussions at the school sites. The survey includes numerous questions about the use of assessments. Here we focus on one question. Table 4.6 presents the teacher, coach and principal responses regarding what assessments teachers use (Questions F6, G6 and G6, respectively). Respondents were directed to check all that apply, so the percentages will not total 100% in any column. Most teachers use multiple sources of information for making instructional decisions, so it is not surprising that teachers would report using additional assessments beyond those required. However, it is interesting that schools would report nonuse of required assessments that must be reported for Reading First. Table 4.6 shows the responses to Item D, use of the 6 – 8 week Skills Assessments, the assessments required in Reading First schools statewide. Though coaches (95%) and principals (91%) largely perceived adherence to this requirement, only 77% of teachers reported using the assessments. However, this figure is inconsistent with the 94% of teachers who indicated that they use the assessments for specific purposes in the following question. This discrepancy raises a possible area of concern regarding either teachers' actual or perceived use of assessments.

Table 4.6: Percentages of Teachers, Coaches and Principals Regarding Assessments Used

If you assess your students every six to eight weeks, which assessments do you (or the teachers in your school) use?	Teachers %	Coaches %	Principals %
a. I (teachers) do not assess students every 6 – 8 weeks	3	0	0
b. I (teachers) use teacher-developed assessments	20	17	12
c. I (teachers) use assessments that come with the adopted program	42	41	48
d. I (teachers) use the 6 – 8 Weeks Skills Assessments	77	95	91
e. I (teachers) use district-developed assessments	30	23	37
f. I (teachers) use assessments other than those above	14	9	15

Overall Effectiveness

Participants were asked to rate the overall effectiveness of the district's adopted reading/language arts program in item I1 from each survey. There was surprising agreement among teachers, coaches, and principals in their opinions. Teacher, coach and principal percentages were, respectively:

(e) Poor (3%, 0%, 0%)
(f) Fair (19%, 13%, 8%)
(g) Good (56%, 59%, 61%)
(h) Excellent (21%, 26%, 29%)

Participants' Perceptions of the Value of Program Elements

Sustainability is an educational concept of significant concern in any large-scale educational improvement effort. Sustainable practices are those that are viewed as critically important by program participants. Foorman and Moats (2004) suggested that sustainability of professional development in reading instruction is more likely to be achieved when teachers perceive that the program is positively impacting student achievement and their own knowledge and skills. Additionally, sustainability is more likely when there is ongoing support and "transactional" support structures such as peer networking, coaching and collaboration that provides opportunities for teachers to apply concepts learned specifically to the curriculum, instructional strategies, and data analysis (Darling-Hammond & McLaughlin, 1995; Desimone, Porter, Garet, Suk Yoon, & Birman, 2002; Novick, 1996).

In this section, we use qualitative research methodology to examine findings from the open-ended question included on the survey, "*If Reading First funding was no longer available, what elements of the program would you want to keep in place at your school and why*?" It is important to examine the perceptions of those on the "front line" of implementation – teachers, coaches and principals at Reading

First schools. They provide insight into the extent to which program elements have been incorporated into the day-to-day practices and routines at the school. The open-ended question regarding sustainable program elements provides an opportunity for participants to voice their opinions about the elements of Reading First that are important enough to sustain.

In this analysis, we first examine the perceptions of teachers, coaches and principals as reported in the open-ended question. We compare the relative perceived importance across respondent groups. Then, we examine differences in perceptions across high implementing and low implementing schools, as determined by the RFII (Reading First Implementation Index, see Chapter 3).

A Note about Qualitative Analysis

The qualitative findings reported in this and subsequent chapters differ from those reported in previous chapters of this report because the data sources are narrative in nature, as opposed to quantifiable data used in other analyses. The advantage of qualitative research is to get an "insider's view" of a phenomenon and "give voice" to participants in that phenomenon (Brantlinger, Jimenez, Klingner, Pugach & Richardson, 2005). Such qualitative data offer rich, descriptive characterizations of participants' perceptions that provide an elaboration on findings from quantitative statistical analyses. When used in conjunction with quantitative statistical analysis, qualitative research can provide a deeper explanation of statistical results. Of the 17,261 teacher surveys collected, 13,244 wrote narrative responses to this question, or 76.7%. Of the 1,028 coach surveys collected, there were 944 narrative comments submitted, or 91.8%. Of the 1,073 principal surveys collected, there were 989 comments submitted, or 92.2%. This is a high response rate for the open-ended question format.

Limitations of this study should be noted. Though qualitative research may provide in-depth insight into phenomena and why they occur among participants, results are viewed as inconclusive. Generalizability of findings beyond the respondents is somewhat limited. Being able to generalize findings requires knowing specific information about the sample and having some assurance that the sample is representative of a particular group. Furthermore, it is difficult to interpret the weight or meaningfulness of findings without the ability to quantify them. Reoccurrences of findings certainly gives some insight into their importance, but they are not weighted or counted as in quantitative methods. In this evaluation study, the qualitative data were examined for converging evidence of sustainable program elements across multiple perspectives (teachers, reading coaches, and principals) and contexts (high and low implementing school sites).

There is an important difference between the closed-ended and open-ended portions of the surveys. In the closed-ended items, respondents were given an array of response options and they selected one or more that best matched their opinions. In the open-ended question format, participants were not prompted to

respond in any certain way and options for responses were not provided. Respondents wrote spontaneously, giving their opinions about a topic. Therefore, when categories or themes of responses reoccur in the data, they take on added meaning as substantial proportions of respondents independently and spontaneously chose to write similar responses. For example, in one category of responses discussed below, approximately one-third of respondents indicated that the curriculum or materials used in Reading First is an element that they would want to continue. This does not mean that two-thirds did value the curriculum or materials. It merely means that, for one-third of the respondents, it was prevalent in their minds at the time of responding and they thought it important enough to write about. It is likely that some portion of the two-thirds who did not elect to write about curriculum or materials would agree that this element is important if asked in a closed-ended format with that option given as a choice to select. This is important to bear in mind when interpreting the percentages of responses that fell within given categories.

Data Analysis Methodology

For this qualitative analysis, the text file extracted from the online survey was subjected to analysis using a qualitative software package. Data reduction involved coding segments, or "chunks" of data that contained meaning related to the study purpose. According to Miles and Huberman (1994), data reduction "is a form of analysis that sharpens, sorts, focuses, discards, and organizes data in such as way that 'final' conclusions can be drawn and verified (p. 11)." For this study, one researcher read through several pages of comments to establish an initial set of codes, or categories of meaning represented in the data, that was consistent with the guiding questions. The researcher and a research assistant familiar with the project then met to discuss code descriptions and coded approximately 5% of the teacher, coach and principal files together. The coding process consisted of assigning one or more codes to each individual response. When there was no obvious existing code for a segment, the coder used a category of "Other." If a recurring pattern of "Other" responses was evident, the lead researcher and coder discussed the possibility of creating a new code and then went back to recapture those already coded in the "Other" category. Periodic checks were conducted to ensure there was consistency of coding. The coder and lead researcher met frequently to create new codes that emerged or alter code definitions. Following the completion of coding, the "Other" category was examined for any recurring theme that might be pulled out and recoded.

Using a grounded theory approach and a recursive coding and analysis process, all segments were coded and categorized using a constant comparison method (Strauss & Corbin, 1990; 1994). Reliability was addressed in this study by the use of frequent conferencing among coders and researchers. Validity was addressed through an audit trail of the analysis process, the examination of confirming and disconfirming evidence, and the high response rate. Once all segments were coded, the software package facilitated refinement of codes and categories to find recurrent patterns and discern themes and their relative strength

(Brent, Slusarz & Thompson, 2002). Segments of data that were irrelevant to the question were discarded. The software includes tools for searching, categorizing, and sorting data as well as hypothesis testing and theory building.

Relative Importance of Program Elements

The codes (or categories of responses) were sorted according to the frequency with which they occurred and are listed in rank order in Table 4.7. Table 4.8 provides a description of each category as well as examples of responses for each. Rankings are listed for the whole data set combined and then for teachers, coaches and principals. This allows the reader to compare the relative importance of program elements by participant group. Note that this table does not depict a rank ordering by teachers, coaches or principals. Rather, it depicts the relative frequency with which the codes occurred and should not be interpreted as ratings. Table 4.7 also lists the percentage of the total respondents that commented within each category. The total number of written responses provided was used to calculate these percentages, not the total number of surveys received by each group. Note that the percentages will not total 100% because in many cases, comments were assigned multiple codes. This occurred when a response included multiple ideas or concepts, or when a response could be interpreted as falling within more than one code. Additionally, some responses were not coded at all because they were irrelevant to the question and the purpose of this part of the study. For example, if a teacher responded that she particularly likes the math curriculum they are using, it was considered an irrelevant comment and not coded.

In this table, we see that the order of frequencies is very similar for the column of All and Teacher respondents. This is because the teachers made up the largest proportion of the response pool. However, there is some variability in the Coach and Principal listings. The highest frequency codes across respondent groups, Curriculum/Materials, Coach/Coaching, and Collaboration/Lesson Studies occurred in the top four frequency listings for all three respondent groups. It is reasonable to assume that these elements, which are considered key to Reading First implementation, were foremost on participants' minds as they responded to this question regarding the sustaining of program elements. It is possible that these elements have become an important part of the routine and fabric of Reading First schools since significant proportions of all three respondent groups spontaneously elected to write about these elements when prompted with this open-ended question. Likewise, Assessment and Data Analysis (5^{th} for teachers, 3^{rd} for coaches and 4^{th} for principals) and Professional Development (6^{th}, 5^{th} and 5^{th}) were high in relative importance. The Negative Comments category captured a variety of comments that expressed a negative view of sustaining Reading First and it is discussed below. Note that the All Aspects category was the 4^{th} highest according to teachers, and 6^{th} and 7^{th} for coaches and principals, respectively.

Table 4.7: Rank Order and Percentages of Responses for Categories

Question: "If Reading First funding was no longer available, what elements of the program would you want to keep in place at your school and why?"

Response Category (Code)	All N = 15,177		Teachers N = 13,244		Coaches N = 944		Principals N = 989	
	Rank	**%**	**Rank**	**%**	**Rank**	**%**	**Rank**	**%**
Curriculum/ Materials	1	34.0	1	35.6	4	21.2	2	24.3
Coach/ Coaching	2	25.8	2	23.4	2	34.3	1	50.6
Collaboration/ Lesson Studies	3	14.9	3	12.6	1	38.8	3	23.3
All Aspects of the Program	4	10.8	4	10.2	6	16.3	7	13.7
Assessment and Data Analysis	5	10.5	5	8.0	3	31.3	4	22.8
Professional Development	6	7.9	6	6.1	5	21.0	5	19.5
Pacing Plan or Guide	7	5.5	7	4.5	7	12.9	6	14.7
Negative Comments	8	3.8	8	4.0	10	4.6	14	2.2
Small Group Instruction/ Universal Access	9	2.6	9	2.3	11	4.3	9	5.4
Instructional Strategies	10	2.5	10	1.7	8	9.7	8	6.7
After School Program	11	1.4	11	1.5	13	1.5	15	0.4
Substitute Days/ Release Time	12	1.3	12	1.0	9	4.7	11	2.7
Structured Teacher Planning Time	13	1.0	13	0.9	14	1.4	12	1.3
Reading/Language Arts Time Block	14	0.6	14	0.4	12	2.4	10	3.0
English Learner Guidelines	15	0.5	15	0.2	15	1.4	13	0.8

Response Descriptions

For each code, or response category, in Table 4.8 below, a brief definition of the highest frequency codes is provided along with major finding and representative comments for each.

Table 4.8: Code Descriptions, Rationales, and Representative Comments for High-Frequency Categories

Question: "If Reading First funding was no longer available, what elements of the program would you want to keep in place at your school and why?"

Code Description	Reasons and Representative Comments
Curriculum/ Materials Indicates the reading curriculum in general, or specific materials	***General*** "I think most teachers would continue to use the adopted curriculum and would still benefit from supplemental materials (Coach)" "The core program direct instruction because I feel that the content being taught is valuable for academic growth and development (Teacher)" "As long as we have the same reading program, I would implement it the same way, with or without Reading First (Teacher)" "The systematic approach to phonemic awareness and the presentation of skills and strategies in reading comprehension. The students are learning to decode and encode. They are learning strategies and skills to help them become good readers (Teacher)" "Consistent instruction of the five reading components (Principal)" ***Specific Curriculum or Materials*** "The Houghton Mifflin reading program because it does a great job of covering the majority of standards using strategies that are effectively teaching students reading skills (Principal)" "I love the sound/spelling cards activities that have been very powerful. In fact, the green section and dictation from the blue section has been the most valuable to me. I can see definite improvement in their ability to read multi-syllabic words and their spelling is more independent. The vocabulary strategies open up abilities to solve comprehension questions on their own (Teacher)" "Phonemic awareness for kindergarten is the most important component of teaching kindergarten (Teacher)" "I would love to keep the anthology sections and leveled reading books to teach comprehension (Teacher)" ***Schoolwide/ Districtwide Curriculum*** "The program that all teachers teach with the same fidelity (Teacher)" "The reading program and its tools. They have proven to keep us all on the same page. We are able to teach and monitor (Teacher)" "The use of a systematic approach to phonics instruction. Using the same set of sound/spelling cards throughout the school seems to have greatly assisted the students' knowledge of sound-symbol correspondence (Teacher)"

	"It has taken a long time to change teachers' opinions of the language arts program. Now they are beginning to understand the basis in research and that when taught explicitly, students will achieve…The program allows students to keep pace academically with other students while learning English (Coach)"
Coach/ Coaching Indicates the coaching model or their specific reading coach	***General Coaching Model*** "I believe that the coaching provides the teachers with a sounding board for their instruction and another set of eyes that validates what they are doing in the classroom (Coach)" "The coach is a huge asset to the implementation of not just the program but of literacy itself (Principal)" ***Specific Aspects of Coaching*** "Having a coach to bring the most current research to the schools and support novice teachers in their implementation of the program (Coach)" "I would want to keep our reading coaches who provide specific support to teachers to improve practice (Principal)" "I would love for my district to fund a reading coach to continue attending workshops and conferences for us and sharing the info with us (Teacher)" "The coach helps with assessments and demonstration lessons (Teacher)" "The intervention guidance from coaches (Teacher)" ***Coaching Linked with Collaboration*** "A full-time literacy coach for K-5 to maintain and plan collaboration for teachers (Coach)" "When I was able to meet with the Reading First coach and my grade level at the same time, I felt this was the most beneficial… I felt reassured that I was teaching the district adopted reading/language arts program appropriately (Teacher)" ***Coaching Linked with Accountability/ Monitoring*** "Coaching—accountability with support (Coach)"

Collaboration/ Lesson Studies May specify collaboration in general or specifically mention lesson studies, grade-level meetings, data analysis sessions	"The teachers have learned that they all have similar problems in the classroom and the same needs. They have learned to work together and rely on each other for support to become better teachers (Coach)" "I would like to keep the data analysis meetings. It is helpful to sit as a grade level with the reading coach to look at scores and discuss the different teaching strategies that were used (Teacher)" "The teachers at the school are open to reflection and refining their practices so Lesson Study is well used for them (Coach)" "Collaborative grade-level meetings are the best part and most valuable for the consistency of the program. Keeps us all together as a team! (Coach)" "Data meetings. The teachers learn how to guide their instruction based on data (Principal)" "We would continue to have lesson study for professional development opportunities even if we didn't receive substitute funding through Reading First (Principal)" "We all have a shared common language through our collaboration (Principal)"
All Aspects of the Program States that all aspects of Reading First are valuable	"I would like for the whole program to be funded at our school because every year our class population is different and our needs are different. We need our reading coach! She is what has made teaching the program more realistic. I sometimes feel tasks are impossible but when she comes in and models a lesson, I feel empowered or more confident to follow through (Teacher)" "All aspects because they seem helpful in guiding my instruction and student progress (Teacher)" "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 (Coach)" "All of the components because they are best practices that should be used all the time (Principal)" "All of the program should continue regardless of funding. It is research-based. It works and teachers do need to be held accountable (Coach)"

Assessment and Data Analysis States the assessment tools, assessment system, processes, progress monitoring, use of data, common data across school/district	"Ongoing assessments to guide instruction and better meet student's needs (Coach)" "We would continue to analyze assessments in order to adjust student's differentiated instruction according to their progress and further, adjust whole class teaching strategies (Principal)" "I would keep the SCOE assessments which the students take every six weeks because they allow me to monitor the progress of the students helping me to plan strategies and adapt instruction based on their academic needs (Teacher)" "I would want to keep the 6-8 week assessments and the OARS data system that gives us such easy access to data results in a color-coded formatting individual student information and grade level comparisons etc. I would want to keep our weekly grade level meetings and Data Conferences that occur every 6-8 week period (Coach)"
Professional Development May specify the Reading First or state-funded professional development, or specify the 80-hour follow-up	"Training is crucial to ensuring that our instruction does not become stagnant but that it continues to evolve and grow (Teacher)" "Professional Development is the most important in ensuring the effective implementation of any reading program (Principal)" "For me, the most important aspect of Reading First funding was the availability of thorough training that provided researched background information (Teacher)" "On-site professional development is crucial to maintaining the momentum of professional growth that supports and augments student learning. It would be a backwards move to eliminate this valuable training (Coach)" "I would like for the Reading First Action Seminars to continue. They have been instrumental in moving my school forward and in helping teachers reflect on their practice and on student data (Principal)" "Reading First is extremely beneficial in providing training to teachers. It is essential that they continue to build their knowledge about the research and fundamentals behind teaching reading (Coach)"

Pacing Plan or Guide Specifies the pacing plan as an important component to maintain	"I believe the pacing charts are beneficial in assuring that sufficient time is allowed for students to learn and practice new skills. They also keep teachers on a schedule which encourages collaboration and grade level planning (Coach)" "I would like to keep the pacing schedule so that everyone is held accountable and there is consistency throughout the grades (Teacher)" "The pacing guide keeps everyone focused. Daily pacing maximizes time and quality of instruction in each class (Coach)" "I would want to maintain the pacing schedule to ensure students receive the breadth of instruction required to prepare them for the subsequent year (Principal)" "The consistency of the program is important in a district like ours where kids move school to school. The pacing plan allows us to be on a schedule which helps those kids not miss any important details from one move to the next (Teacher)" "The pacing guide had brought consistency and more universal access to our students (Coach)"
Small Group Instruction/ Universal Access States that instructing students in small groups is to be maintained; may mention program-specific small group structures	"I would like to keep the direct instruction because I can reach more students and reach the students who need extra help when I pull them for U.A. (Teacher)" "Differentiation during designated UA time ensures that the needs of all learners are addressed (Principal)" "I like the UA time. It provides time for me to offers support to small groups of students while allowing other students to practice needed skills in a manner different than the everyday lessons (Teacher)" "I would keep the use of Universal Access Time because this concentrated block of time is crucial to our intensive strategic and English language learners in being pre-taught and re-taught HM lessons. UAT also allows the teacher to work with the mentally gifted students in helping excel to their fullest potential (Coach)"

