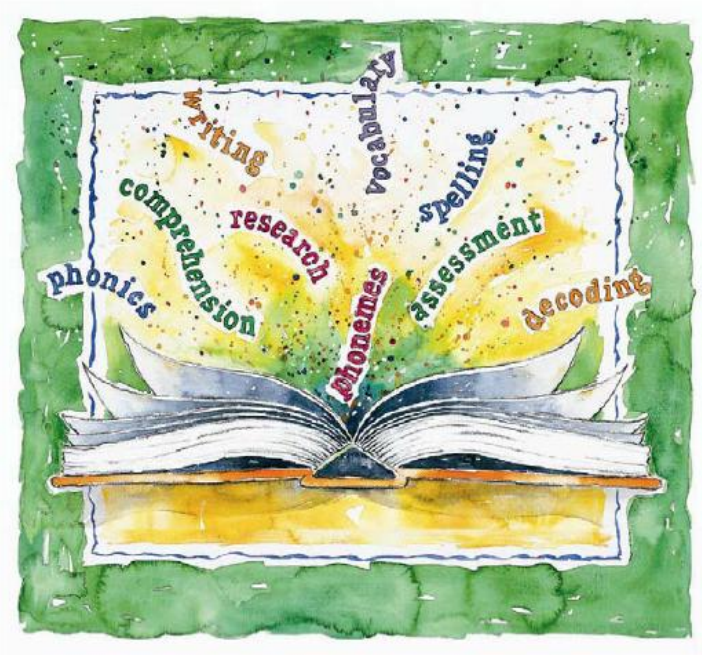
The California Reading First Year 6 Evaluation Report

December 2008



Prepared by:

Diane Haager, Ph.D.
Co-Principal Investigator

Craig Heimbichner
Co-Principal Investigator

and

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

EDUCATIONAL DATA SYSTEMS

Executive Summary

For six 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 six years. The current federal Reading First program will cease providing funding to states at the end of fiscal year 2008 and there will be one more report following this one.

The key findings, which remain similar to prior reports and suggest sustainability of effect, are summarized below. The body of the report contains the detailed analyses to support these conclusions. It should be kept in mind, as noted in Chapter 1, that this report examines Reading First in terms of its unique programmatic model for the State of California according to the California Reading First Plan. The findings for "Reading First" referenced below refer to this statewide study, a particularly significant fact in light of a national Reading First impact study which published findings that are not consistent with those presented here.

Finding #1: Reading First has had a significant impact on student achievement in California. Reading First schools have grown significantly since the inception of the program. They have grown significantly more than a statistical control group and significantly more than non-Reading First schools. High implementing schools have grown significantly more than low implementing schools. The overall impact of Reading First as determined through a meta-analysis of all achievement results since 2003 controlling for school demographic differences, is 0.082 with a standard error of 0.004. (This is an average standardized beta coefficient calculated using multiple regression analyses, not to be confused with Cohen's *d* or other measures of effect size.) This 0.082 effect is more than 16 standard errors greater than zero, where 2 standard errors greater than zero would be sufficient to claim a statistically significant effect with 95% confidence. (Chapter 2)

Finding #2: The Reading First effect is meaningful. Translated into real terms, the 0.082 Reading First effect means that a school that increases its Reading First Implementation Index statistic (RFII) by 25 points for a given time-span can expect, approximately, to double its achievement gains on a variety of grade 2-5 metrics over the same time-span – a 100% increase in achievement gains. A school that increases its RFII by just 5 points can expect a 10-20% increase in achievement gains. Viewed another way, the Reading First standardized effect size of 0.082 is 63% as large as that of demographic variables widely thought to be important predictors of achievement. Using the example of migrant students, the decrease in grade 2 scale scores that would occur if the percentage of migrant students in a school increased by 10 percentage points would be almost exactly counterbalanced by increasing the school's

RFII by 11 points. Thus the negative effect of a percentage point increase in a demographic variable like Percent of Migrant Students is almost entirely canceled by a one point increase in the RFII. (Chapter 2)

Finding #3: Growth remains significant. The Reading First Achievement Index (RFAI), a composite of K-3 achievement metrics for Reading First schools that ranges from 0 to 100, has risen an average of 3.1 points per year, equivalent to 18.6 points over 6 years relative to a starting year. On the grade 2 CST achievement metric Reading First schools have grown 30 scale score points since 2002, indicating significant growth for grade 2 in California since the program began. (Chapter 2)