Instructional Strategies Specifies that the instructional approach or methodology learned in the professional development and used in the curriculum is to be maintained	"I would like to continue with the instruction of the comprehension strategies and skills (Coach)" "The amount of time spent on instruction has led to great gains (Principal)" "Blending and dictation strategies are helpful in assisting students to access text (Teacher)" "I would keep the sound/spelling card strategies, the comprehension and phonetic skills instruction and the IWT (small group time) because these all help to improve instruction and lead to reinforcing the state standards (Teacher)"
After School Program	"I would keep the Special Ed. Reduction Program, After School Program. I believe it really helps the children grasp concepts in small group settings (Principal)" "I would want to continue tutoring after and before school, even if I wasn't being paid in order to help my students' be successful (Teacher)" "I would like to keep interventions going- this has helped cut down the number of students referred to special ed (Coach)" "The before and after-school interventions because the students are all at the same level and get a chance to delve into concepts that as a teacher, I don't have time to devote to during class (Teacher)"
Substitute Days/ Release Time	"I love the monthly grade level release days we are able to have because of the Reading First funding. It gives me time to really collaborate with my colleagues to plan how to effectively reach our students (Teacher)" "Our Reading First sub days provided me the time I needed to really dig in to the core program at each grade level (Coach)" "Sub days have been beneficial for observing colleagues. I believe that novice and veteran teachers alike can learn from observing same-grade as well as cross-grade teachers both on and off site. Students benefit from different experiences so teachers need to continue to learn new and specific strategies to best support their learning styles (Teacher)" "The substitute days allocated have allowed for our teachers to participate in on-site professional development sessions that have resulted in collaboration amongst grade levels as they continuously refined their craft (Principal)"

Structured Teacher Planning Time	"I would like to continue with STPT (structured teacher planning time) as it necessitates reflecting on one's own teaching practices and it also guides instruction. STPTs provide a safe place to ask for help and share successes and challenges (Coach)" "STPT has been a value because it allows for a structured forum to analyze data and discuss instructional strategies. It truly assists in refining instructional delivery (Principal)" "All day planning is definitely an important piece of our planning and implementation of the reading program at our school. It allows us to work as a grade level to plan enriched lessons to implement the reading program which in turn facilitate student learning (Teacher)"
Reading/Language Arts Time Block	"I would like to keep my Reading Block the same because I think that it provides everything that students need to decode words and learn spelling patterns (Teacher)" "I believe the number of minutes per day devoted to Language Arts instruction and the inclusion of Universal Access are key (Coach)"
English Learner Guidelines	"The Extra Support and ELL Handbooks are helpful in reading instruction (Teacher)" "We are also reclassifying a greater percentage of ELs since we've made the EL Handbook the program during ELD (Principal)" "The Extra Support Handbook and ELL Handbook are invaluable! (Coach)"
Negative Comments Negative comments relevant to the question.	Note: Many of the negative comments were vague and did not fall into subcategories. These subcategories represent relatively small numbers of respondents. They are included here only to demonstrate contrasts to the positive comments included in the other categories in this table. ***Writing*** "I would keep most of the components. The only issue I have is with the writing. The writing section of Open Court 2000 is the weakest part (Coach)" "I would keep all but the writing component because I feel it is not well developed (Teacher)" ***Assessment*** "I would keep all the elements, but would like to reduce the number of assessments to allow more time for instruction (Principal)" "There is no need to assess students so often (Teacher)"

	English Learners/Waivered Classrooms (These topics are discussed in more depth in Chapters 6 and 7) "I think we may be doing our students a disservice by providing an unclear transition plan into English Only classrooms in 4th grade (Teacher)" "Teachers are not satisfied with the fact that the Spanish reading program is a translation of the English program and the research used for this program was the same used for the English program (Coach)" ***General Frustration*** "I feel that teachers should have been given Open Court training prior to ever having been expected to teach the program. This was the district's decision (Teacher)" "The time requirements have made it nearly impossible to teach other subjects such as science and social studies. (Teacher)" ***Curriculum Control*** Some teachers expressed a desire to teach what they want rather than have to follow a curriculum and pacing plan: "Reading First detracts from my teaching. I have to spend an inordinate amount of time testing and 'teaching to the test' (Teacher)" "Reading First comes attached with extra teacher responsibilities that take away from my classroom planning and teaching what I know my students need (Teacher)" "Reading First is the Reading Police. Go away now! (Teacher)"

Conclusions

To answer the question, what program elements are of most importance, we have examined data from the survey results and narrative responses from an open-ended question. Examining implementation provides an opportunity to determine what specific strengths exist in the program as well as what areas may need improvement.

Implementation of the Reading First program is fairly strong in the areas of professional development institutes, curriculum use, adherence to a pacing schedule, collaborative teacher meetings, time allocation (kindergarten) and leadership support; these areas could still be improved. Implementation may need to be strengthened further in the areas of follow-up professional development, time allocation (grades 1-3), and assessment use. Coaches and principals report higher levels of adherence to program requirements than do teachers.

Overall, program participants view the Reading First program as effective in improving reading achievement at their schools. Curriculum/materials, Reading First coaching, and collaborative teacher meetings were viewed by nearly 75% of participants (teachers, coaches and principals) as high in importance and sustainable elements. Participants' comments provide further description of the importance and sustainability of program elements.

References

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

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

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.

Darling-Hammond, L., & McLaughlin, M. (1995, April). Policies that support professional development in an era of reform. *Phi Delta Kappan*, *76*(8), 597-604.

Desimone, L., Porter, A., Birman, B., Garet, M., & Yoon, K. (2002, October). How Do District Management and Implementation Strategies Relate to the Quality of the Professional Development That Districts Provide to Teachers?. *Teachers College Record*, *104*(7), 1265-1312.

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.

Novick, R. (1996, January 1). Actual Schools, Possible Practices: New Directions in Professional Development. *Education Policy Analysis Archives*, *4*(14), 1-15.

Power, T., Blom-Hoffman, J., Clarke, A., Riley-Tillman, T., Kelleher, C., & Manz, P. (2005, May). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools*, *42*(5), 495-507

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.

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

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

Chapter 5: Evaluation of Reading First Coaching

The purpose of this chapter is to examine the perceptions of Reading First participants regarding the coaching model used in Reading First schools. LEAs receiving funding have had the option of using part of the funding to hire, train and support reading (or literacy) coaches, out-of-the-classroom teachers assigned to provide classroom-level support for program implementation. Extensive resources have been allocated to training and supporting a vast statewide network of coaches. This chapter addresses the question, "What elements of Reading First coaching are beneficial and why?"

In this chapter, we use information from the Reading First surveys to examine the roles of coaches in depth. We use selected items from the survey to determine teachers', coaches' and principals' views of the role and function of coaches. In addition, we examine the responses of participants to an open-ended question that provided teachers, coaches and principals opportunities to express their views of the aspects of coaching that are beneficial.

This chapter yields the following key findings:

- Reading or literacy coaches are an integral part of the Reading First program in California. They are highly valued by program participants.
- Coaches serve important functions in supporting implementation of the Reading First program and maintaining a school's focus on improving student achievement.
- The Reading First initiative has provided extensive training and support to coaches, an effort that has built capacity at the district, school and classroom levels. Coaches have reached a high level of expertise to the point that many are now qualified to provide training at Reading First institutes.
- The most important functions served by coaches in California are providing demonstration lessons, serving as a source of resource and support for teachers, and facilitating collaboration focused on student achievement and fidelity of implementation.
- Most coaches have ready access to classrooms to provide support with curriculum, instruction, and assessment.

Research on Coaching

The concept of coaching has emerged from criticism of traditional professional development that offers workshop-type sessions provided by an expert with little or no follow-up support (Darling-Hammond & McLaughlin, 1995; Novick, 1996). The research on how to expand and sustain professional development is just beginning to document coaching as an important element. Garet and colleagues (2001) identified

aspects of professional development that significantly correlated with change in teachers' knowledge and practices. They posit that effective professional development should focuses on specific content knowledge; provide ongoing training at the school site; integrate training with the daily work of teachers; promote the collective participation of teachers; align with instructional goals, instructional practices and local standards; and provide opportunities for active participation and learning. Sustained professional development that focuses on specific academic subject matter and gives teachers integrated opportunities for 'hands on' practice during their instructional day is more likely to increase teacher knowledge and produce positive student outcomes. Formats that promoted implementation included coaching, in-class modeling and observations, and reflective meetings, all elements of the Reading First coaching model.

A recent report on Reading First coaching in five states (Deussen, Coskie, Robinson, Autio & Institute for Education Sciences, 2007) indicates that coaching can mean different things and documents five categories of coaches' roles as implemented through Reading First: a) data-oriented coaches, focusing mainly on data and data results; b) student-oriented coaches, spending a great deal of time working directly with students; c) managerial coaches, focusing on the system of meetings, paperwork, etc.; d) teacher-oriented coaches who focus mainly on whole-group coaching activities such as running meetings or providing training; and e) teacher-oriented coaches who focus mainly on working with individual teachers. This report also found that Reading First coaches may not be fulfilling expectations of the state project personnel, reporting that coaches often spent less time than expected in classrooms or working directly with teachers and report a high level of paperwork and data management. This report also highlights the promise of the coaching model to impact teachers and ultimately student achievement.

Data Sources

For this chapter of the evaluation report, we examined selected items from the survey related to Reading First coaches. Teachers and principals completed survey items reporting on the extent and nature of coach support. Additionally, coaches self-reported on their roles. Additionally, information was obtained from C-TAC personnel regarding the training and expertise of coaches.

To gather additional insight into perceptions of Reading First coaching, teachers, coaches and principals had the opportunity to write in responses to an open-ended question, Teachers and principals responded to the question, "*In your opinion, what aspects of Reading First-funded coaching do you view as most valuable or beneficial and why?*" while coaches responded to, "*In your opinion, what are the most valuable or beneficial aspects of your role as a Reading First coach and why?*" The responses were compiled by respondent group in a text file and used in a qualitative analysis, described later in this chapter.

Qualifications of Coaches

Coaches were asked to report on their level of experience and preparation. Table 5.1 shows the number and percentage of coaches' responses.

Table 5.1: Coach Survey Results Regarding Experience and Preparation

How many years of experience do you have with your district's adopted reading/language arts program?	**N**	**% of total**
Less than 1 year	8	1
2 years	8	1
3 years	23	2
4 years	284	28
5 or more years	609	59
How many years will you have taught or provided support in the primary grades (K-3) as of July 2006?	**N**	**% of total**
Less than 1 year	13	1
1 year	23	2
2 years	27	3
3 – 5 years	207	20
6 – 10 years	279	27
11 – 20 years	273	26
21 – 25 years	87	8
26 or more years	109	11
How long have you been a Reading First coach?	**N**	**% of total**
This is my first year	255	25
This is my second year	204	20
This is my third year	227	22
This is my fourth year	312	30
What qualifications does your school leadership require of its reading coaches? Check all that apply.	**N**	**% of total**
A valid California teaching credential	999	97
Three years or more of successful classroom teaching experience	966	94
Recent, relevant training in scientifically-based reading instruction	773	75
Demonstrated skill in working with adult learners	730	71

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

Additional information about coach qualifications was provided by C-TAC personnel. In California, 1,320 Reading First reading/literacy coaches have received specific coach training aligned with their duties related to implementing the Reading First Assurances. From 2004 through 2006, about 300 coaches completed the Commission on Teacher Credentialing (CTC) Reading Certificate program, qualifying them to be reading experts. An additional 110 coaches matriculated into the CTC Reading and Language Specialist Credential program in the past year, qualifying them to be reading experts at the district level by the summer of 2008. Both of these CTC programs have been partnered with UCLA Education Extension.

The C-TAC has developed and coordinated semi-annual, 2-day professional development for all Reading First coaches. This program, begun in 2003, was designed to enhance the skills of the coach to provide demonstration lessons in classrooms and school site professional development, instruct individual teachers, and facilitate grade level collaborative meetings twice a month. In 2006-07, the C-TAC developed teacher modules for coaches to use with teachers to improve the quality of teaching in specific skill areas. The C-TAC professional development program for coaches was modified to offer much of the same content as the courses for the Reading Certificate program. Given the turnover of Reading First coaches, LEAs had the opportunity to send new coaches to an additional two sessions per year beyond the semi-annual, coach trainings to further build capacity among the coaching force. In 2006-07, 1278 coaches attended these sessions, of which 462 were Reading First coaches.

This extensive training provided to coaches has served to build capacity in LEAs around the state. Based on their acquired expertise, many of the coaches applied for and were accepted as instructors for the teacher summer institutes under the auspices of the statewide network of the Reading Implementation Centers, the authorized Reading First professional development provider. Of the instructor pool of 550, 184 were active coaches, who conducted almost 600 5-day professional development courses for teachers during the summer of 2007. These courses ranged from Year 1 (SB 472), beginning level, to Year 5, most advanced level.

In sum, Reading First coaches, through the semi-annual professional development program of 4 days and their involvement with the professional development summer institutes, either as instructor or participant, underwent from 14 days to 24 days of instruction. The level of expertise and experience among the state's reading coaching force has risen steadily, building capacity at the district, school and classroom levels. It is clear that coaching is an important element of Reading First for sustaining the infrastructure of support to teachers.

Perceptions of Coaching Roles

Accountability

Teachers, principals and coaches were asked on the survey to indicate who held primary responsibility for implementation of the district's adopted reading/language arts program (Question D10 for teachers, E7 for coaches and principals). Table 5.2 presents the percentages of responses from each group. Though the proportion of principals and coaches responding in each category seemed to be similar, the proportions of teachers' responses differed. A higher percentage of teachers (than coaches and principals) indicated that the coach assumed primary responsibility and a lower proportion of teachers indicated that the principal assumed primary responsibility. The Reading First program encourages the principals to take primary responsibility, in collaboration with the coach. This is perhaps an area of implementation that could be strengthened.

Table 5.2: Percentages of Teachers', Coaches', and Principals' Responses Regarding Responsibility for Program Implementation

Who takes responsibility for teachers using the district's adopted reading/language arts program?	Teachers %	Coaches %	Principals %
Neither the principal nor the coach take much responsibility	2	1	0
The principal takes primary responsibility	13	37	36
The principal and coach share equal responsibility	45	47	50
The principal gives the coach primary responsibility	38	14	10

Access

Access to coaching is an important concern in a coaching model. Teachers need to feel they can get the assistance they need, when they need it. To be effective, coaches need ample opportunity to communicate with and work with teachers. Teachers were asked to report their level of access to coaches (Question E1). A similar question was asked of coaches (E12) and principals (E11) regarding coaches' access to teachers. Table 5.3 reports teachers', coaches' and principals' responses. Principals' and coaches' perceptions were similar. A high level of access to coaching is reported across teachers, coaches and principals. This is particularly notable, considering that this is a shift in practice from the traditional model of teacher independence with little involvement of peers.

Table 5.3: Percentages of Teachers', Coaches', and Principals' Responses Regarding Access to Coaches

Teacher Survey: What is your access to a reading coach?	**Teachers %**	**Coaches %**	**Principals %**
The coach is often unavailable	10	-	-
The coach is usually available	53	-	-
The coach seeks me out to assure that I have the support I need	35	-	-
Coach and Principal Survey: How much access do you (do coaches) have to teacher classrooms?			
Not applicable	-	-	0
Coaches need teacher or principal permission to visit a classroom	-	3	1
Coaches have free access to classrooms, but only a few teachers welcome my (the coach's) presence	-	4	3
Coaches have free access to classrooms, but only about half of the teachers welcome my (the coach's) presence	-	13	14
Coaches have free access to classrooms, and almost all of the teachers welcome my (the coach's) presence	-	78	78

Coach as a Resource for Teachers

An important role of the coach is to serve as a resource for teachers; to answer questions, find information or materials, help interpret data and demonstrate instructional strategies. Coaches were generally perceived to be effective in these roles, but not always to provide specific assistance. Table 5.4 provides the percentages of teachers, coaches and principals for relevant items. The majority of respondents indicated that coaches provide help by answering questions or conducting demonstration lessons. In a later section of this chapter, it is apparent that teachers highly value the demonstration lessons but her we see that only 33% of teachers felt the demonstrations significantly improved their teaching. Improving the quality of demonstration lessons provided by coaches may be an area in need of strengthening in the Reading First implementation.

Table 5.4: Percentages of Teachers', Coaches', and Principals' Responses Regarding the Coach as a Resource

Teacher (E2) /Principal Surveys (F5): How helpful is your coach in answering questions about how to teach the program? **Coach Survey (F5): How helpful do you feel you are in answering teacher questions about how to teach the program?**	**Teachers** **%**	**Coaches** **%**	**Principals** **%**
Coach often doesn't know more than the teachers about how to teach the program	7	1	0
Coach gives general answers to questions	24	12	8
Coach gives specific, detailed answers that teachers can use	66	86	87
Teacher (E3) /Principal Survey (F6): If the coach has conducted demonstration lessons, how helpful were they? **Coach Survey (F6): If you conduct demonstration lessons, how helpful are they?**			
Coach does not conduct demonstration lessons	27	5	2
Coach's demonstrations do not help much	6	2	2
Coach provides adequate demonstrations	32	35	25
Coach provides demonstrations that significantly improve teaching	33	58	68

Coach as Facilitator

Coaches are expected to facilitate grade-level meetings, lesson studies, and data analysis sessions with teachers. Table 5.5 presents the percentages of teachers, coaches and principals who indicated varying levels of coach involvement in the facilitator role. Teachers reported a lower level of facilitation than did coaches and principals, but overall the results are positive. The majority of respondents viewed coaches as facilitating meetings and keeping the meetings focused on instructional needs. Both of these roles are important.

Table 5.5: Percentages of Teachers', Coaches', and Principals' Responses Regarding the Coach as a Facilitator

Does the coach (do you) facilitate regular grade-level teacher meetings related to your district's adopted reading/language arts program?	Teachers %	Coaches %	Principals %
Coach is not involved with the grade-level meetings	23	12	7
Coach helps facilitate the meetings regularly	46	37	33
In addition to facilitating meetings, coach keeps the focus on instructional needs of teachers	29	49	56

Perceived Value of Coaching

In this section, findings are reported from the qualitative analysis of the open-ended responses to a survey question designed to obtain further information about perceptions of coaching. Teachers and principals responded in narrative form to the question, "*In your opinion, what aspects of Reading First-funded coaching do you view as most valuable or beneficial and why?*" while coaches responded to, "*In your opinion, what are the most valuable or beneficial aspects of your role as a Reading First coach and why?*" Similar to the format used in Chapter 4, in this chapter we use qualitative research methodology to examine findings from the open-ended question regarding coaching. In this analysis, we gain insight from school personnel who are most directly involved with implementing the Reading First coaching model.

In this analysis, we first examine the perceptions of teachers, coaches and principals as reported in the open-ended question. We compare the relative perceived importance of resulting categories of responses across respondent groups. Then, we examine differences in perceptions across high implementing and low implementing schools, as determined by the RFII (Reading First Implementation Index, see Chapter 3). The reader is referred to Chapter 4 for information about the nature and benefits of qualitative methodology for understanding educational phenomena, such as coaching.

Of the 17,261 teacher surveys collected, 12,243 wrote narrative responses to this question, or 70.9%. Of the 1,028 coach surveys collected, there were 928 narrative comments submitted, or 90.3%. Of the 1,073 principal surveys collected, there were 947 comments submitted, or 88.3%. This high response rate lends validity to the findings, ensuring that the opinions expressed are likely to be representative of all Reading First participants.

As stated in Chapter 4, limitations in the generalizability of qualitative findings should be noted. Though qualitative research may provide in-depth insight into phenomena and why they occur among participants, results are viewed as inconclusive. Interpretability is also somewhat limited without the ability to quantify findings. In this study, we seek to report findings that are verified by high occurrence in the dataset or

confirmation across respondent groups. The key advantage of using qualitative methodology is the ability to derive findings and interpret them within a specific context. In this evaluation study, the qualitative data were examined for converging evidence of elements of the coaching model that are high in importance across multiple perspectives (teachers, reading coaches, principal) and contexts (high and low implementing school sites).

Data Analysis Methodology

The methodology used in this chapter was identical to that reported in Chapter 4. We refer the reader to Chapter 4 for an explanation of the coding and categorization procedures.

Relative Importance of Aspects of Coaching

The codes (or categories of responses) were sorted according to the frequency with which they occurred and are listed in rank order in Table 5.6. Descriptions of these categories and sample comments are provided in Table 5.7. Rankings are listed for the whole data set combined and then for teachers, coaches and principals. This allows the reader to compare the different participants groups' perceptions in terms of relative importance of elements of coaching. Note that this table does not depict a rank ordering by teachers, coaches or principals. Rather, it depicts the relative frequency with which the codes occurred and should not be interpreted as ratings. Table 5.6 also lists the percentage of the total respondents that commented within each category. The total number of written responses provided was used to calculate these percentages, not the total number of surveys received by each group. Note that the percentages will not total 100% because in many cases, comments were assigned multiple codes. This occurred when a response included multiple ideas or concepts, or when a response could be interpreted as falling within more than one code. Additionally, some responses were not coded at all because they were irrelevant to the question and the purpose of this part of the study.

In this table, we see that the order of frequencies is similar for the columns of All and Teacher respondents, but not identical. This is because the teachers made up the largest proportion of the response pool. There is some variability in the Coach and Principal listings. Demonstration by Coaches was in the top three rankings across respondent groups indicating that this is a highly valued aspect of coaching. Teacher Support, which was highest in frequency among coaches and principals, was sixth in frequency for teachers; however, Coach as a Resource (third highest for teachers) is a very similar category, so the function of coaches in which they provide support and resources could be thought of as highly valued also. Other high-frequency categories of responses across groups included Instructional Strategies, Program Implementation Support, and Collaboration/ Grade-Level Planning. The Negative Comments category captured a variety of comments but occurred at a relatively low frequency.

Table 5.6: Rank Order and Percentages of Responses for Categories

Response Category (Code)	All N = 14,118		Teachers N = 12,243		Coaches N = 978		Principals N = 947	
	Rank	**%**	**Rank**	**%**	**Rank**	**%**	**Rank**	**%**
Demonstration by Coaches	1	24.2%	1	17.5%	3	28.3%	2	36.1%
Teacher Support	2	17.2%	6	8.1%	1	40.7%	1	38.9%
Instructional Strategies	3	15.4%	2	10.8%	5	25.9%	8	17.7%
Program Implementation Support	4	14.6%	8	5.5%	2	35.1%	3	35.9%
Collaboration/ Grade-Level Planning	5	13.9%	5	9.0%	4	26.8%	5	21.2%
Knowledge and Skills Provided by Coach	6	12.2%	4	9.4%	10	16.1%	12	11.6%
Coach as a Resource	7	11.9%	3	10.0%	12	13.2%	13	7.3%
Data Analysis/ Assessment	8	11.9%	7	6.8%	6	22.1%	4	23.7%
Professional Development	9	9.8%	9	5.3%	8	21.1%	7	19.4%
Qualities of Coach	10	7.2%	11	4.3%	16	6.5%	6	19.3%
Observation and Feedback	11	7.2%	13	3.1%	11	14.9%	9	15.3%
Expertise of Coach	12	6.5%	12	3.3%	14	8.2%	10	14.7%
Negative Comments	13	5.8%	10	4.8%	17	4.2%	17	1.9%
Improvement of Lesson Quality	14	5.3%	15	1.2%	7	21.2%	11	11.6%
Comments re Waivered/Bilingual Classes	15	5.1%	17	1.0%	9	19.8%	19	0.6%
Improved Student Achievement	16	3.5%	18	0.9%	13	12.0%	15	3.9%
Coach in Non-Judgmental Role	17	2.7%	14	1.6%	18	3.6%	14	4.9%
Improved Awareness of Research	18	1.7%	19	0.6%	15	7.3%	16	2.2%
Increased Accountability of Teachers	19	1.3%	16	1.1%	19	1.2%	18	0.6%

Code Characterization

For each code, or response category, in Table 5.7 below, a brief definition is provided along with representative comments from the respondents. These are listed in the order of frequency occurring within all respondent groups combined. Descriptors are provided for all, but comments were included only from those categories that occurred within 10% or greater of any respondent group.

Table 5.7: Code Descriptions and Representative Comments

Code Description	Reasons and Representative Comments
Demonstration by Coaches States that demonstration or modeling of lessons and teaching techniques by coaches is a valuable aspect of Reading First.	"Lesson demonstrations by the coach are the most beneficial. 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 (Teacher)" "I feel the demonstration lessons really provide a 'hands on' approach to learning. Teachers can sit and see what good teaching looks like from the modeling of a coach (Principal)" "Through demonstration lessons teachers have the opportunity to observe the coach's delivery and pacing of lessons and use of effective strategies to enhance student engagement and implementation of scaffolds needed for student learning (Coach)" "The coach is very helpful and gives me suggestions as well as demonstrates teaching techniques that will improve my teaching (Teacher)" "Demonstrations, observations and feedback to teachers is very beneficial because teachers need continuous support, reflection and practice time to make instructional changes (Principal)"
Teacher Support States that coaches are supportive to the teachers in providing many different types of assistance to the teachers including cognitive planning and reading practices.	"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)" "The most valuable aspect of my role is to provide support to my teachers in implementing the program, clearing up any misunderstandings and helping them reach the needs of their students. I have also helped teachers to cognitive plan their lessons (Coach)" "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 (Principal)" "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)" "Our coaches are very supportive and consistently look for ways to assist both the teachers and the students. They model lessons, conduct workshops and implement action plans for improvement (Teacher)" "The Reading First coach is providing direct support for teachers in the classroom. She is able to focus on teachers that administrators have identified as needing extra support. This is something that we would not be able to provide otherwise (Principal)"

Instructional Strategies States that coaches provide teachers with guidance and planning of instructional practices and strategies. Further indicates that coaches are knowledgeable in the area of instructional strategies.	"The coach's professional ideas for bettering the teaching in the classroom and daily support the coach gives to teachers is invaluable (Teacher)" "I am able to assist teachers in improving teaching practices and guiding instruction to reach all the students in the class (Coach)" "One of the most valuable aspects is the bank of strategies that I can provide to teachers. Through demo lessons and collaborative meetings teachers can incorporate new strategies in their lessons and continuously improve (Coach)" "Help with strategies and ideas is most beneficial. The coach is always on target as to what I need to include in my instruction at the time (Teacher)" "Coaches and teachers are able to dialogue about program implementation as well as effective teaching techniques and strategies in order to improve instruction in the classroom (Principal)"
Program Implementation Support Indicates that coaches monitor, support and guide implementation of the reading program.	"The most valuable aspect of coaching is being able to provide a highly trained person to improve teachers' program implementation through observation and feedback, demo lessons and elbow coaching (Principal)" "The coaches provide needed assistance and guidance to help implement the curriculum and make it most beneficial for the students (Teacher)" "The most valuable aspect of my role as a Reading First coach is the support I provide in helping teachers understand the purpose of the components of the program and how to implement them to achieve the highest success with the students (Coach)" "Coaching provides guidance to stay on track with the program and focus on student achievement (Principal)"
Collaboration/ Grade-Level Planning States that coaches are important in facilitating meetings between teachers, administrators and school staff. Coaches are connectors of people in the school environment.	"Cognitive coaching and grade level collaboration are powerful tools that can shape and strengthen a staff's professional and instructional development (Coach)" "A well-trained coach on site to assist, facilitate and collaborate with teachers is essential to the program's success. The ability to collaborate and assist teachers in analyzing data, targeting student achievement and teaching strategies has been valuable (Principal)" "The lesson study planning and implementation have been very helpful. The colleague feedback time and time for reflection have also been extremely beneficial to my teaching (Teacher)" "I feel that coordination of collaborative meetings has helped my staff to share ideas and become stronger with implementation. I also feel that facilitating Action Plan meetings has helped to focus our goals (Coach)" "Having a literacy coach at the school full time provides an open forum for administrators and coaches to deepen the collaboration discussion and broaden their knowledge base. It gives support for administrators to evaluate and implement the district Reading/Language Arts program and helps support teachers by providing a focus for grade

	level collaboration (Principal)" "I think that our unit planning time has been very valuable. It gives us an opportunity to meet as a grade level with our coach and plan out specifics for each unit (Teacher)"
Knowledge and Skills Provided By Coach Indicates that the coach supports the development of teachers' knowledge and skills; deepens and expands teachers' expertise. Further states that coaches clear up any misunderstandings about the reading program or Reading First.	"It is very helpful to have a coach on site who is always available to answer questions I may have regarding the program (Teacher)" "Having staff that are knowledgeable of all the program details and at all grade levels is invaluable for teachers and administrators (Principal)" "The most valuable aspect of my role as a Reading First coach is the ability to work with teachers and support teachers in their continuous learning and improvement as professional. As I coach, I impact hundreds of students by sharing my knowledge with their teachers (Coach)" "Reading First funded coaches are valuable when coaches are knowledgeable in the program and are able to provide assistance to new teachers when needed (Teacher)"
Coach as a Resource States that the coach is a resource of materials, information and ideas for teachers and their specific classroom needs.	"The coaches are always there to support us in many ways. They provide great model lessons and materials and if we have questions which are unknown at the moment, they go out of their way to find answers for us (Teacher)" "The most valuable aspect of my role is being the resource person at my school site. I am able to find activities and present new ways to teach by providing training and support (Coach)" "I like having an onsite person to ask questions of and get help when needed. The coach is available to meet with me and help me design lessons. Having someone who actually understands the students as well as the material is very helpful (Teacher)" "I value being a resource to support and answer any questions or concerns that the teachers may have about the program implementation or research that guides the program (Coach)"
Data Analysis/ Assessment Describes that coach's role in supporting and guiding teachers in the area of data collection and analysis of student data is vital.	"I find it very valuable that I can sit with my reading coach and analyze data and determine the needs of my students. We are able to brainstorm ideas to better reach those students (Teacher)" "I believe the Reading First grant has really brought us together as a school. We now collaborate every week and discuss data (Coach)" "My role as a coach has helped student achievement at my school by helping teachers with looking at data and reflecting on it (Coach)" "The literacy coach will review the 6-8 Weeks Skill Assessments with me and make me aware of what my students need overall and individually to improve my success in teaching reading (Teacher)" "Facilitating the teachers in their analysis of data, assisting them in recognizing areas of strength and areas that need growth, and identifying changes in instruction and intervention to achieve that growth (Principal)"

Professional Development States that coaches provide professional development for teachers and receive their own professional development to become experts. Professional development is a vital aspect of the Reading First program.	"I find that the caches receive ongoing training that they are able to share with our faculty. This helps us keep up with better ways to teach language arts. They also provide excellent professional development that assists us in implementing better practices in the classroom (Teacher)" "I believe my most valuable role as a coach is to share what I have learned at Coach Institutes and other trainings through demonstration lessons, staff meetings, grade level meetings and workshops (Coach)" "Providing extensive training to the coach is important so that the coach is able to support teachers across grade levels in all components (Principal)" "I can provide ongoing research-based training and support for our teachers. I work collaboratively to plan staff development and promote school-wide instructional improvement (Coach)"
Qualities of Coach The quality of the coaching makes an impact on a school. Responses indicate that high quality coaching has a positive impact and low quality coaching does not.	"The teachers and I have immediate access to someone who is knowledgeable in literacy to model lessons, facilitate grade level meetings, provide demonstrations and observe teachers (Principal)" "The coach is an integral part of what we do at the school. Her expertise and support of teachers help make them better teachers. Teachers feel comfortable with her. All of these aspects, no doubt, lead to improved student achievement (Principal)" "Teachers are understanding why some practices are better than others. I am able to help teachers by passing along best practices, which many teachers never get to see. It also makes teachers feel good about what they have done (Coach)" "Our coach is very organized and efficient. She works very hard to answer any questions we have and keeps us motivated. She puts a lot of time and effort into her job and it shows in our student data (Teacher)"
Observation and Feedback States that coaches conduct regular observations of reading lessons and help teachers improve by giving feedback.	"Monitoring the implementation of the program ensures fidelity. The coaches providing feedback to teachers with acknowledgement of components and strategies that are in place as well as areas to has helped to improve instruction (Principal)" The most valuable aspects of the Reading First coaching (for me) are the immediate feedback on instructional practices which allows me to be more effective, and strategizing together about what instructional practices to use in a workshop/intervention when the 6-8 week assessments results come back (Teacher)" "I try to take each teacher to the next level of teaching in terms of their implementation and their knowledge. I meet with teachers in a variety of settings, individually, small groups or whole staff. Each venue allows me to customize and individualize my coaching. I succeed in my mission when teachers believe all students can learn from them and they analyze and response to their students' needs. I succeed when I see all students learning to read and progress in language arts (Coach)"

Expertise of Coach States that because coaches have developed expertise to be a specialist in reading curriculum or instruction, they are valuable because of this expertise.	"Having a knowledgeable highly trained coach is of great value because it supports the teaching and it helps with refinement of the core Language Arts Program (Principal)" "Our coaches are experts in reading and can apply their knowledge to actual situations occurring in the classrooms. They are not rule-stickers but thinkers. They provide depth of understanding to teachers of why they are doing what they are doing and why some things do not work (Principal)" "Our coach is highly qualified, exceptionally well prepared, exceptionally conscientious and has a remarkable can-do attitude in the face of tedious tasks (Teacher)"
Improvement of Lesson Quality Shows that when coaches help teachers with their lessons, the quality of instruction is improved.	"The coach takes the time to come into my classroom and teach a lesson to help me find new ways to teach that lesson. I like to have the coach's honest input, not criticism (Teacher)" "The coach helps teachers to enhance their knowledge and their lesson structure (Principal)"
Comments re Waivered/Bilingual Classes Comments discuss bilingual education or waivered classrooms	"If they (coaches) are truly capable, they can model bilingual lessons (Teacher)" "Having literacy coaches at our site has really been a great benefit for all our teachers and especially to help guide instruction, monitor student progress and help our English learners (Principal)" "I advocate for our English learners and continuously put their needs at the forefront (Coach)"
Improved Student Achievement Comments refer to improved student achievement as a goal or result of Reading First coaching	"The use of a reading coach as a support to the grade levels when interpreting student data and determining the next steps for instruction has helped increase student achievement (Principal)" "Having a coach is very beneficial. When coaches really support teachers, they answer questions and guide teachers and students to get better results in language arts (Teacher)" "The coach is a trusted knowledgeable partner for all teachers and myself to work with to continue to improve student achievement (Principal)" "Our coach is a valuable part of our efforts to raise student achievement (Teacher)" "I collaborate with teachers to develop action plans to improve student achievement (Coach)"

Coach in Non-Judgmental Role Indicates that coaches are seen as non-judgmental, non-evaluative, or non-threatening. When coaches go into classrooms to observe, they can be objective and non-evaluative.	"I value having a coach that is non-evaluative and helpful in any way that can help the teacher (Teacher)" "The coaches have an ability to assist the teachers in a way that is not evaluative so teachers take more advantage as it is less threatening (Principal)" "Our coach is always willing and available to help. She is non-threatening in my classroom and I know she is there to help, not criticize (Teacher)" "Teachers have a colleague in the classroom who is there to assist in implementation of the program. Any corrective action that is taken in the implementation is not viewed as punitive (Coach)"
Improved Awareness of Research Indicates that coaches improve teachers' awareness of research and how it applies to instruction	"Reading First funded coaches are a valuable resource to tap for up-to-date cutting edge research information and instructional strategies (Teacher)" "The strong knowledge base and rich coaching support. She provides strong research foundations for what we are doing and how to best do it (Principal)" "The coaches keep everybody updated on current data and resources (Teacher)" "I explain the effectiveness of certain strategies, materials and organizational structures since many teachers are not familiar with scientifically-based reading research (Coach)"
Increased Accountability of Teachers The coach helps the teachers to be accountable for full implementation of the program	"I believe Reading First funded coaching helps teachers stay on track and meet the targeted goals for our district (Teacher)" "Our coach stays focused on the accountability component which has provided the teachers with a deeper understanding and focus of the standards and the purpose for their instruction (Principal)"

Negative Comments Any negative response about Reading First coaching	Note: Many of the negative comments were vague and did not fall into subcategories. The subcategories listed here represent relatively small numbers of respondents. They are included here only to demonstrate contrasts to the positive comments included in the other categories in this table. ***Lack of Administrator or Structural Support*** "If I were supported in the implementation of the program, then I could be beneficial (as a coach). It would be valuable for teachers to have someone to discuss assessment results with, but I am not allowed (Coach)" "The messages or mandates we receive from the coach could just as well be sent by email. I don't feel that the support is really there for implementation of the program. We don't really have coaching (Teacher)" "I am disappointed that Reading First did not keep up the administrator and coach monthly seminars (Principal)" ***Lack of Availability*** "Coaches are shared with schools. We need our coach to be here at our school for more time in order for our needs and students' needs to be met (Teacher)" ***Lack of Buy-in from Teachers*** "I have been less valuable than I would like to be. This is my first year at this site and the teachers feel they have no need for a literacy coach (Coach)" "About 50% of the teachers at this school have resisted taking the time to debrief after I demonstrate or observe a lesson (Coach)" "I think that the Reading First coaching is a waste of time after two years of the adopted program. The teachers at my school are well trained and capable of understanding the program. It is insulting to pay for a coach (Teacher)" "I don't think coaching is valuable at all. We are teaching our regular Language Arts program like we would anyway (Teacher)" ***Need for More Demonstration Lessons*** "The teachers did not request model lessons so there was not much coaching this year (Teacher)" "The weekly lesson modeling would be great, but it just hasn't happened here (Principal)"

Conclusions

In conclusion, this chapter finds that reading or literacy coaches are an integral part of the Reading First program in California. They are highly valued by program participants. The positive perceptions of coaches and the importance of their role in supporting the curriculum and implementation are almost universal. Coaches serve important functions in supporting implementation of the Reading First program and maintaining a school's focus on improving student achievement.