Finding #4: The Reading First effect generalizes across student performance levels. Students in all performance levels show a boost from Reading First implementation. The advantage over non-Reading First schools is especially pronounced for students in the "Below or Far Below Basic" categories. On the grades 2-5 California Standards Test (CST) achievement metrics, the migration of students into "Proficient and Above" is matched or exceeded by a migration of students out of "Below or Far Below Basic". The migration of students out of "Below and Far Below Basic" is more than twice what it is for non-Reading First schools. (Chapter 2)

Finding #5: Reading First significantly impacts grade 4 and grade 5 performance. The above findings are replicated in grades 4 and 5, even though Reading First is a K-3 program. There is an approximate 10 point scale score difference between high and low implementing Reading First schools for grade 5. The grade 5 movement of students out of "Below or Far Below Basic" in Reading First schools is more than twice that seen in non-Reading First schools, three times more in high implementing schools. This finding indicates a sustainable and replicable effect of the program once students no longer have grade-level access to it due to funding and programmatic limitations (K-3). (Chapter 2)

Finding #6: 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. (Chapter 6)

Finding #7: Achievement among high implementers in the first cohort of schools is flattening. In line with predictions from previous reports, the growth rate of schools that have been in the program 6 years (Years in Program, or "YIP", = 6) is not as great as in previous years, causing a flattening of achievement scores in high implementing schools (mirrored in the trend-lines of non-Reading First schools) and a narrowing of the difference between high and low implementation schools. This indicates a possible "plateau effect," a phenomenon observed in mature programs. However, while the YIP 6 schools

may be encountering such an effect, the Reading First population as a whole is not. For the YIP 4 and YIP 5 schools the effectiveness of Reading First has been increasing since 2006. (Chapter 2)

Finding #8: Implementation of Reading First principles remains adequate but could be higher. Most schools in the Reading First program are implementing the program adequately, with modest growth in implementation since 2004. The average Reading First Implementation Index statistic (RFII) across all schools was 36 in 2004 and 2005, 39 in 2006, 2007 and 2008. The level of implementation appears to have hit a plateau. There is ample room for further growth in implementation, which could significantly increase achievement gains. (Chapter 3)

Finding #9: Principal participation and teacher program evaluations are strong predictors of achievement. Of the dimensions measured in the Reading First surveys, school-level implementation by the principal and school staff, and teacher evaluation of Reading First, are the two strongest predictors of achievement gains. This suggests that active principal participation and positive teacher perceptions of the program are likely to increase program effectiveness to a significant degree. (Chapters 3, 4)

Finding #10: The Reading First program has led to the development of a sustainable, well-integrated structure and process of providing reading/language arts instruction in California. The program elements outlined in the Reading First assurances are integral parts to a whole that is more than the sum of its parts. Taken individually, each program element would not likely have the effect of impacting reading instruction nor would they individually be sustainable. Together, the use of state-adopted curricula, professional development, coaching, leadership support, protected time blocks, and other program elements have impacted reading achievement in the state. (Chapter 4)

Finding #11: The 6-8 Week Skills Assessments are helpful. The 6-8 Weeks Skills Assessments were consistently used to monitor student progress and guide instruction during the past four years. Teachers, special education teachers, coaches and principals generally found them to be useful for monitoring student progress, guiding instruction, helping to identify students who need additional assistance, and helping to plan reading intervention for small-group follow-up instruction. Suggestions for improving the 6-8 Weeks Skills Assessments included improving the alignment of the assessments with the specific skills taught and the state standards; improving the test format, procedures, or timing; and specific suggestions regarding how fluency, comprehension and vocabulary are assessed. (Chapter 5)

Finding #12: Most special education teachers use their district's adopted reading/language arts curriculum. Over 50% of the special education teachers reported using their district's adopted reading/language arts curriculum for the majority of their instruction, while others use alternative curriculum materials or partially use the core materials. (Chapter 7)