The Reading First initiative has provided extensive training and support to coaches, an effort that has built capacity at the district, school and classroom levels. Coaches have reached a high level of expertise to the point that many are now qualified to provide training at Reading First institutes.

The most important functions served by coaches in California are providing demonstration lessons, serving as a source of resource and support for teachers and facilitating collaboration focused on student achievement and fidelity of implementation. Most coaches have ready access to classrooms to provide support with curriculum, instruction and assessment.

References

Darling-Hammond, L., & McLaughlin, M. (1995, April). Policies that support professional development in an era of reform. Phi Delta Kappan, 76(8), 597-604.

Deussen, T., Coskie, T., Robinson, L., Autio, E., & Institute of Education Sciences (ED), W. (2007, June 1). Coach Can Mean Many Things: Five Categories of Literacy Coaches in Reading First. Issues & Answers. REL 2007-No. 005. Regional Educational Laboratory Northwest, (ERIC Document Reproduction Service No. ED497517) Retrieved December 1, 2007, from ERIC database.

Garet, M., Porter, A., Desimone, L., Birman, B., & Yoon, K. (2001, January 1). What Makes Professional Development Effective? Results from a National Sample of Teachers. American Educational Research Journal, 38(4), 915

Novick, R. (1996, January 1). Actual Schools, Possible Practices: New Directions in Professional Development. Education Policy Analysis Archives, 4(14), 1-15.

Chapter 6: Impact of Reading First on English Learners

The purpose of this chapter is to examine the impact of the Reading First program on English learners (ELs). In this chapter, we examine the reading achievement of this subgroup of students. We also examine the responses of participants to an open-ended question that provided teachers, coaches and principals opportunities to express their views of the impact of the Reading First program on ELs. It is important to note that in California (and in the Reading First program), the instructional model is not uniform for ELs. These students may receive instruction in English, with an emphasis on immersion into the English language, or in a bilingual setting (waivered classrooms), with the transition from Spanish to English occurring during the primary grades. This chapter looks at the impact of Reading First for EL students as a group, regardless of instructional setting. Chapter 7 examines the waivered classroom setting further.

This chapter yields the following key findings:

- Achievement gains for English learners in Reading First schools are positive for grades 2, 3 and 4.
- Achievement gains are higher for English learners in Reading First schools than for English learners in non-Reading First schools for grades 2 and 3.
- Achievement gains are higher for English learners in high implementing Reading First schools than ELs in low implementing Reading First schools and non-Reading First schools.
- The EL subgroup is more sensitive to differences in Reading First implementation than the student population as a whole. The EL subgroup in low implementing Reading First schools is particularly at risk for low growth, whereas ELs in high implementing Reading First schools often grow more than the student population as a whole.
- The effect of Reading First implementation on EL achievement is reproduced for ELs in grade 4. However, the non-Reading First EL subgroup shows higher growth than the EL subgroup in Reading First schools. We hypothesize that this may be a statistical artifact of EL reclassification criteria that reclassify ELs to English-fluent based on grade 3 California Standards Tests (CST) results.
- In open-ended survey comments, teachers, coaches and principals reported overall positive regard for the Reading First program and its appropriateness and support for ELs.
- In open-ended survey comments, teachers, coaches and principals reported significant improvement in the curriculum and instruction for EL students due to their schools' participation in Reading First.

- In open-ended survey comments, teachers, coaches and principals noted evident and significant improvement in the vocabulary, language development and reading achievement of ELs as a result of their schools' participation in Reading First.
- Though there was generally a positive perception of the impact of Reading First on EL students, some participants expressed concerns regarding the amount of time needed to effectively teach ELs, specific aspects of the curriculum and materials, the pacing of instruction for ELs and the need for more systematic English language development to better meet the needs of ELs.

Research on Reading Instruction for English Learners

There has been a significant rise in the number of EL students in schools during the past decade in California. Over 25% of the state's K-12 students are considered ELs, but the percentages range from 30% to 40% in the primary grades. In the Reading First program, the percentage of ELs was 53.7% for Cohort 1, 54.7% for Cohort 2, 58.5% for Cohort 3, and 31.2% for Cohort 4. Studies and databases continually demonstrate the pervasive academic difficulties of ELs. Many students who enter school with a primary language other than English score below competency markers on academic achievement; over 50% score in the bottom third in reading or mathematics with a continued gap between EL and non-EL reading achievement (National Center for Education Statistics, 2007).

There are important questions that remain to be fully addressed in the research literature regarding effective instruction for ELs. What do we really know about effective reading instruction for ELs? Are the same curricula, practices and assessments used with non-ELs as effective with ELs? We are left with discerning the best of what we know from a limited research base. In the most comprehensive effort to date to examine the research on EL literacy development, Snow (2006) summarized the work of the National Literacy Panel on Language-Minority Children and Youth: "The literature we reviewed reveals remarkably little about the effectiveness of different aspects of instruction, and provides only limited guidance about how good instruction for second-language speakers might differ from that for first-language speakers (p. 638)… Most discouraging, the research we reviewed provides little basis for deciding whether or what kinds of accommodations or adaptations are most helpful to second-language learners (p. 639)." Additionally, constraints imposed by the politics of educational policy, most notably arguments over bilingual versus English-only instruction, draw attention to the lack of definitive answers from research (Gersten, 2006; Gersten & Baker, 2000). However, there is some evidence that word-level instructional components prevalent in the Reading First program are effective with ELs, such as explicitly teaching phonological awareness, letter-sound relationships and decoding, especially when taught along with meaningful experiences in engaging text (Chiappe & Siegel, 2006; Chiappe, Siegel, & Wade-Wooley, 2002; Snow, 2006). Not only do we need to know what practices are effective, but as Klingner

and colleagues state, we need to know what works "with whom, in what contexts, and under what circumstances (Klingner, Sorrells & Barrera, 2006. P. 223)." The Reading First program is the first comprehensive effort to date in California to provide instruction that relies on the best of what we know from research.

What constitutes effective reading instruction in the primary grades? In a series of observational studies in first grade California EL classrooms, there were specific instructional practices that correlated significantly with EL reading gains (Baker, Gersten, Haager & Dingle, 2006; Haager, Gersten, Baker & Graves, 2003; Gersten, Baker, Haager & Graves, 2005; Graves, Gersten & Haager, 2004). These included such practices as modeling, making instruction explicit, and prompting students, instruction geared toward low performers, explicit phonemic awareness and decoding instruction, monitoring student performance, extensive vocabulary development, and sheltered English techniques.

A recent report, "Similar English Learner Students, Different Results: Why Do Some Schools Do Better?" examined school and instructional factors related to positive outcomes for EL students (Williams, Hakuta, Haertel, et al., 2007). Using schools' Academic Performance Index (API) and students' California Standards Tests (CSTs) and California English Language Development Test (CELDT) scores, the report found several practices that are similar to those promoted in Reading First to be strongly correlated to improved outcomes for ELs. One factor, the extensive use of assessment data, is a cornerstone of the Reading First initiative. The coherence and consistency of the curriculum and instruction, and the focus of a school on achievement gains were two additional strong correlates of EL achievement. These factors would also be considered to characterize the Reading First initiative.

Data Sources

For this chapter of the evaluation report, we extract relevant student achievement results that were reported in Chapter 2 and examine them to determine the extent to which the Reading First program has differentially impacted ELs in California.

To gather additional information about the impact of Reading First on ELs, teachers, coaches and principals had the opportunity to write in responses 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."* The responses were compiled by respondent group in a text file and used in a qualitative analysis, described later in this chapter.

EL Student Achievement

For a full discussion of the rationale for using the various achievement metrics presented in this report, consult Chapter 2. In the current chapter, we use two of the previously described achievement metrics to measure school progress or growth (achievement gains) for the English learner (EL) subgroup of students as classified using the California English Language Development Test (CELDT) and recorded in the California STAR file. The two achievement metrics are the percentage of EL students in a school that are in the "Proficient" or "Advanced" CST performance categories ("% Proficient and Above") and the average CST English language arts scale score of EL students in the grade ("Mean Scale Score"). Because percentages of EL students in the Below Basic and Far Below Basic proficiency levels were not available for the 2006-2007 school year in the STAR research file, we are not able to address questions regarding the extent to which ELs in Reading First schools migrate out of the bottom CST performance categories. However, migrations into or out of the bottom categories do have an effect on the Mean Scale Score.

The number of schools reported in this chapter is lower than that reported in Chapter 2 because some schools do not have CST data for the English learner subgroup. This is especially noticeable for the group of non-Reading First schools. As in Chapter 2, the number of schools reported for grade 4 is much less than for other grades because they are confined only to schools that have been in the program for five years.

We report achievement gain scores as our indicator of EL progress. As in Chapter 2, the CST gain score reported in the tables of this chapter is the 2007 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 EL 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.

To provide context for studying the EL Reading First Gains, we compare the achievement gains of ELs in Reading First schools to the gains of ELs in non-Reading First schools. The upward trend seen for the Reading First schools is mirrored in the rest of the state, but we reiterate that the non-Reading First group of schools is demographically dissimilar to the Reading First group, and caution should be exercised when comparing them. In the trend-line charts presented later in this chapter, the All Non-Reading First Elementary Schools group (which has a starting point significantly higher 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. It should also be noted that when comparing schools using the English learner subgroup, the count of non-Reading First schools is about half the count obtained when using the entire student population. This is because schools with fewer than 11 English learners are not included in the

STAR file for purposes of EL subgroup analysis. This substantially complicates the interpretability of the non-Reading First population.

We also compare the achievement gains of ELs in high and low implementation Reading First schools. Chapter 3 of this report describes how the Reading First Implementation Index (RFII) was computed in order to measure the degree to which the Reading First program is being implemented in each school. The RFII was used to divide Reading First schools into two groups labeled High Implementation Schools and Low Implementation Schools, and the school classification in this chapter is the same as in Chapter 2. We define a high implementation school as 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 is one with an average yearly RFII less than 36.0.[1] This classification scheme leaves out schools between 36.0 and 41.4 from the high and low groups, but they continue to be represented in the "All Reading First Schools" category.

The following pages present a series of tables and trend-line charts that parallel the analysis presented in Chapter 2. The tables and charts provide starting scores, ending (2007) scores, and gains on each of the two achievement metrics available for the EL subgroup. They are the basis for our conclusion that Reading First is an effective program for English learners. Before presenting the achievement results, we repeat two points useful in interpreting the tables:

1. Interpreting Significance Tests. The statistics in the achievement tables provided in this chapter are sometimes accompanied by superscripts "a", and "b". 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 ELs in the non-Reading First schools 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 that the new group average (ending year, 2007) is significantly higher than where it started, 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.

[1] 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.

2. Rounding Errors. Sometimes we report a gain score that does not appear to 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 was computed at more than eight decimal places. Thus the reported gain is (slightly) more accurate than the difference between the reported starting and ending scores.

3. Trend-lines of Non-Reading First Schools. When graphing the trend-lines for ELs in non-Reading First schools, we continue the convention of adjusting their trend-lines downward to have the same starting point as the ELs in Reading First schools.

Summary Gains (Table 6.1)

Table 6.1 reports the achievement gains of English learners across all Reading First schools (all YIPs) in terms of average yearly gain in the mean "% Proficient and Above" achievement metric and the mean CST scale score metric. In other words, it reports the difference between a school's starting score (in the year previous to entry into Reading First) and its ending year (2007) score, divided by the number of years it has been in the program. This difference is averaged across all applicable schools. There is no trend-line chart because the starting point is different for each YIP. The "All Reading First Schools (All Students)" column repeats data presented in Table 2.1 of Chapter 2. The gains in the four columns headed "English Learner Students," including the "All Non-Reading First Elementary Schools" column, are computed using *only* data for the EL subgroup. The first column is computed using data for both EL and non-EL students.

Table 6.1: Summary Gains for English Learners, All YIPs Combined, All Grades, Mean Yearly Gain

All Schools, All Grades, Average Change Per Year	Reading First Schools				
		English Learner Students			
	All Reading First Schools (All Students)	All Reading First Schools (EL Only)	High Implementation Schools, RFII > 41.4 (EL Only)	Low Implementation Schools, RFII < 36.0 (EL Only)	All Non-Reading First Elementary Schools (EL Only)
Grade 2, CSTs	(N=831)	(N=786)	(N=132)	(N=284)	(N=2103)
% Proficient and Above	3.8	3.3[b]	3.8[b]	3.1[b]	3.2
Gains in Scale Score	4.5	4.2[b]	5.1[ab]	3.8[b]	3.9
Grade 3, CSTs	(N=832)	(N=779)	(N=127)	(N=283)	(N=2026)
% Proficient and Above	1.6	1.2[ab]	1.5[ab]	0.9[ab]	0.3
Gains in Scale Score	2.9	2.9[ab]	3.3[ab]	2.5[ab]	1.2
Grade 4, CSTs	(N=255)[1]	(N=235)	(N=25)	(N=96)	(N=1869)
% Proficient and Above	3.2	2.2[ab]	2.6[b]	1.8[ab]	3.0
Gains in Scale Score	4.1	3.7[b]	4.8[b]	3.0[b]	3.8

[a]Significantly different ($p < 0.05$) relative to English learners in "All Non-Reading First Elementary Schools".
[b]Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.
[1]The grade 4 sample includes only YIP 5 schools, hence the smaller N.

A casual examination shows that achievement growth for English learners is substantial for grades 2 and 3, and more modest for grade 4 but still significantly greater than zero. The grade 2 and 3 English learners in Reading First schools show higher gains than ELs in non-Reading First schools. The English learners in high implementation Reading First schools have higher achievement gains than ELs in either low implementation Reading First schools or non-Reading First schools. High Reading First implementation greatly affects school growth rates, both for ELs and the student population as a whole. In general, the relationship between implementation and achievement for ELs appears to confirm that found in Chapter 2 for all students. Reading First works for ELs as well as non-ELs.

However, when Table 6.1 is compared side-by-side to Table 2.1 from Chapter 2 – average yearly gains for all students (not just ELs) – a more complex story emerges. First, we note that the non-Reading First schools post stronger gains on the two achievement metrics when we focus on the EL subgroup alone than when we look at the student population as a whole (though the relationship is complicated by the lower number of non-Reading First schools, since schools with less than 11 EL students are dropped from the STAR file). We see that EL instruction appears to be improving more rapidly than instruction for the student population as a whole. However, the Reading First schools do not repeat this pattern. For example, the non-Reading First EL subgroup in grade 3 grew 1.2 scale score points versus 0.5 scale score points for the "all" non-Reading First student group (which is 0.7 scale score points *more*), whereas the Reading First EL subgroup grew 2.9 scale score points, which is the *same* as the growth rate for the all students group in Reading First schools. We do not know definitely why the "EL versus All Students" difference should be higher in non-Reading First schools than in Reading First schools, but note that the Reading First schools have much higher concentrations of predominantly Hispanic EL students and related demographic groups as reported in Table 1.1. Thus, we might hypothesize that in a context where non-Reading First schools are showing achievement growth for all students, the relatively small EL student populations in those schools might experience stronger pressure to "keep up" with the rest of the school population and thus post higher total gain rates, as compared to the EL student population in Reading First schools where the EL population generally exceeds 50% of the total student population.

In comparing Table 6.1 to 2.1, we also see that the grade 4 growth rates for the EL subgroups in Reading First schools are substantially lower than for the population as a whole, and that while the relationship between implementation and achievement continues for the EL population, the absolute growth rates are much less impressive. This is supported by the fact that while grades 2 and 3 show *somewhat* lower growth rates for Reading First EL students relative to all students (e.g., in Grade 2 a 3.8 percentage point gain for all students versus a 3.3 percentage point gain for ELs), grade 4 shows *much* lower growth rates (3.2 percentage points versus 2.2 percentage points). This observation is tempered by the finding that the grade 4 drop in growth rates is much more pronounced in the "% Proficient and Above" achievement

metric than in the mean scale score metric. In grade 2 the mean scale score gain is 4.5 scale score points for all students versus 4.2 scale score points for ELs, whereas in grade 4 the mean scale score gain is 4.1 scale score points for all students versus 3.7 scale score points for ELs – not as striking a discrepancy as that found for the "% Proficient and Above" achievement metric.

One hypothesis that could explain the grade 4 anomaly (relative to other grades, and relative to all students) is the possibility that in a substantial number of districts high performing ELs are being reclassified as fluent in English based on their grade 3 CST scores. This would substantially lower the performance of the EL subgroup in grade 4 relative to earlier grades since the grade 4 EL subgroup no longer includes the high performing ELs from grade 3. It would also lower the EL subgroup performance relative to non-ELs. The artifact would be more pronounced in grade 4 since it has been reported (anecdotally) that many districts prefer not to reclassify their students until CST scores have been obtained from both grades 2 and 3. The artifact would be more pronounced in the "% Proficient and Above" achievement metric than in the mean scale score metric because scoring "Proficient" on the grade 3 CSTs is often used as a prerequisite for reclassifying ELs. The mean scale score metric, on the other hand, counts students at all ability levels, including the great majority of students at the lower performance levels who are not reclassified.

EL reclassification criteria differ substantially across LEAs. Without data regarding the reclassification criteria used in Reading First LEAs, our proposed explanation can be no more than a hypothesis. Other hypotheses are possible, but they are harder to reconcile with all the facts.

A third finding emerges when comparing Table 6.1 with 2.1, perhaps the most important of all. The contrast in growth rates between High and Low Implementation schools is more pronounced for the EL subgroup than for the "all students" group. For example, Table 2.1 shows that for "all students" the high implementing schools growth rate was 5.1 scale score points versus 4.1 points for the low implementing schools. Table 6.1 shows that for EL students the high implementing schools growth rate was 5.1 scale score points versus *3.8 points* for low implementing schools – 0.3 scale score points lower than for "all students." In other words, the EL subgroup is disproportionately *penalized* in low implementing Reading First schools. This finding is reproduced in grade 3, less so in grade 4. The lesson is plain. The EL subgroup responds well when Reading First is well implemented but is vulnerable to lower growth rates when Reading First is poorly implemented. It is imperative that Reading First schools with high concentrations of ELs faithfully implement the program.

In general, taking into account differences that pertain to ELs in non-Reading First elementary schools and to performance in grade 4, we see that ELs have shown remarkable growth as a subgroup since 2002 statewide, and that this growth is magnified when Reading First is implemented rigorously. However,

ELs are especially vulnerable when Reading First schools do not faithfully implement the program, a vulnerability that is less pronounced in non-Reading First schools due presumably to the demographic advantage of having much fewer ELs per school.

CST Results for Grade 2 (Table 6.2, Figures 6.2a and 6.2b)

Table 6.2 and the accompanying trend-line charts shows the CST results for grade 2, YIP = 5, Reading First schools. The table includes the starting and ending mean scores for grade 2 in schools that have been in the program for five years. The first column of achievement gains duplicates the "All Reading First Schools" data that is reported in Table 2.3 (p.29). The gains in the four columns headed "English Learner Students" were computed using *only* data for the EL subgroup. English learners in high implementation schools show the strongest achievement gains.

Table 6.2: CST Metric, YIP = 5, Grade = 2

Years in Program: 5 Grade: 2	Reading First Schools				
		English Learner Students			
	All Reading First Schools (All Students)	All Reading First Schools (EL Only)	High Implementation Schools, RFII > 41.4 (EL Only)	Low Implementation Schools, RFII < 36.0 (EL Only)	All Non-Reading First Elementary Schools (EL Only)
Number of Schools	259	244	27	98	2103
% Proficient and Above					
2002	15.4	11.1	9.7	11.6	17.9
2007	34.2	26.8	32.7	24.4	33.7
Change Since Starting Year	**18.9**	**15.7**[b]	**23.0**[ab]	**12.8**[ab]	**15.8**
Mean Scale Score Per Student					
2002	299.8	292.7	290.2	293.3	304.8
2007	324.7	314.6	324.3	310.3	324.4
Change Since Starting Year	**25.0**	**22.0**[b]	**34.1**[ab]	**17.0**[b]	**19.6**

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

Table 6.2 is consistent with summary gains Table 6.1, but highlights several important issues. We see that EL gains are smaller than for "all students" (e.g., 22.0 scale score points versus 25.0 scale score points). However, when comparing Table 6.2 with Table 2.3 from Chapter 2, we see that the ELs actually grew *more* in High Implementation schools than the "all students" population in High Implementation schools. The lower EL gains across the Reading First population are thus a result of very low EL gains in Low Implementation schools. This highlights what we observed in the summary table, that Reading First works for ELs so long as the program is strongly implemented. ELs are much more vulnerable to low growth than the "all students" population when the schools are low implementers of Reading First.

A comparison with Table 2.3 repeats the finding that ELs in non-Reading First schools experienced higher gains than the "all students" population but that this pattern is not reproduced in Reading First schools. As was pointed out, this is probably due to the fact that Reading First schools are composed primarily of ELs whereas ELs are a relatively small minority in non-Reading First schools. In addition, because ELs are particularly sensitive to low Reading First implementation, when there are a large number of low implementing Reading First schools this drags down the overall growth of the Reading First population.

Figures 6.2a and 6.2b show the trend lines for the EL subgroup for grade 2 in the YIP 5 schools, on the "% Proficient" and "mean scale score" achievement metrics. As noted earlier, the trend-lines for non-Reading First schools have been adjusted downward to have the same starting point as "All Reading First Schools."

Figure 6.2a: English Learner CST % Proficient and Above, YIP = 5, Grade = 2

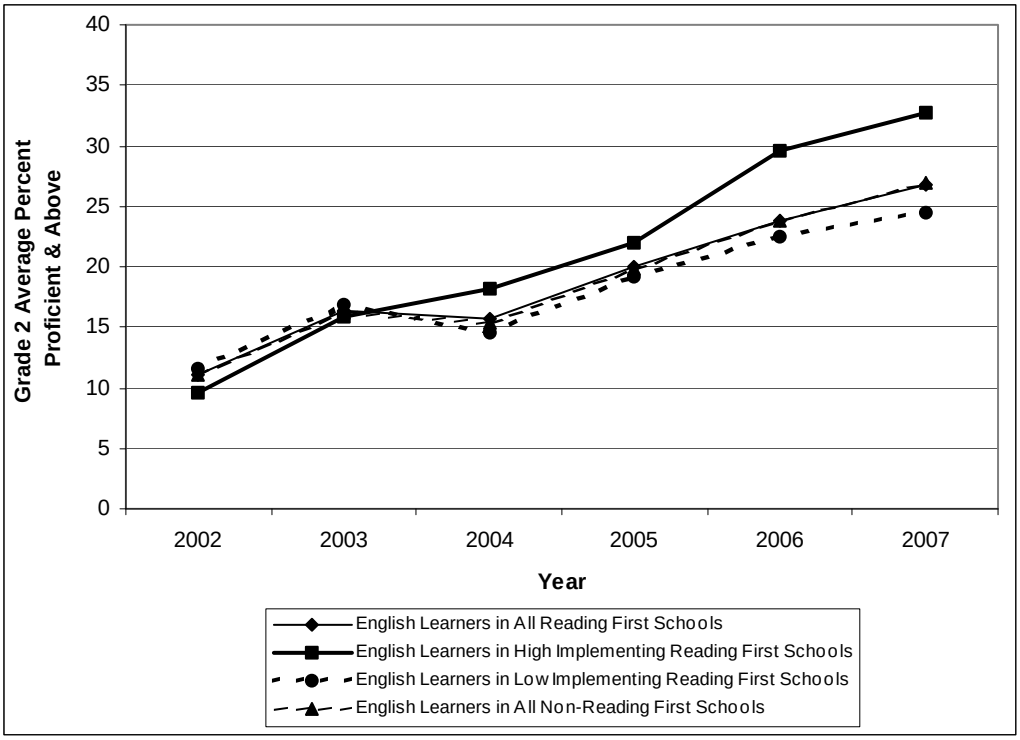
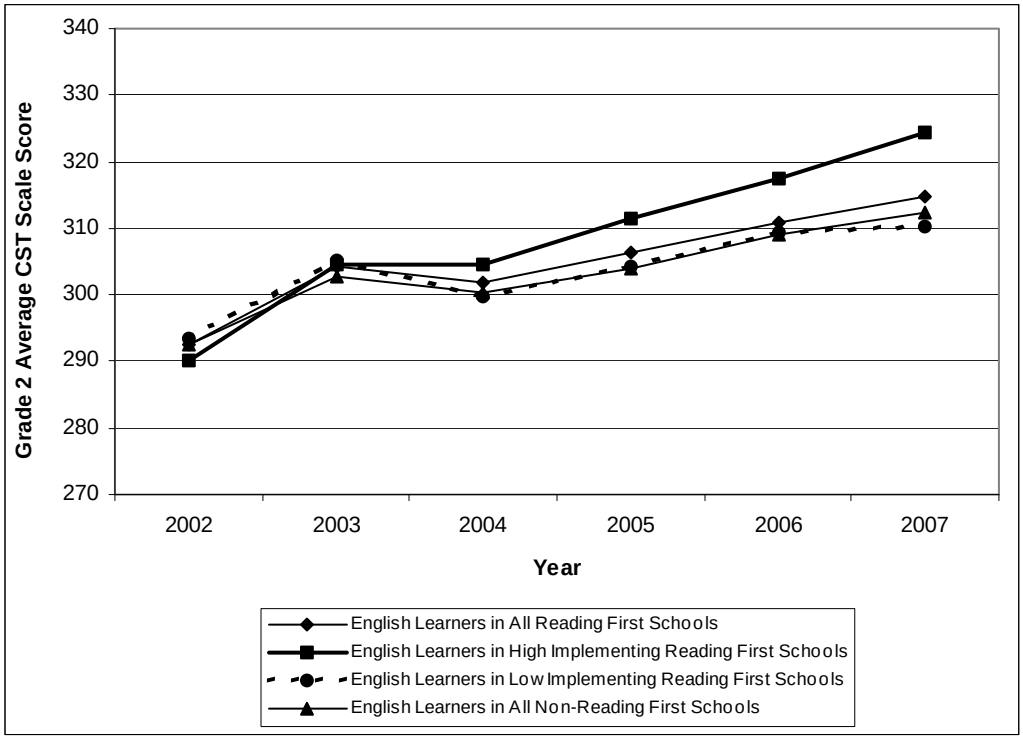


Figure 6.2b: English Learner CST Mean Scale Score, YIP = 5, Grade = 2



The relative steepness of the High Implementing trend-lines in Figures 6.2a and 6.2b emphasize how important high implementation is for the EL subgroup. Without high implementation, Reading First schools are not much more effective than non-Reading First schools for the EL subgroup. When the program is faithfully implemented, schools show remarkable gains in their ability to serve their populations of English learners.

CST Results for Grade 3 (Table 6.3 and Figures 6.3a and 6.3b)

Table 6.3 contains the CST achievement gains for grade 3 English learners in Reading First schools which have been in the program for 5 years. The first column of achievement gains duplicates the "All Reading First Schools" data that is reported in Table 2.4 (p.34). The gains in the four columns headed "English Learner Students" were computed using *only* data for the EL subgroup.

Table 6.3: CST Metric, YIP = 5, Grade = 3

Years in Program: 5 Grade: 3	Reading First Schools				
		English Learner Students			
	All Reading First Schools (All Students)	All Reading First Schools (EL Only)	High Implementation Schools, RFII > 41.4 (EL Only)	Low Implementation Schools, RFII < 36.0 (EL Only)	All Non-Reading First Elementary Schools (EL Only)
Number of Schools	259	239	27	97	2026
% Proficient and Above					
2002	14.8	8.3	6.0	8.8	15.6
2007	20.8	10.0	12.1	9.6	16.9
Change Since Starting Year	**6.0**	**1.7**[b]	**6.1**[b]	**0.8**	**1.4**
Mean Scale Score Per Student					
2002	294.5	283.3	279.9	284.4	297.2
2007	307.4	291.8	295.9	290.3	303.3
Change Since Starting Year	**12.9**	**8.5**[b]	**16.0**[ab]	**6.0**[b]	**6.2**

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

The patterns observed in Tables 6.1 and 6.2 are reproduced here, yielding similar conclusions. However, there are some notable anomalies. For instance, the difference between the "all student" population and the EL subgroup is quite dramatic – a 6.0 gain in "% proficient" for all students versus only a 1.7 percentage point gain for the EL subgroup. Because the corresponding difference is not nearly so pronounced in Table 6.1, which includes schools from all YIPs, it appears that the large difference is peculiar to the YIP 5 schools. YIP 5 includes Los Angeles Unified School District (LAUSD), noted for its extremely high concentration of ELs. Therefore, it is possible that the demographic consequences of high EL concentrations are more pronounced in the YIP 5 schools than in the other YIPs that go into the summary gains Table 6.1. We also cannot ignore the possibility that EL reclassification starts in grade 3 rather than grade 4 for LEAs in this cohort.

Such anomalies aside, however, we find that the implementation effect is prominent in grade 3 and that EL students are well served by Reading First.

Figures 6.3a and 6.3b present the corresponding trend-lines on the "% Proficient" and "Mean Scale Score" achievement metrics. As usual, non-Reading First schools have been adjusted downward.

Figure 6.3a: English Learner % Proficient and Above, YIP = 5, Grade = 3

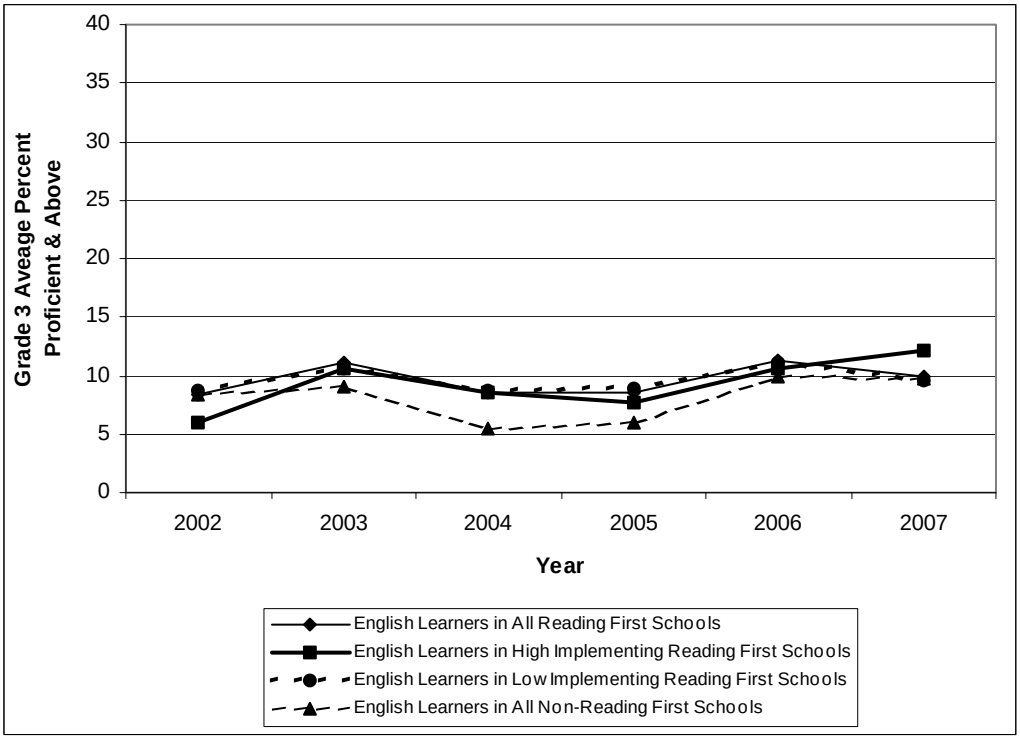
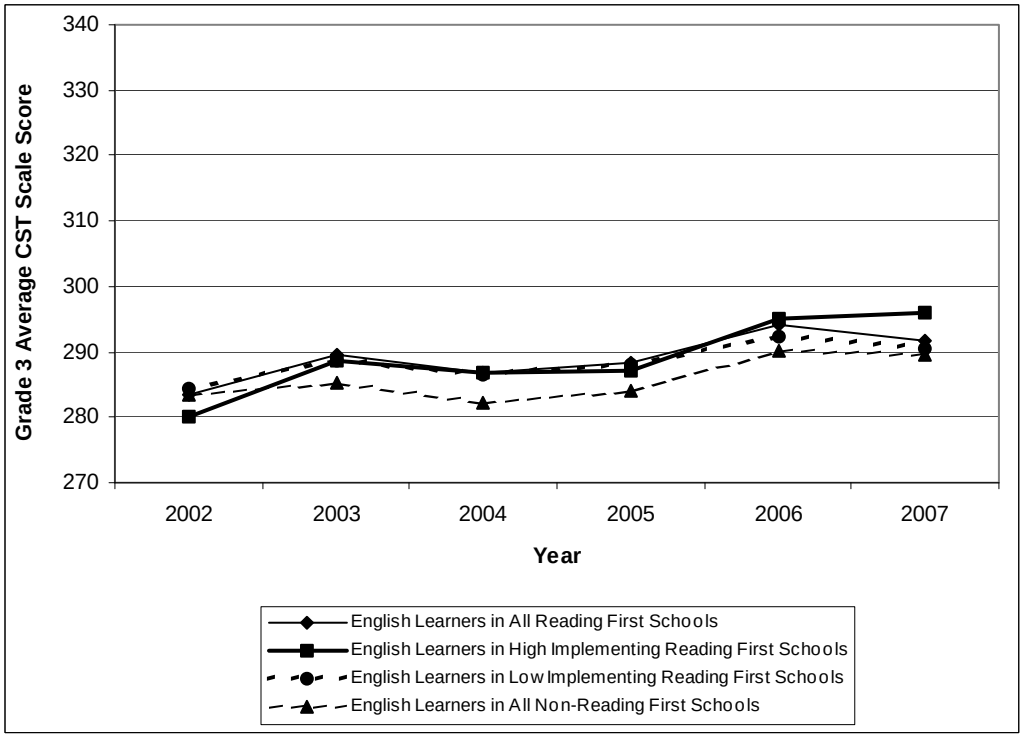


Figure 6.3b: English Learner CST Mean Scale Score, YIP = 5, Grade = 3



While the trend-lines in Figures 6.3a and 6.3b do not show as dramatic an implementation effect as the other grades (for many of the same reasons discussed in detail in Chapter 2), we see a reiteration of the basic finding that High Implementation Reading First schools are much more effective with the EL subgroup over time than Low Implementation Reading First schools and non-Reading First schools.

CST Results for Grade 4 (Table 6.4 and Figures 6.4a and 6.4b)

Table 6.4 reports the CST achievement results for grade 4 English learners in Reading First schools that have been in the program for five years. The first column of achievement gains duplicates the "All Reading First Schools" data that is reported in Table 2.5 (p.40). The gains in the four columns headed "English Learner Students" were computed using *only* data for the EL subgroup.

Table 6.4: CST Metric, YIP = 5, Grade = 4

Years in Program: 5 Grade: 4	Reading First Schools				
		English Learner Students			
	All Reading First Schools (All Students)	All Reading First Schools (EL Only)	High Implementation Schools, RFII > 41.4 (EL Only)	Low Implementation Schools, RFII < 36.0 (EL Only)	All Non-Reading First Elementary Schools (EL Only)
Number of Schools	255[1]	235	25	96	1869
% Proficient and Above					
2002	15.2	6.3	4.6	7.0	11.9
2007	31.3	17.2	17.4	16.1	26.6
Change Since Starting Year	**16.1**	**10.9[ab]**	**12.8[b]**	**9.1[ab]**	**14.8**
Mean Scale Score Per Student					
2002	306.8	292.8	287.3	294.7	303.7
2007	327.3	311.1	311.1	309.6	322.5
Change Since Starting Year	**20.5**	**18.2[b]**	**23.8[b]**	**14.9[ab]**	**18.8**

[a]Significantly different ($p < 0.05$) relative to English learners in "All Non-Reading First Elementary Schools".
[b]Significantly different ($p < .05$) relative to the starting year, i.e., significantly different from a gain of zero.