Finding #13: Special Education remains disconnected from the general education environment. Despite legal mandates for inclusion and improving access to the general education environment for students with disabilities, this analysis suggests that communication barriers continue to exist regarding students with disabilities and their participation in the grade-level reading/language arts curriculum. Participants most frequently reported that they were not aware of how the Reading First program has impacted special education teachers and students while others reported a generally positive impact but did not elaborate. (Chapter 7)

Finding #14: Schools have not yet begun to implement Response-to-Intervention (RTI). Many schools either have not yet begun to implement Response-to-Intervention (RTI) or are in the beginning stages of implementation. Yet, some participants reported implementation of intervention for struggling readers and were able to describe specific elements of their RTI approach. (Chapter 7)

Although there is ample room for improvement in program implementation and in the program itself, there are no significant negative findings to report regarding California's Reading First program.

Background

Reading First is a federal initiative aimed at improving reading instruction in America. Authorized in 2001 as part of the No Child Left Behind (NCLB) Act, Reading First promotes the use of scientifically based reading practices in grades K-3. The initiative provides a significant amount of federal funding for improving reading instruction for large proportions of students experiencing academic difficulty and socio-economic disadvantage. This funding ceases as of the end of fiscal year 2008.

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

- Use of a state-adopted reading program
- Access to training programs authorized by state legislation and focused on research-based reading instruction, including Senate Bill (SB) 472 teacher and coach professional development and Assembly Bill (AB) 430 principal professional development, as well as extended follow-up professional development
- Access to assessment tools that measure students' skills every six to eight weeks
- Hiring of reading coaches, expert teachers who support program implementation

[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."

Anecdotal evidence indicates that many non-Reading First schools have voluntarily been adopting some or all of these components over the same 6-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 six 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. In addition, this chapter comments on the April 2008 National Study for Reading First to provide context for reading this California statewide study. Finally, this chapter briefly examines Reading First in the context of federal accountability measures under NCLB (Program Improvement), since such sanctions are increasingly widespread and important both statewide and nationwide under NCLB's accountability timeline.

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. In addition, this chapter contains "lessons learned" about the program elements in Reading First.

Chapter 5 focuses on the use of assessment in Reading First schools by examining participants' reported use of the assessments, perceived benefits of the data provided, and suggestions for improvement.

Chapter 6 provides achievement statistics and trend-lines showing the growth of the English learner subpopulation in Reading First schools since 2002 and an examination of participants' perceptions of the impact of the Reading First program on the English language development of English learners.

Chapter 7 looks at the impact of Reading First on the special education programs in Reading First schools through an analysis of special education teacher surveys and an open-ended question on the teacher, coach and principal surveys regarding the impact of Reading First on special education. This analysis is a new dimension assessed for the first time this year in the statewide evaluation study.

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 with their RFAI and RFII scores for 2005-2008 (Appendix F)

A Data Example from Grade 2

Our core findings are exemplified in the tables and figures presented below, representing the growth of reading achievement in Reading First schools on grade 2 achievement metrics since these schools entered into the program six 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 also report here the results for grade 5.

Table ES.1.0 reports the grade 2 CST starting and ending scores for schools that have been in the program 6 years (Cohort 1). Statistics are reported for the program as a whole, for the program broken out by high and low implementing schools, for a statistical control group, and for all of the non-Reading First elementary schools in the state.

Table ES.1.0: CST Metric, Years in Program = 6, Grade = 2

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

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."
[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.
Note: Numbers reporting change since starting year were rounded and may not appear to be an exact difference between 2002 and 2008 figures.

Referring to the "All Reading First Schools" column, we note the following. On the percent Proficient and Above achievement metric, an average of 15.5 percent of students in these schools scored Proficient and Above in 2002. By 2008, this percentage had increased so that 35.8 percent of students were scoring Proficient and Above. The size of the gain was 20.3 percentage points. (Note that rounding accounts for any seeming discrepancies in computing the change from 2002 to 2008.) The superscripts "[abc]" tell us this gain was "significantly" greater than the gains of the "statistical control group"[2] and the non-Reading First schools in California, and that the gain is significantly greater 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 or Far Below Basic in 2002 and in 2008, and the subsequent change. This change is negative because it refers to students moving *out*

[2] The "statistical control group" is a construct defined using multiple regression to hold the effects of school population characteristics constant 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.

of the bottom two performance level categories. Reading First schools particularly distinguish themselves from the rest of the state in this column, moving twice as many students out of the lower performance levels.

Then we see the average CST grade 2 scale score (a test score ranging roughly from 250 to 450) for students in 2002 and in 2008, and the difference between them. Remember that there was one group of students who were in grade 2 in 2002, and there was another group of students who were in grade 2 in 2008. On average the 2008 students scored an average of 30 scale score points higher than their 2002 predecessors. For context, that is more than 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" and "low implementing" (using an implementation index calculated from teacher, coach, and principal responses to the annual Reading First implementation survey). The "Statistical Control Group" column estimates the likely growth of schools that are similar to Reading First schools but not implementing the program.[2]. The last column reports the same statistics for the remaining 4,057 elementary schools in California that are not in the Reading First program. Since this population has much higher starting scores than the Reading First schools, the starting points for "All Non-Reading First Schools" have been adjusted downward in the trend-line charts to coincide with the starting points of the Reading First groups.

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.

These findings support the efficacy of Reading First and extend and confirm the findings from the Year 5, Year 4 and Year 3 California Reading First Evaluation Reports. The trend-lines corresponding to Table ES.1.0 are presented below, in Figures ES.1.0, ES.1.1, and ES.1.2. We call attention to three points:

- There is evidence of a "plateau effect" for high implementing schools that have been in the program 6 years. The trend-lines show that the percent Proficient and Above for the high implementing schools was flat in 2007 and 2008 and that the gap in achievement between high and low implementing schools has been decreasing since 2006. This effect has been anticipated in previous evaluation reports and reflects a possible natural limit in the amount of additional increase that can be expected of any educational program after the program has become fully absorbed. This plateau effect is not observed for later cohorts of schools or for the Reading First population as a whole.

- Despite the flattening in percent Proficient and Above, the average scale score achievement gains remain steady for all Reading First schools, presumably due to the strong and ongoing movement of students out of the lower performance levels.
- The achievement of non-Reading First schools has reached a plateau for students at all performance levels, causing their growth to lag even further behind that of Reading First schools.