In contrast to all reading first students, the ELs in Reading First schools – even those in high implementation schools – show lower gains in the "% Proficient and Above" metric than their EL counterparts at non-Reading First schools. The gains in mean scale score show a similar pattern, with the exception that the ELs in high implementation schools have higher gains than their non-Reading First counterparts. It would thus appear that Reading First ELs in Grade 4 are substantially lagging their counterparts in non-Reading First schools.

While we do not yet know why the grade 4 ELs grow so much less than the "all students" population, we hypothesize, as discussed in detail with summary Table 6.1, that it is at least in part an artifact of LEA reclassification policies that reclassify high performing ELs as English-fluent based on their grade 2 and

grade 3 CST results. Thus, high-performing ELs may be systematically under-represented in these grade 4 statistics. This hypothesis is supported by the fact that the anomaly is much more pronounced for the "% Proficient and Above" achievement metric than for the "Mean Scale Score" metric. Students are often reclassified based on whether they scored "Proficient" or above in the previous grade, so this achievement metric would be much more sensitive to reclassification effects. Because the mean scale score metric averages scale scores from all performance levels, including ELs who score Basic or below, reclassification effects would be somewhat dampened. However, without data regarding the reclassification policies of Reading First LEAs, such theories are conjectural.

Figures 6.4a and 6.4b display these anomalies graphically for ELs. It is instructive to compare them to their counterparts in Chapter 2 for all students, Figures 2.5a and 2.5c. The patterns are reversed.

Figure 6.4a: English Learner % Proficient and Above, YIP = 5, Grade = 4

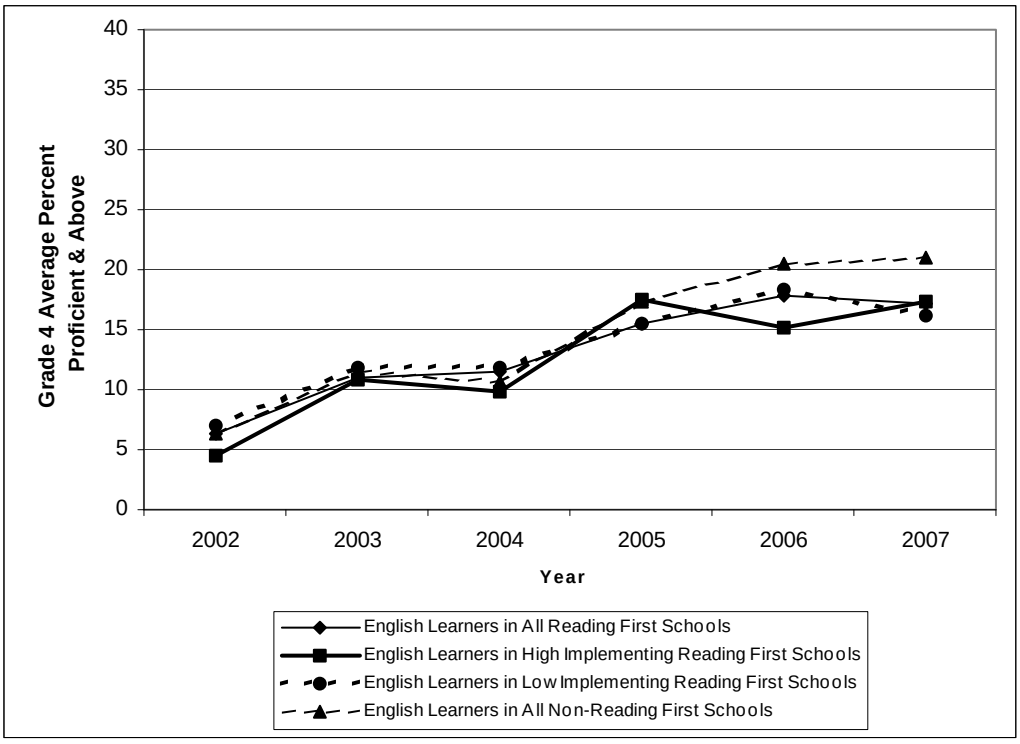
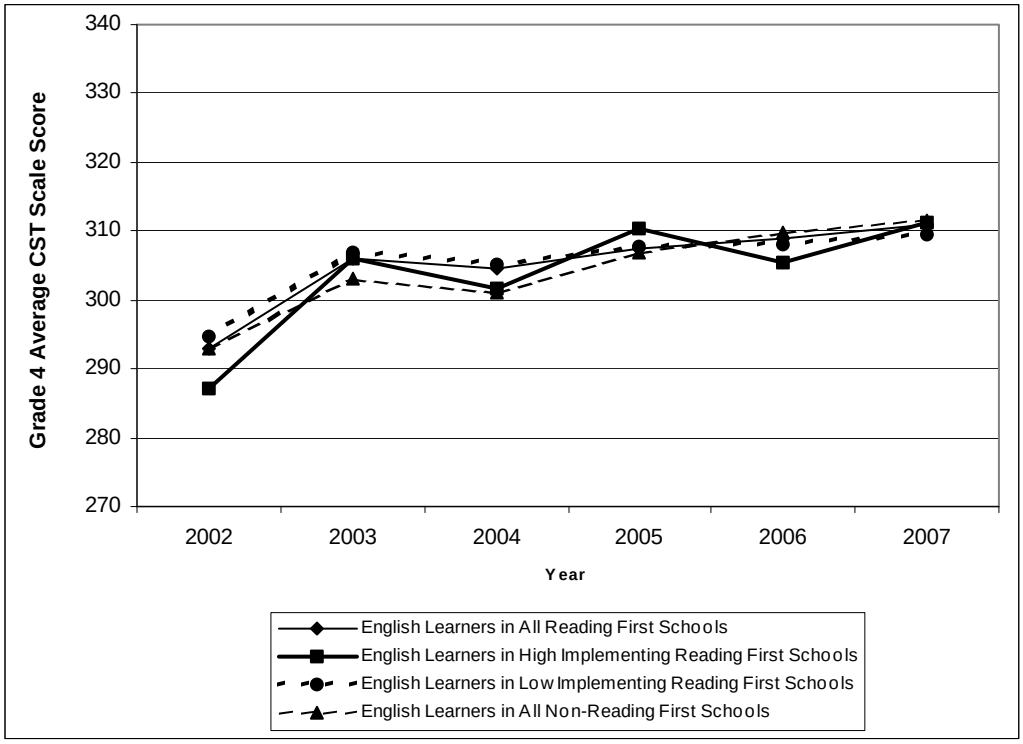


Figure 6.4b: English Learner CST Mean Scale Score, YIP = 5, Grade = 4



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

Teachers, coaches and principals responded to the question, "In what ways has your school's participation in Reading First impacted the learning of English learners in your school? Explain your response." Similar to the format used in Chapters 4 and 5, in this chapter we use qualitative research methodology to examine findings from the open-ended question regarding ELs. This analysis yields the perspectives of school personnel who are most directly involved with implementing the Reading First coaching model.

In this analysis, we first examine the perceptions of teachers, coaches and principals as reported in the open-ended question. We compare the relative perceived importance of resulting categories of responses across respondent groups. Then, we examine differences in perceptions across high implementing and low implementing schools, as determined by the RFII (Reading First Implementation Index, see Chapter 3).

Of the 17,261 teacher surveys collected, 11,466 wrote narrative responses to this question, or 66.4%. Of the 1,028 coach surveys collected, there were 884 narrative comments submitted, or 86.0%. Of the 1,073 principal surveys collected, there were 989 comments submitted, or 92.2%. This response rate is

somewhat lower than that of the open-ended questions discussed in Chapters 4 and 5, but may reflect the fact that some schools have few or no EL students.

Chapter 4 provides a description of the qualitative methodology and a discussion of the limits on generalizability of the findings. We refer the reader to Chapter 4 for an explanation of the coding and categorization procedures.

Relative Importance of Factors Associated with the Impact of Reading First on ELs

"Time and again I have heard teachers claim that at first they thought the reading program was way beyond the capabilities of our EL learners. Yet, as the teachers' implementation and sophistication and refinement of practice increased, they have seen remarkable results in their students' abilities. Reading First provided the focus for reading instruction via the teacher dialogue, content and skills training, and continued self-monitoring of grade-level instruction. It is an excellent opportunity (Response from a Reading First principal)"

This response from a principal on the survey reflects the change in thinking evident in the open-ended survey responses regarding reading instruction for ELs. Many of the comments submitted by teachers, coaches and principals expressed the general positive regard for the support provided by Reading First for teaching ELs. This section describes categories of responses and their relative importance.

The codes (or categories of responses) were sorted according to the frequency with which they occurred and are listed in rank order in Table 6.5. (Descriptors are provided in the following section.) Rankings are listed for the whole data set combined and then for teachers, coaches and principals. This allows the reader to compare the participant groups' perceptions in terms of relative importance of factors in Reading First that impacted ELs. Note that this table depicts the relative frequency with which the codes occurred and should not be interpreted as ratings.

Table 6.5 also lists the percentage of the total codable responses that occurred within each category. The percentages will not total 100% because in many cases, comments were assigned multiple codes. This occurred when a response included multiple ideas or concepts, or when a response could be interpreted as falling within more than one code. Additionally, some responses were not coded at all because they were irrelevant to the question and the purpose of this part of the study. It is important to note that these were narrative responses to open-ended questions rather than prompted or closed-ended questions that would direct participants to respond to specific aspects or issues. Therefore, the percentages represent the proportion of respondents who spontaneously chose to respond to a particular issue. For example, the top-ranked category, "Instruction improved for ELs," occurred in 10.9% of the teacher responses, 50.9% of coach responses and 37.9% of principal responses. This does not mean that approximately 90% of

teachers did not feel that instruction was improved. It only means that 10.9% of the teachers who wrote responses spontaneously elected to write about the improvement of instruction. It is logical that a large proportion of the coach responses would focus on the improvement of instruction because that is related to their primary responsibility as coaches to work for improvement in instruction.

Table 6.5: Rank Order and Percentages of Responses for Categories

Response Category (Code)	All N = 13,339		Teachers N = 11,466		Coaches N = 884		Principals N = 989	
	Rank	%	Rank	%	Rank	%	Rank	%
Instruction Improved for ELs	1	9.9%	4	10.9%	1	50.9%	1	37.9%
Curriculum or Materials for ELs	2	7.2%	1	15.1%	5	18.0%	5	13.9%
Vocabulary or Lang. Improvement	3	6.8%	2	12.7%	7	15.7%	2	18.7%
Achievement Gains or Improvement	4	6.6%	3	12.2%	4	18.6%	3	15.9%
Negative Comment	5	5.9%	5	10.7%	6	17.4%	4	14.4%
Awareness of EL Needs	6	3.9%	7	4.6%	2	23.9%	6	10.4%
Professional Development	7	3.3%	8	3.4%	3	20.9%	7	9.9%
El Guidelines or Handbook	8	3.1%	7	4.6%	9	11.5%	8	9.8%
Small Group Instruction	9	2.1%	9	2.6%	11	8.5%	10	8.3%
No Impact	10	2.1%	6	5.1%	12	3.2%	12	1.4%
Assessment Practices W/ ELs	11	1.7%	12	1.2%	10	10.2%	9	8.4%
Early Intervention Prog. for ELs	12	1.5%	7	3.7%	13	1.2%	11	2.4%
Bilingual/Waiver Classroom Impact	13	1.2%	14	0.3%	7	15.7%	13	1.3%
Transition to English	14	0.7%	10	2.2%	8	14.7%	15	0.0%
Collaboration/Lesson Study	15	0.7%	11	2.0%	14	0.0%	15	0.0%
Accountability	16	0.4%	13	0.9%	13	1.2%	14	0.3%

In Table 6.5, the highest-ranking categories focused on the positive changes that occurred in Reading First classrooms as a result of Reading First. The first two categories focused on improvements in instruction and curriculum. Improving the curriculum and instruction is the central focus of the Reading First initiative. In this dataset, we see evidence that the most important changes occurring as a result of Reading First are carrying over to the EL population. In an unprompted, open-ended question, the largest proportion of spontaneous responses of participants focused on the impact on curriculum and instruction. Some of these comments focused on access issues; that Reading First had made the core curriculum and standards accessible and available for ELs. There was a widespread feeling that, prior to Reading First, schools could not expect ELs to achieve success with grade-level curriculum and standards, and that

Reading First had provided the strategies and materials to move in this direction. Some comments indicated that only when they fully implemented the program did they see substantial improvement. Other comments focused on the quality and importance of research-based instruction and curriculum materials. Additionally, the next two highest occurring categories focused on the improvement of vocabulary and language development or overall academic improvement of ELs. Comments in these two categories indicated that Reading First participants saw noticeable improvement in the students' day-to-day classroom performance as well as in their assessment outcomes. Table 6.6 provides sample comments from all categories.

In Table 6.5, there are some inconsistencies in the rankings across respondent groups. The category, Instruction Improved for ELs, occurred with highest frequency for the all, coach and principal categories, but was 4th highest for teachers. The teacher response category that occurred with highest frequency was Curriculum/Materials for ELs, ranked 5th for coaches and principals. Second highest for teachers and principals, was the improvement of vocabulary and language skills for ELs, but 7th highest for coaches. All groups commented on notable achievement gains for ELs as a result of Reading First. Improved Awareness of EL Needs was ranked second in the coach group, but 7th and 6th for teachers and principals. For other categories, perceptions occurred with the same relative frequency. The following response categories were considered of high importance across rating groups: Instruction Improved for ELs, Curriculum/Materials for ELs, Vocabulary/Language Development, Achievement Gains/Improvement, Negative Comments, Awareness of EL Needs, and Professional Development. The other categories occurred with fairly low frequency and should not be considered to represent a significant finding among participant groups.

Code Characterization

For each code, or response category, in Table 6.6 below, a brief definition is provided along with representative comments from each respondent group. These are listed in the order of frequency occurring within all respondent groups combined.

Table 6.6: Code Descriptions and Representative Comments

Code Description	Reasons and Representative Comments
Instruction Improved for EL s Indicates that Reading First has helped improve instruction and led to better instructional strategies for ELs.	"Front loading sentence frames give Ells great opportunities for oral practice and reading comprehension skills (Teacher)" "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 appropriate scaffolds (Coach)" "It has positively impacted the students because Reading First has helped with guiding my instruction and helped me to re-teach concepts that are missed (Teacher)" "Teachers are more aware that differentiated instruction must be included into their lessons and each lesson must accommodate the ELL (Principal)" "It has provided teachers with the tools necessary to improve their teaching strategies which in turn helps with student progress (Teacher)"
Curriculum/ Materials for ELs Responses show that Reading First has provided curriculum and materials specifically designed to meet the needs of ELs. May mention the Spanish language reading programs.	"Reading First has provided us with additional tools and materials to build and develop learning styles as well as become more sensitive to the cultural differences among our EL students. This helps to ensure the creation of an environment wherein they can be successful (Coach)" "The materials provided by Reading First have given the students tools needed to learn the English language (Teacher)" "Reading First has allowed our English learners to have access to the core curriculum (Principal)" "All students are getting the same material. Also EL students are learning more because there is specific intervention for them in the program and teachers do not have to go outside to get materials specific for them (Coach)" "The additional funding has made possible the purchase of materials to enhance our students' learning (Teacher)"

Vocabulary/ Language Improvement States that the Reading First program has helped ELs to develop vocabulary and oral language skills sometimes referring to English language development	"They have expanded their vocabulary as they are explicitly taught the words, meanings and usage of it (Teacher)" "For those students who started in kindergarten, they have been able to reach benchmark and continue to be successful in future grades. They have been able to move up in ELD levels and re-classify by third grade (Principal)" "One area English learners have difficulty with is vocabulary. Since participating in the Reading First program I feel English learners have improved in this area (Coach)" "Being immersed in the heavy phonics emphasis and vocabulary instruction correlated with the reading gives our English learners a rich language experience (Teacher)"
Achievement Gains or Improvement Results report that Reading First has improved reading achievement for ELs more so than before implementing the Reading First program.	"In my opinion, my school's participation in Reading First has significantly impacted our ELL students. Ell students are learning to read more quickly and gaining more vocabulary because their need for assistance is being targeted (Coach)" "For ELL specifically, we are seeing better skills in reading and comprehension than when we were trying to teach Spanish and transitioning slowly into English. They are learning English skills and maintaining them with this effort (Teacher)" "Standardized testing results have shown an increase in proficiency for EL students (Principal)" "We have seen language arts skills improve across the board. All students have improved their reading and writing since Reading First was implemented (Coach)" "Our English learners have improved and made great progress in reading as a result of the excellent reading program (Teacher)"
Awareness of EL Needs States that Reading First has led to increased teacher awareness of the needs of ELs and how to address them. May also include increased awareness of state standards for ELs.	"We have a high number of Ells at our site and Reading First has provided increased attention to the needs of Ells (Coach)" "We're aware of the needs of our EL students and try to adapt our teaching to help those students (Teacher)" "Reading First has impacted our English learners because our level of conversation has changed as to how we are going to meet this particular subgroup's needs. Reading First keeps the struggling readers in the forefront so our staff is directed in providing effective strategies and interventions to meet their needs. We are seeing positive results (Principal)" "Reading First has impacted the ELL learners because we can focus on them and see where their strengths and weaknesses are based upon the SCOES (assessments) (Teacher)" "Certainly the tiering has brought the struggle of ELL to the forefront. Teachers are more aware of the difficulty facing these students and are working together to help meet their needs in a more focused way (Coach)"