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

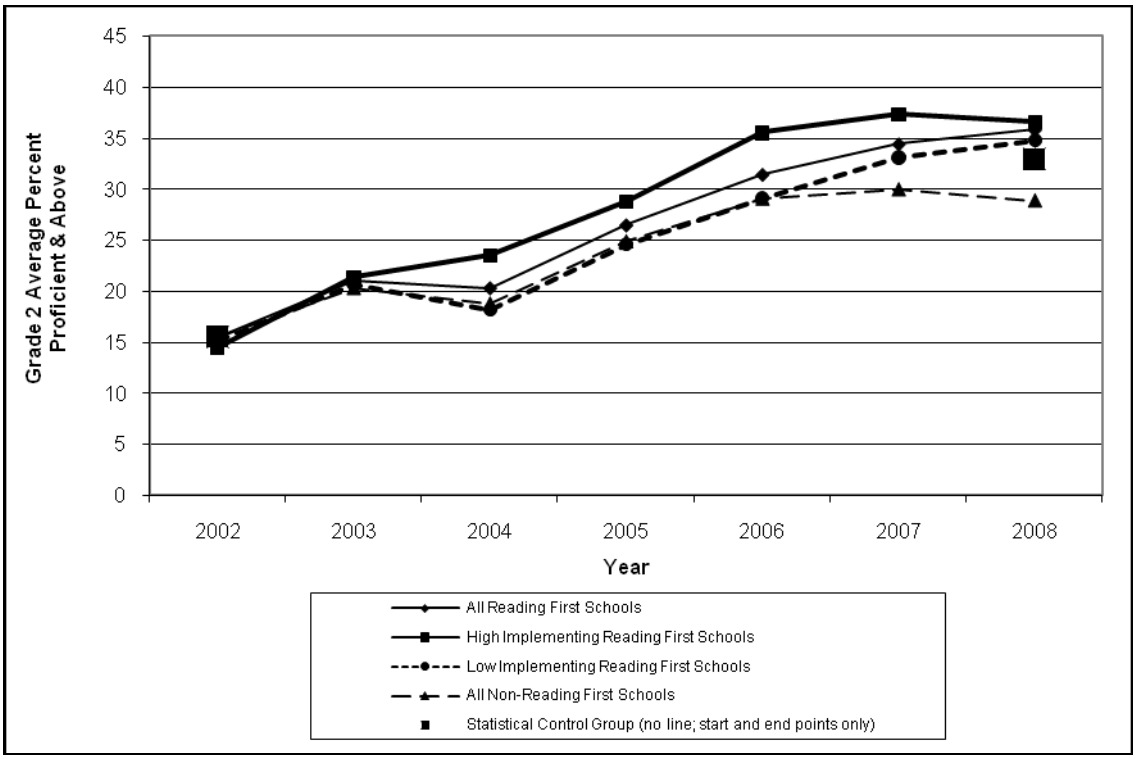


Figure ES.1.1: CST % Below Basic & Far Below Basic, YIP = 6, Grade = 2

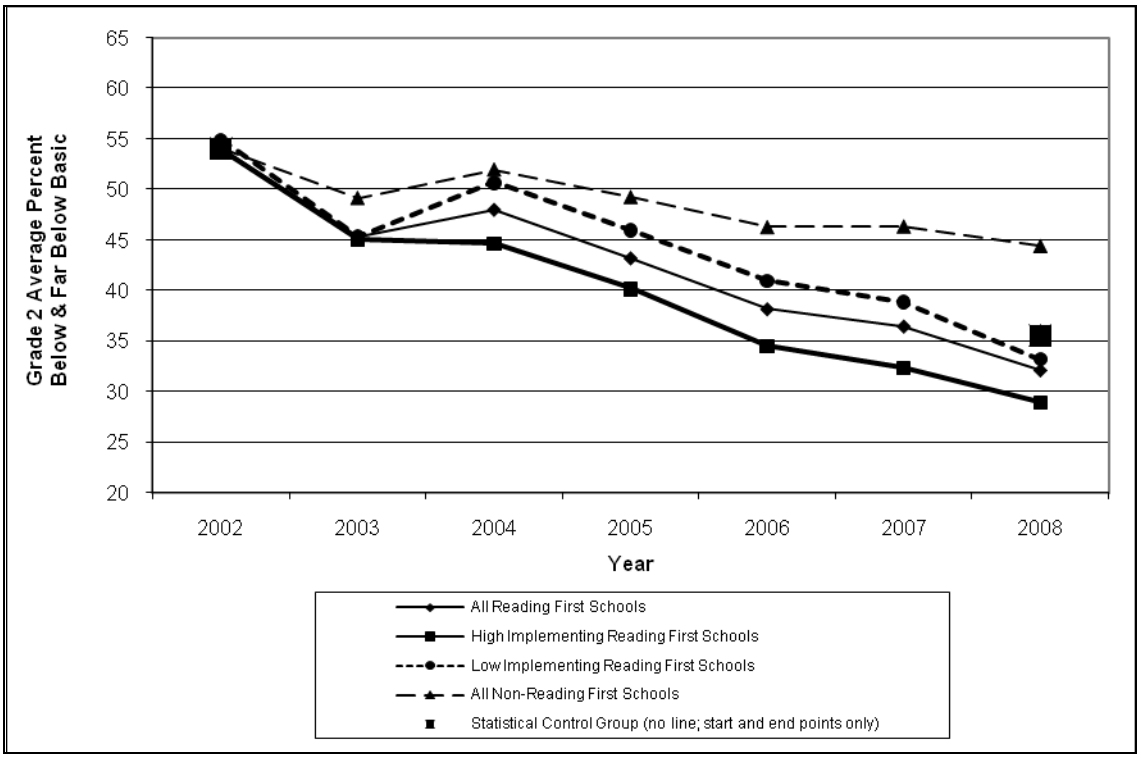
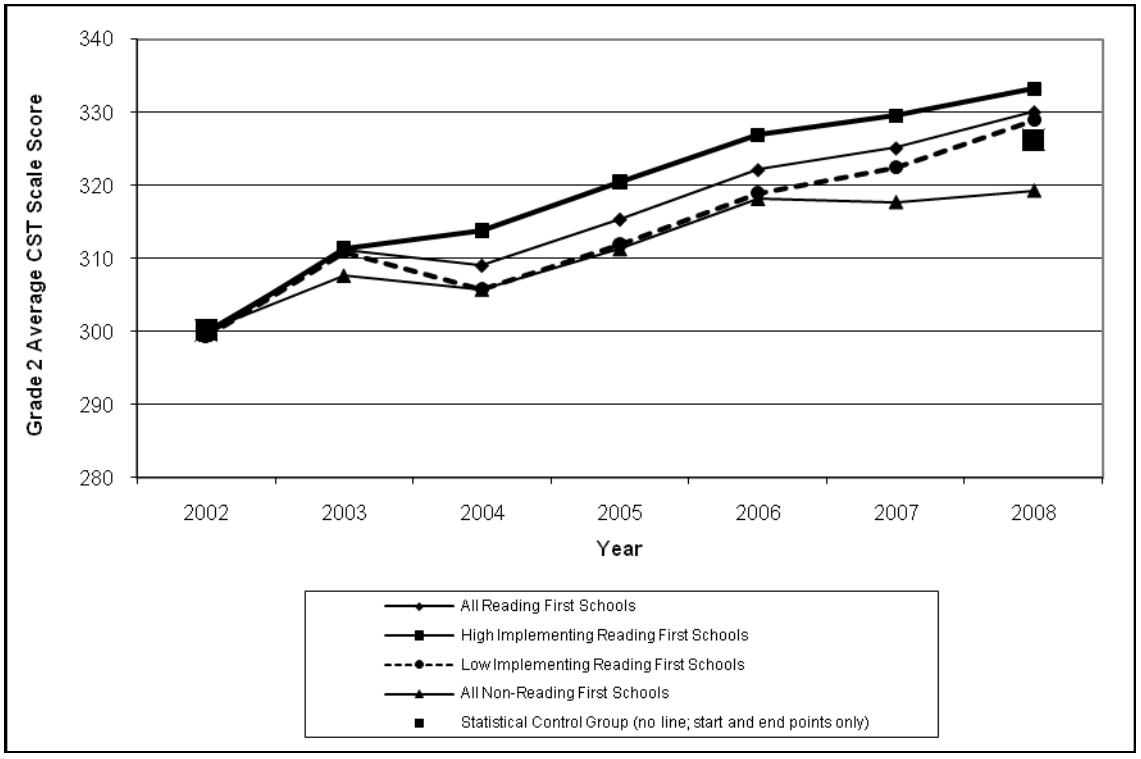