Professional Development Indicates that professional development for teachers, coaches and principals has focused on EL needs and has helped personnel to acquire the knowledge and skills needed to address such needs.	"We have had more extensive training in meeting the needs of our English learners. Through lesson studies we have targeted our areas of need and have received professional development to help us reach our teaching goals (Teacher)" "We have provided some staff development with in-class follow up addressing the vocabulary needs and reading comprehension of English language learners. This year our goal is to provide this training to all our teachers (Coach)" "We have been provided with knowledge and strategies to help us work with English learners. The principal, coaches and teachers have been given staff development and all have worked to implement what they have learned (Principal)" "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 (Coach)"
El Guidelines or Handbook Indicates that Reading First has provided schools with guidelines or a handbook that provides strategies and ideas for meeting the needs of ELs.	"The implementation of the ELL handbook has helped most EL students grasp the material in a way that is easier than I would have been able to do it (Teacher)" "The materials such as the English Learner Support Guide, Reteach, Intervention, etc. have been a … great assistance to our English learners (Teacher)" "English learners' needs are taken into consideration more because of the availability of the ELL handbook and giving students access to the core program (Coach)" "The English Learner Support Guide was very useful for providing specific, detailed vocabulary support for EL students (Principal)"
Small Group Instruction States that ELs receive support with reading or language development through opportunities for small group instruction or intervention.	"Our first grade team has agreed to divide children by CELDT (language proficiency) level and offer differentiated, small group instruction based on those levels for the next school year (Teacher)" "I feel the EL students are comprehending more when they are put in smaller groups during Universal Access time (Coach)" "In small groups, teachers focus on ELs during Universal Access time, which provides more structured and focused access to the core curriculum by using targeted EL strategies (Principal)" "Teachers have allotted time to work with EL students in small groups on specific language development skills that will help them more successfully access the content (Principal)"

No Impact Indicates that Reading First has not had an impact on or changed the instruction provided for ELs.	"I am not sure Reading First has made an impact on English learners. We have always differentiated instruction for ELs and have always had an English language period of 45 minutes (Teacher)" "The English learners are not being addressed any differently than the English only students are. No consistent differentiation is taking place (Teacher)" "I don't think Reading First has made an impact on EL learners because the pacing plan remains the same regardless of whether the students are English only or English learners (Coach)" "Reading First has not greatly impacted our English learners. Whether we have Reading First or not, we would be using Open Court, our district adoption (Principal)" "Our participation has not changed our approach to English learners (Principal)"
Assessment Practices W/ ELs States that Reading First assessment practices or tools are beneficial for teaching EL students, the data provide meaningful information, or the assessments allow schools to track progress.	"We keep the ELL students in focus and discuss them whenever we have collaborative meetings. We study their data carefully and provide lessons to meet their specific needs (Principal)" "ELL students are learning to read more quickly and gaining more vocabulary because they are being targeted for assistance and their assessment results are checked to assure they are on track (Coach)" "Using data on a regular basis and setting performance goals for all students has caused teachers to better focus and cognitively plan for ELLs (Coach)" "We have created an assessment-driven program that helps identify the needs of our EL students so we can target and address areas of improvement (Teacher)"
Early Intervention for ELs States that Reading First has provided impetus for catching reading difficulties early and providing appropriate intervention for ELs.	"EL students have received extra support and early in their development (Teacher)" "Reading First has provided early intervention opportunities for our EL students (Teacher)" "Teachers are focused on how to do immediate corrective intervention to close the gaps in English skills for EL students. The principles of 'teach, practice, apply' and universal access have been the keys to upward movement of our EL students (Principal)"
Bilingual/Waiver Classroom Impact Though the question asked about ELs in general, some comments specifically mentioned the impact on bilingual teachers or instruction in bilingual or waivered classrooms.	"Our teachers are teaching their Language Arts in Spanish, but beginning in first grade, we have implemented the parallel English instruction in the area of phonics (Coach)" "The instruction in waivered classrooms is as top-notch as in non-waivered classrooms (Coach)" "Reading First has supported the bilingual instruction as well as the English (Teacher)"

Transition to English Indicates that the Reading First program has improved the process of transitioning EL students from their primary language into English instruction	"There are fewer EL students because they transition out of the ELD program more quickly (Teacher)" "The evidence of the impact of Reading First is the increased numbers of students who FEP (transition to Fluent English Proficient) out of our ELD program (Coach)" "Reading First has provided English learners the opportunity to learn the basics of English better than any other program we have implemented before (Principal)"
Collaboration/Lesson Study States that the process of collaboration and collaborative planning has improved the instruction specifically for ELs.	"Reading First has provided collaboration time with our Reading First meetings to discuss ways to help our second language learners achieve. In these meetings, we have been able to decide on a focus or target area and implement plans made together to accomplish the goals. We then can revisit and assess the results to help guide us further or modify our plans. Through these meetings and the professional development, we have been able to really understand the program components on a deeper level and raise our EL achievement (Teacher)"
Accountability States that the Reading First program has led to an increased sense of accountability or holding the students to high standards	"Reading First makes us all more accountable and schedule more time to meet the needs of our EL learners (Teacher)" "Our EL students and their parents have become more accountable for reading success (Teacher)" "In the past, I think teachers had somewhat given up the responsibility of helping EL students in their classrooms. Now, many teachers are doing in-class 30 minutes of universal access in addition to a district-mandated 30-minute support block which moves EL students to different groups. The in-class universal access time is helping teachers take more responsibility for those students (Coach)"

Negative Comment Any negative comment or criticism regarding Reading First and ELs is indicated here.	***More Instructional Time*** "They also need more time in the morning to work on transferring skills into English. This is a tall order and many students are capable but simply need more instructional time. If anyone is listening we need more time in the instructional day. That is the biggest complaint amongst our staff. They feel that they can do this but need more time in the day. Yes, a longer day." ***Curriculum*** "The reading writing connection is not strong enough (Coach)" "The program comes with supplemental support but the teachers still have to sort through it to find appropriate material with the help of the coach (Coach)" "There is not a lot of ELD support written within the program (Teacher)" ***Concern for Struggling ELs*** "Early on in our implementation of OCR, not enough attention was given by program professional development to the unique needs of ELs. This was true especially in regard to lesson pacing and whole group instruction. However, now the experiences of EL learners involved in this program have been more intensely considered (Coach)" "The same expectations for all sometimes backfires. We need a better intervention program for k-3 ELs (Coach)" "It is difficult in the beginning for ELs. They do not have the oral language to be successful in the program (Principal)" ***Pacing Plan*** "We are in our second year. I don't think Reading First has made a difference for ELs because the pacing plan remains the same regardless of whether the students are English-only or ELs. There is no differentiation. I believe it is best to teach with quality in mind rather than quantity (Teacher)" "For the majority of ELL students the program goes way too fast for them. I find it very frustrating both for the teachers and the students (Teacher)" ***Lack of Systematic English Language Development*** "The lack of rigorous ELD tied to the core curriculum is the primary obstacle to student success in meeting English Language Arts Standards (Coach)" "Overall I feel like RF has not adequately addressed the needs of ELs. There needs to be more talk of frontloading vocabulary and language for ELs and more of an understanding of the separation between content area teaching (ELA) and language teaching (ELD)(Principal)" "I have to supplement the ELD section. It is more work for me (Teacher)"

Conclusions

In conclusion, this chapter finds that achievement gains for English learners in Reading First schools are positive for grades 2, 3 and 4. Additionally, achievement gains are higher for English learners (ELs) in Reading First schools than for English learners in non-Reading First schools for grades 2 and 3.

Implementation is an important factor for ELs as it is in general for Reading First schools. Achievement gains are higher for ELs in high implementing Reading First schools than ELs in low implementing Reading First schools and non-Reading First schools. The EL subgroup is more impacted by differences in Reading First implementation than the student population as a whole. The EL subgroup in low implementing Reading First schools is particularly at risk for low growth, whereas ELs in high implementing Reading First schools often grow more than the student population as a whole.

The effect of Reading First implementation on EL achievement in grades K-3 is reproduced for ELs in grade 4. However, the non-Reading First EL subgroup shows higher growth than the EL subgroup in Reading First schools. We hypothesize that this may be a statistical artifact of EL reclassification criteria that reclassify ELs to English-fluent status based on grade 3 CST results.

In open-ended survey comments, teachers, coaches and principals reported overall positive regard for the Reading First program and its appropriateness and support for ELs. Teachers, coaches and principals reported significant improvement in the curriculum and instruction for EL students due to their schools' participation in Reading First. Specifically, teachers, coaches and principals noted evident and significant improvement in the vocabulary, language development and reading achievement of ELs as a result of their schools' participation in Reading First.

Though there was generally a positive perception of the impact of Reading First on EL students, some participants expressed concerns regarding the amount of time needed to effectively teach ELs, specific aspects of the curriculum and materials, the pacing of instruction for ELs and the need for more systematic English language development to better meet the needs of ELs.

References

Baker, S. K., Gersten, R., Haager, D. & Dingle, M. (2006). Teaching Practice and the Reading Growth of First-Grade English Learners: Validation of an Observation Instrument. *Elementary School Journal, 107 (2),* 199-220.

Chiappe, P. & Siegel, L. S. (2006). A longitudinal study of reading development of Canadian children from diverse linguistic backgrounds. *The Elementary School Journal, 107*(2), 135-152.

Chiappe, P. Siegel, L. S., & Wade-Wooley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading, 6*, 369-400.

Gersten, R. (2006). Introduction: Scientific research on English learners. *The Elementary School Journal, 107*(2), 133-134.

Gersten, R., & Baker, S. (2000). What we know about effective instructional practices for English-language learners. *Exceptional Children, 66*, 454-470. Gersten, R., Baker, S. K., Haager, D. & Graves, A. (2005). Exploring The Role of Teacher Quality in Predicting Reading Outcomes for First Grade English learners: An Observational Study. *Remedial and Special Education, 26*, 197-206.

Graves, A., Gersten, R., & Haager, D. (2004). Literacy Instruction in Multiple Language First Grade Classrooms: Linking Student Outcomes to Observed Instructional Practice. *Learning Disabilities Research and Practice, 19*, 262-272.

Haager, D., Gersten, R., Baker, S., & Graves, A. (2003). The English-Language Learner Classroom Observation Instrument: Observations of Beginning Reading Instruction in Urban Schools (p. 111-144). In S.R. Vaughn & K.L. Briggs (Eds.), Reading in the Classroom: Systems for Observing Teaching and Learning. Baltimore MD: Brookes Publishing.

Klingner, J. K., Sorrells, A. M., & Barrera, M. T. (2007). Considerations when implementing response to intervention with culturally and linguistically diverse students. In D. Haager, J. K. Klingner & S. Vaughn (Eds.), *Evidence-based reading practices for response to intervention* (pp. 223-244). Baltimore, MD: Paul H. Brookes Publishing.

National Center for Education Statistics (2007). *The Nation's Report Card: Reading*. Washington, DC: Institute for Education Sciences, U. S. Department of Education. Available at http://nces.ed.gov/nationsreportcard/pdf/dst2007/2008455_1.pdf.

Snow, C. (2006). Cross-cutting themes and future research directions. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners (pp. 631-651)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Williams, T., Hakuta, K., Haertel, E., et al. (2007). Similar English Learner Students, Different Results: Why Do Some Schools Do Better? A follow-up analysis, based on a large-scale survey of California elementary schools serving low-income and EL students. Mountain View, CA: EdSource.

Chapter 7: Impact of Reading First on Waivered Classrooms

California's Proposition 227 requiring instruction in K-12 schools to be provided in English allows the option for parents to sign a waiver indicating that they prefer their children to be educated in their primary language. Statewide, the majority of waivered K-3 classrooms are providing instruction in Spanish. Additionally, AB 1485 in 2003 mandated that Spanish language instructional materials and support for teachers be available in Reading First schools. This bill also specified that students must be tested in English by the end of third grade. Thus, waivered classrooms must include a transition from Spanish to English instruction during the K-3 years.

In the 2004-2005 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. 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*. The goal is for students receiving bilingual reading instruction in Spanish and English to become proficient in English by the end of grade 3 as evidenced by the Standardized Testing and Reporting (STAR) test. 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. The Reading First program included professional development for coaches, teachers and principals regarding the use of the Spanish language materials and assessments were developed in Spanish to monitor student progress.

The purpose of this chapter is to examine the efficacy of the waivered classroom option as part of the Reading First program for English learners (ELs). Chapter 6 reported the achievement results for the EL subgroup and we include relative findings extracted from that chapter here. We also examine the responses of participants to an open-ended question that provided teachers, coaches and principals opportunities to express their views of the impact of the Reading First program on waivered classrooms.

This chapter yields the following key findings:

- English learners in non-waivered classrooms show significantly higher grade 2 and grade 3 STAR scores than English learners who have been in waivered classrooms for 2 or 3 years.
- Many participants with experience in waivered classrooms served by the Reading First program expressed positive perceptions of the program.

- Participants generally had positive opinions of the state adopted curriculum materials used in waivered classrooms but expressed concerns regarding grammatical or typographical errors or problems with translation from English to Spanish in the materials.
- Participants perceived that Reading First has resulted in improved outcomes for EL served in waivered classrooms, primarily as a result of setting high expectations and accountability for ensuring students are proficient by the end of third grade.
- Though Reading First support is attributed to facilitating the transition from Spanish to English instruction for EL students in waivered classrooms, participants expressed a need for further guidance on how to effectively conduct the transition.

Data Sources

For this chapter of the evaluation report, we compare student achievement results of students in waivered Reading First classrooms with achievement results of English learners in non-waivered Reading First classrooms to determine their relative efficacy.

To gather additional information about the impact of Reading First on ELs, teachers, coaches and principals had the opportunity to write in responses to an open-ended question, *"If you have waivered classrooms in your school, where instruction is provided in Spanish, what is your opinion of the impact of the Reading First program on the instruction and learning of the students in waivered classrooms? (If you do not have direct experience with waivered classrooms, please leave this question blank.)"* The responses were compiled by respondent group in a text file and used in a qualitative analysis, described later in this chapter. In this chapter, we do not compare perceptions at high and low implementation sites due to the small number of schools with waivered programs that fell into the low implementation category.

English Learner Student Achievement

Tables 7.1 and 7.2 report the percentage of English learners from waivered and non-waivered classrooms who scored Proficient or above on the grade 2 and grade 3 CSTs, as well as their mean scale scores. English learners are defined to be from a waivered classroom if they received the Spanish form of the 6-8 Week Skills Assessments during the 2006-07 school year and if they attended such a classroom for 2 or 3 years. English learners are defined to be from a non-waivered classroom if, as of the 2006-07 school year, they have never attended a waivered classroom.

Table 7.1: Waivered vs. Non-waivered EL students, Grade 2 CSTs

Grade 2 English learners	ELs in Waivered Classrooms 2 or 3 years, 2006-07	ELs in Non-Waivered Classrooms, 2006-07
Number of EL Students	4,386	29,637
% Proficient and Above	15.5[a]	26.5
Mean Scale Score	297[a]	315

[a] Significantly lower than ELs in non-waivered classrooms, $p < 0.05$.

Table 7.2: Waivered vs. Non-waivered EL students, Grade 3 CSTs

Grade 3 English learners	ELs in Waivered Classrooms 2 or 3 years, 2006-07	ELs in Non-Waivered Classrooms, 2006-07
Number of EL Students	2,412	26,721
% Proficient and Above	8.2[a]	10.7
Mean Scale Score	285[a]	293

[a] Significantly lower than ELs in non-waivered classrooms, $p < 0.05$.

Tables 7.1 and 7.2 show that English learners in waivered classrooms do not score as high on the grade 2 and grade 3 CSTs as the English learners in non-waivered classrooms do. In grade 2, the non-waivered percent proficient and above is 11 percentage points higher than the waivered percent proficient and above, translating to an 18 scale score point advantage. In grade 3 the differences are less dramatic, symptomatic of the grade 3 CSTs in general, but the pattern is similar. ELs in non-waivered classrooms score higher than ELs in waivered classrooms.

While the pattern of higher achievement of ELs in non-waivered classrooms seems clear, it is less clear what the cause of this pattern is. The simplest hypothesis is that reading instruction in a student's primary language is less effective than instruction in an English-only learning environment (immersion). However, this is a complex issue and other hypotheses have not been ruled out. For instance, it may be that the Spanish versions of the Open Court and Houghton-Mifflin reading programs are not as well-written or conceived as the English versions, a finding reported in the qualitative analysis in a later section of this report. Some teacher comments suggest that the Spanish versions of the state-adopted programs do in fact suffer from typographical errors and translation problems. On the other hand, teachers generally found the Spanish materials to be adequate. Another possibility is that the teachers of waivered classrooms may not be as generally effective as their colleagues in non-waivered classrooms, a factor that is not possible to address in this report.

It is also possible that the two student samples are not fully comparable. For instance, our study is forced to assume that the ELs in waivered and non-waivered classrooms enter Kindergarten at roughly the same average level of English ability. Perhaps students in waivered classrooms come from different socio-economic backgrounds. Perhaps they are dominated by a migrant population that does not receive the same degree and consistency of educational exposure. Perhaps their parents are not as educated or as motivated. Given the limitations of the data that are available to us, we have no way to rule out these hypotheses. Nonetheless, the size of the waivered/non-waivered difference is sufficiently large to support a strong preliminary presumption that instruction in one's primary language is less effective in teaching English to ELs than immersion in English-only classrooms based on kindergarten through grade 4 findings.

Our findings do not allow us to make conclusions regarding when students should transition from waivered to non-waivered classrooms. However, our data suggests that CST scores are slightly higher for students who have been in a waivered classroom for 3 years than for 2 years, though the difference is not large. (Of course, students who have been in a waivered classroom for 0 years – i.e., are in non-waivered classrooms -- score highest of all, confounding efforts to infer a positive relationship between performance and years in a waivered classroom.)

Note that these findings only compare two types of Reading First programs. They do not assess the impact that Reading First has on bilingual education relative to non-Reading First schools. For that, we rely on qualitative data.

Participants' Perceptions of the Impact of Reading First on Waivered Classrooms

This section reports the qualitative analysis of an open-ended question included on the survey regarding waivered classrooms. Teachers, coaches and principals responded to the question, *"If you have waivered classrooms in your school, where instruction is provided in Spanish, what is your opinion of the impact of the Reading First program on the instruction and learning of the students in waivered classrooms? (If you do not have direct experience with waivered classrooms, please leave this question blank.)"* Similar to the format used in Chapters 4, 5 and 6, in this chapter we use qualitative research methodology to examine findings from the open-ended question regarding waivered classrooms. This analysis yields the perspectives of school personnel who have experience with waivered classrooms.

In this analysis, we first examine the perceptions of teachers, coaches and principals as reported in the open-ended question. We compare the relative perceived importance of resulting categories of responses across respondent groups. The reader is referred to Chapter 4 for information about the nature and benefits of qualitative methodology. Chapter 4 provides a description of the qualitative methodology and

a discussion of the limits on generalizability of the findings. We refer the reader to Chapter 4 for an explanation of the coding and categorization procedures.

Of the 19,362 surveys collected in total, 11.8% respondents provided written comments to the open-ended question regarding waivered classrooms. Of the 17,261 teacher surveys collected, 1,772 wrote narrative responses to this question, or 10.3%. Of the 1,028 coach surveys collected, there were 246 narrative comments submitted, or 23.9%. Of the 1,073 principal surveys collected, there were 260 comments submitted, or 24.2%. This response rate is lower than that of the open-ended questions discussed in Chapters 4, 5, and 6 due to the fact that respondents were directed to refrain from responding if they did not have direct experience with waivered classrooms.

Factors Associated with the Impact of Reading First on ELs

There was a generally positive impression of the impact of Reading First on waivered classrooms in California. This comment from a teacher reflects the enthusiasm for Reading First and the feeling that the support is welcomed by those responsible for providing the instruction:

"I teach a waivered biliteracy class and I think that having this program in Spanish is VERY valuable. My students are learning to read in Spanish quicker than I've ever seen in the 11 years I've been a bilingual teacher. This allows them to become English readers much quicker. My district has been working hard to teach us how to teach transferability so I am seeing more biliterate kindergarten students than ever before. I think it's great! (Open-ended response from a Reading First teacher)"

The value of the explicitness of instruction as well as the structure and guidance are evident in this and many other comments. Despite the enthusiasm and positive regard for the program expressed by many, the topic of waivered classrooms generated more negative comments than any other open-ended question. It appears that though Reading First has brought needed resources and support to the bilingual programs around the state, improvement in some areas is still needed. For example, there were positive and negative opinions about the curriculum materials (See Tables 7.3 and 7.4 below). Participants were generally pleased to receive much-needed curricular materials. The negative comments focused more on translation errors or difficult vocabulary in the materials rather than generally negative perceptions about the instructional approach or curriculum.

The goal in waivered programs is for students to be proficient in academic skills in English by the end of third grade. AB 1485 requires that schools use the English state assessments for EL students receiving instruction in waivered programs. Generally, participants felt that having the professional development, curriculum and support provided by Reading First made it feasible to transition students to English competency by the end of third grade. They felt that the AB 1485 requirements set the expectation and the

Reading First program provided needed support. However, many comments expressed concern about the transition in preparation for end-of-third-grade English testing. They expressed the need for more explicit guidance on how to do so.

This section describes categories of responses and their relative importance. Table 7.3 provides a listing of the codes, or categories of responses, in rank order in (descriptors are provided in the following section). Rankings are listed for the whole data set combined and then for teachers, coaches and principals. This allows the reader to compare the participant groups' perceptions in terms of relative importance of the response categories relative to waivered classrooms. Note that this table depicts the relative frequency with which the codes occurred and should not be interpreted as ratings. Table 7.3 also lists the percentage of the total codable responses that occurred within each category. Note that the percentages will not total 100% because in many cases, comments were assigned multiple codes. This occurred when a response included multiple ideas or concepts, or when a response could be interpreted as falling within more than one code. Additionally, some responses were not coded at all because they were irrelevant to the question and the purpose of this part of the study.

The responses to this open-ended question reflect unprompted reactions of participants. The responses were not prompted and there were no choices to select from that would direct participants' thoughts. Therefore, the percentages represent the proportion of respondents who spontaneously chose to respond to a particular issue. For example, with regard to participants' perceptions of the materials, 15.4% of the responses (in the All Participants column) were coded as "Materials Positive" while 6.9% were coded as "Materials Negative." These figures do not add to 100% because there were many responses that did not mention materials at all.

In this table, we see general agreement across respondent groups for the codes of highest relative importance. However, notable differences occurred. For principals, the code Same as Regular Classes occurred with higher frequency than for teachers or coaches. For coaches, two codes occurred with relatively higher frequency than for teachers and principals: Equity of Program and Professional Development. The following response categories were considered of high importance across rating groups: Program Positive, Materials Negative, Academic Positive, and Transition Easier. Other categories occurred with varying frequency across groups. A few categories occurred with low frequency and should not be considered to represent a significant finding among participant group: Suggestions for Improvement, Academic Negative, Describe their Program, Transition Negative, Assessments Negative and Time is Problem.

Table 7.3: Rank Order and Percentages of Responses for Categories

Response Category (Code)	All N = 2,278		Teachers N = 1,772		Coaches N = 246		Principals N = 260	
	Rank	%	Rank	%	Rank	%	Rank	%
Program Positive	1	33.9%	1	34.7%	1	31.3%	1	31.2%
Materials Negative	2	15.4%	2	17.7%	3	17.1%	2	14.2%
Academic Positive	3	13.9%	3	15.3%	2	21.1%	3	13.8%
Transition Easier	4	10.0%	4	12.9%	4	16.3%	5	11.9%
Bilingual-Biliteracy Positive	5	8.0%	5	8.8%	8	5.3%	6	8.1%
Materials Positive	6	6.9%	6	8.4%	6	9.3%	11	5.0%
Bilingual vs English-Only	7	6.5%	8	6.8%	12	2.8%	12	3.8%
Equity of Program	8	4.6%	7	7.3%	5	11.8%	10	5.8%
Same as Regular Classes	9	4.1%	11	3.4%	9	4.9%	4	12.3%
Suggestions for Improvement	10	4.0%	13	2.8%	7	6.1%	12	3.8%
Academic Negative	11	3.9%	10	3.7%	11	3.3%	14	3.1%
Describe their program	12	3.6%	9	4.1%	9	4.9%	8	6.5%
Transition Negative	13	3.6%	12	3.0%	7	6.1%	7	7.7%
Assessments Negative	14	2.9%	15	2.6%	13	0.4%	15	1.5%
Time is Problem	14	2.9%	11	3.4%	10	3.7%	13	3.5%
Professional Development Positive	15	2.8%	14	2.7%	5	11.8%	9	6.2%

Code Characterization

For each code, or response category, in Table 7.4 below, a brief definition is provided along with representative comments from each respondent group. These are listed in the order of frequency occurring within all respondent groups combined. Only codes that occurred within 10% or more of a respondent group are described.

Table 7.4: Code Descriptions and Representative Comments

Code Description	Reasons and Representative Comments
Program Positive Expresses a general positive opinion about Reading First and its impact on English learners, specifically for waivered classes, or regarding Spanish language instruction in reading as part of Reading First	"As the Spanish teacher for kindergarten, I feel the impact of RF programs is vital for our students to succeed in both languages (Teacher)." "I think that adopting *Foro Abierto* for K-5 was an excellent decision. This will provide consistency and coherence to the Language Arts program. Teachers will claim that it is not good...but from my perspective it creates uniformity and consistency (Principal)." "Well, based on our data, our waivered students are experiencing greater success than many of our non-waivered classes. I have also been told that we have some of the highest scores in the district when it comes to Spanish language arts assessment. I think the fact that our waivered teachers attended the 5-day training the last two years has impacted their delivery tremendously (Coach)." "Reading First has been beneficial for the waivered classrooms because the teaching strategies are consistent in either classroom regardless of the language. The teachers benefit from the professional development planning time lesson studies etc. as do all staff. Support from the coaches is available to all teachers and this empowers everyone (Principal)."
Materials Negative Indicates a negative perception about curriculum materials used in waivered classrooms as part of Reading First	"Because of the number of errors in reading books, errors in the practice book pages, and the fact that the program is a direct translation of English puts the *Lecture* classes at a severe disadvantage (Teacher)." "*Foro Abierto* was implemented this past year and teachers are not happy with quality level of the program (Principal)" "The translations into Spanish have to improve. There are a lot of spelling, syntax and grammar mistakes in the materials in Spanish. Also, assessment questions cannot be literally translated from English. That seems to be the case in some assessments. Some of those questions don't make sense. Also, it seems to me that the words in the vocabulary lists are simply translated from the English version. If there is an English vocabulary word that's appropriate for a second grader, that doesn't mean that the translation into Spanish is also appropriate for a second grader (Teacher)." "I have listened to major complaints that the program uses language from other countries that have little to do with either Mexican or High Spanish (Coach)."
Academic Positive Expresses a positive opinion about the academic outcomes or impact of RF on academic gains, including reading, for Spanish speaking students	"I think that the impact [of RF] is positive. Our students who are receiving Spanish instruction are gaining the knowledge and skills that they need to succeed not only in Spanish but in English as well (Teacher)." "Students that have been provided instruction in Spanish are developing skills in Spanish that are being transferred into English. These students have shown great progress due to the fact that they have been able to develop a strong foundation in a language they understand (Principal)" "I think that the impact is positive. Our students who are receiving Spanish instruction are gaining the knowledge and

	skills that they need to succeed not only in Spanish but in English as well (Teacher)." "In general, students in my class are becoming better readers with this curriculum as opposed to the program that we used to use (Teacher)" "The Bilingual students are learning to read and write in a more systematical way using the program. As in the English classes, using the program gives consistency throughout the grade levels and allows for conversation among the teachers about theme-based instruction and strategies for teaching the program (Coach)."
Transition Easier Indicates Reading First helps EL students in waivered classrooms to transition to English instruction. The transition from Spanish to English instruction is easier because of what is taught or how it is taught	"Students are more successful in their learning because they understand the core curriculum. As the students learn English, they transfer the learned information (skills, concepts, vocabulary, etc.) allowing them to build academic experiences that they will so much need in the later years of school (Teacher)." "It has positively impacted instruction for our 2nd grade waiver class. After 2nd grade, we don't have any waiver classes. English Language Arts is formally introduced to these students in the second semester. Reading First support has been key in the facilitating the transition for these students into English Language Arts classes in the 3rd grade (Principal)." "I am a Bilingual second grade teacher who has been teaching *Foro Abierto* this academic year. I find that students who are provided instruction in their primary language (Spanish) receive a foundation in reading. When given an ELD program and transferability strategies, the students are able to transfer or read in English more quickly and efficiently. The training and strategies given to me through Reading First have definitely helped me with my instruction (Teacher)." "The children in the bilingual programs consistently have higher scores than those in the English classrooms where the majority of children are still English learners. The *Lectura* program supports the child's learning in their home language and allows them to transfer these skills to English (Coach)." "The biggest plus in my opinion is the continuity of instruction and the similarity of skills being taught in English and in Spanish. This allows for easier transfer to English. Students in the Waiver programs generally outperform students in the English programs on the SCOE assessments because they are learning in their primary language and because Spanish is so phonetically regular (Coach)."
Bilingual-Biliteracy Positive Expresses a positive opinion about the value or merits of teaching children to be bilingual or biliterate	"The instruction on the children's primary language helps them to succeed in English because they learn the vocabulary and language terms in the language they speak. Later on they transfer everything they know to the second language (Teacher)." "Based on our data, our waivered students are experiencing greater success than many of our non-waivered classes. I have also been told that we have some of the highest scores in the district when it comes to Spanish language arts (Coach)"

Materials Positive Expresses an opinion that the curriculum materials used in waivered classrooms are helpful or appropriate for Spanish speaking students	"The *Lectura* program is definitely a program that teaches all the aspects of reading and writing thoroughly. The students are receiving a strong base in their native language and it is helping them as the years progress. I have been able to notice the difference in language acquisition since we began using the Extra Support Handbook for 30-45 minutes of English Language Arts instruction. This component is definitely a benefit to the students and their bilingual education (Teacher)." "I teach L1 and the impact has been effective, again, because we know have a complete program – books, assessment materials, etc., that we have never had before (Teacher)." "The positive impact of Reading First on our bilingual classes is having all the materials for students in both languages (Principal)" "This school has Spanish dual-immersion classes. The impact of Reading First is the same for *HMR* and *Lectura* classrooms. Having CORE materials that are identical in both languages has been very helpful (Coach)"
Bilingual vs English-Only Provides an opinion of whether it is better to have a bilingual program or teach in English.	"The students are more well behaved because they are learning in their native language and are ahead in many ways in the *Foro Abierto* program. The students in the English-only classes will have a head start however when it becomes all English (Teacher)" "I feel that students should be taught in English only because the state tests required are given in English. The state test in Spanish is given but does not count as does the English counterpart (Teacher)" "I do see that the instruction in primary language is very necessary. Some students don't have English support at home. They need to learn in any language. I completely disagree with the idea of testing them in English. I think it's criminal to test them in a language in which they are not comfortable (Teacher)" "The bilingual RF program forced our school to look at how we run our Bilingual classes, how we direct instruct in English, when we instruct in English, and when we provide Spanish support (Principal)" "I have mixed feelings about this topic. I have in the past been a strong supporter of the bilingual classroom. I have listened to major complaints that the *Foro Abierto* program uses language from other countries that have little to do with either Mexican or High Spanish. I am slowly coming to the conclusion that unless a program is dual immersion it may not be the support our children need (Coach)"
Equity of Program Expresses an opinion that there is now equity in the Reading First program for bilingual or waivered classes. May state that waivered classrooms were excluded from training and support but now are included	"This is the first time the waivered classrooms have had the same program as the English classes. We now can hold all programs to the same expectations and level of instruction by the teachers (Coach)" "Equal access in all areas is provided in these classroom settings (Teacher)" "The Reading First program provides Spanish-speaking students with the same opportunities as the English program to excel in school since they learn the same skills and strategies used in English instruction (Teacher)." "We have treated both programs equitably through Reading First with its leadership and assistance. All the various aspects of the Reading First Program remain the same for our few waivered classes (Principal)."

Same as Regular Classes States that there is no difference in the impact of Reading First on waivered classes compared with regular classes	"The *Lectur*a program aligns with the English program in teaching the comprehension strategies and skills vocabulary development spelling grammar writing etc. (Teacher)" "The RF Spanish program really is the same as the English program. There are equal services. (Teacher)" "Our Bilingual classrooms have received the same support as their counterparts in English under Reading First. I think the impact has been positive because it has helped us focus on teaching the components of reading (Principal)." "The student achievement expectation is the same in these classrooms. The expectations are the same; rigorous teaching is expected (Coach)." "The impact is equitable for all classes, both waivered and non-waivered (Coach)."
Suggestions for Improvement Provides suggestions for improving the reading instruction or Reading First activities relative to waivered classrooms.	***Examples of General Suggestions for Improvement*** "We need to provide teachers with an explicit plan for getting all students proficient by the end of grade three and the training they need to do so (Coach)" "We need more guided practice and more visuals (Teacher)" ***Examples of Program-Specific Suggestions*** "HM has too many comprehension skills that are not tied closely enough with the heavyweight language arts standards. Some of the comprehension skills need to be eliminated and replaced with multiple exposures to a single focus such as predicting and making inferences which are heavyweight standards (Teacher)" "I wish there were more guidelines on how to introduce the English sound-spelling cards and reading into the Spanish program (Teacher)"
Academic Negative Expresses a negative opinion about the academic outcomes in waivered classrooms	"Since my teachers know the non-Spanish version of Open Court, they help transition their students into reading in English. However, low comprehension and vocabulary scores are a problem for the Spanish speakers who transition into English-only classrooms (Coach)" "Our students in the bilingual program have a very difficult time transitioning into English and are usually very low in third grade. It is impossible for them to catch up (Teacher)" "When students who have been in Spanish classrooms in kinder to second grade come into our third grade classes, we have a hard time teaching them the third grade standards. We have to go back to first and second grade standards in order to teach them the third grade standards (Teacher)"

Describe their School's Program Provides an anecdotal description of how their school operates waivered classrooms	"Because there are limited resources to provide interventions in Spanish, we are having the bilingual classes only for those students who are on grade level in Spanish. If students are below grade level, we place them in English instruction with structured English immersion (Coach)" "We have dual immersion programs that consist of 50/50 (Spanish/English) and 90/10. In addition, we have a developmental biliteracy program (Coach)" "It has made us reflect on the delivery of instruction in our two-way immersion program. Our students now interact with the other students during deployment time, however, not at the expense of the core instruction (Principal)"
Transition Negative Describes problems with transitioning students from Spanish instruction to English instruction	"There is no strong plan for transition from Spanish to English. That component should be built in and begun as early as first grade. The English Language Development curriculum is not enough, especially if we want the EL students college bound (Coach)" "The Reading First program needs to provide clearer guidance on the transitioning process. We need specific assessment tools to measure their growth as they transition (Principal)" "I do not feel that Reading First and the Houghton Mifflin help transition students from one language to another. The text in Spanish should also be introducing English elements as the years progress to help our students transition. Instead, students are forced from Spanish to English with little ease of transition (Teacher)"
Assessments Negative Expresses a negative opinion of the assessment requirements, procedures or tools for students in waivered classes	"Some of the assessments in Spanish have not been accurately translated and that makes it difficult for students to be and feel successful in this program (Teacher)" "Our students receiving Spanish instruction outperform their English speaking peers when taking the state's Spanish language assessments. However, when they take the English assessments, they are behind. It is apparent that EL students need a fairer system when taking state required tests (Principal)" "There are significant burdens for teachers who must assess their students in both English and Spanish at the expense of instructional time (Teacher)"
Time is Problem Indicates that there is not enough instructional time to teach reading and language arts in waivered classes	"The two and one half hours of Spanish instruction leaves insufficient time to cover information in English (Coach)" "Spanish scores have gone down since our district began parallel instruction in English with the support of the RTAC. Students in bilingual classes are no longer getting enough instructional time in Spanish (Principal)" "There are not enough instructional minutes in a day to provide an adequate Spanish program in addition to an English program (Teacher)"

Professional Development Positive Expresses a positive opinion about the professional development provided for teachers of waivered classrooms.	"Participating in Reading First has helped our teachers teaching in waivered classrooms to receive the same training as the English component. The training and coaching have helped increase student achievement in the waivered classrooms (Coach)" "Because they (waivered classroom teachers) went to the same training as their English-instruction counterparts, they had increased knowledge of both programs such that there was increased dialogue amongst teachers at the same grade level when they were discussing the best ways to help their EL students (Coach)" "The Reading First program has impacted our alternative program in Spanish positively. Teachers of waivered classrooms participate in all the articulations and professional development and follow pacing guides the same way teachers in English settings do. Consequently, students in waivered classrooms have the same learning opportunities and expectations (Principal)"

Conclusions

In conclusion, this chapter finds that English learners (ELs) in non-waivered classrooms show significantly higher grade 2 and grade 3 STAR scores than English learners who have been in waivered classrooms for 2 or 3 years.

Many teachers, coaches and principals with experience in waivered classrooms served by the Reading First program expressed positive perceptions of the program. Participants generally had positive opinions of the state adopted curriculum materials used in waivered classrooms but expressed concerns regarding grammatical or typographical errors or problems with translation from English to Spanish in the materials.

Teachers, coaches and principals perceived that Reading First has resulted in improved outcomes for ELs served in waivered classrooms, primarily as a result of setting high expectations and accountability for ensuring students are proficient by the end of third grade.

Though Reading First support is attributed to facilitating the transition from Spanish to English instruction for EL students in waivered classrooms, participants expressed a need for further guidance on how to effectively conduct the transition.