Figure ES.1.2: CST Mean Scale Score, YIP = 6, Grade = 2



A Data Example from Grade 5

This Year 6 Reading First Evaluation Report continues the Year 5 Report's expansion of the scope of the evaluation by adding grade 5 CST performance as an achievement outcome. Because Reading First is administered only in grades K-3, the grade 5 results shed light on whether student exposure to Reading First in the earlier grades improves ability to read in grades 4, 5 and above. Table ES.2.0 and Figures ES.2.0, ES.2.1, and ES.2.2 show that it does, most dramatically for students who were in grade 5 in 2008. 2008 is the first year for which students who have been in the program since kindergarten reached grade 5.

Table ES.2.0: CSTs, YIP = 6, Grade = 5

Years in Program: 6 **Grade: 5**	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	238	26	87	N/A	3,988
% Proficient and Above					
2002	11.1	11.5	11.8	11.1	35.8
2008	29.9	35.5	28.6	26.7	52.3
Change Since Starting Year	**18.8**[abc]	**24.0**[abc]	**16.8**[c]	**15.7**	**16.5**
% Below or Far Below Basic					
2002	50.0	49.1	48.7	50.0	24.3
2008	31.3	25.0	33.8	34.3	16.6
Change Since Starting Year	**-18.7**[abc]	**-24.1**[abc]	**-15.0**[bc]	**-15.7**	**-7.7**
Mean Scale Score Per Student					
2002	303.4	303.7	304.8	303.4	334.5
2008	324.9	331.7	322.9	321.0	352.3
Change Since Starting Year	**21.5**[abc]	**28.0**[abc]	**18.1**[c]	**17.6**	**17.9**

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."
[b] Significantly different ($p < 0.05$) relative to "All Non-Reading First Elementary Schools."
[c] Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.
Note: Numbers reporting change since starting year were rounded and may not appear to be an exact difference between 2002 and 2008 figures.

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.

Figure ES.2.0: CST % Proficient & Above, YIP = 6, Grade = 5

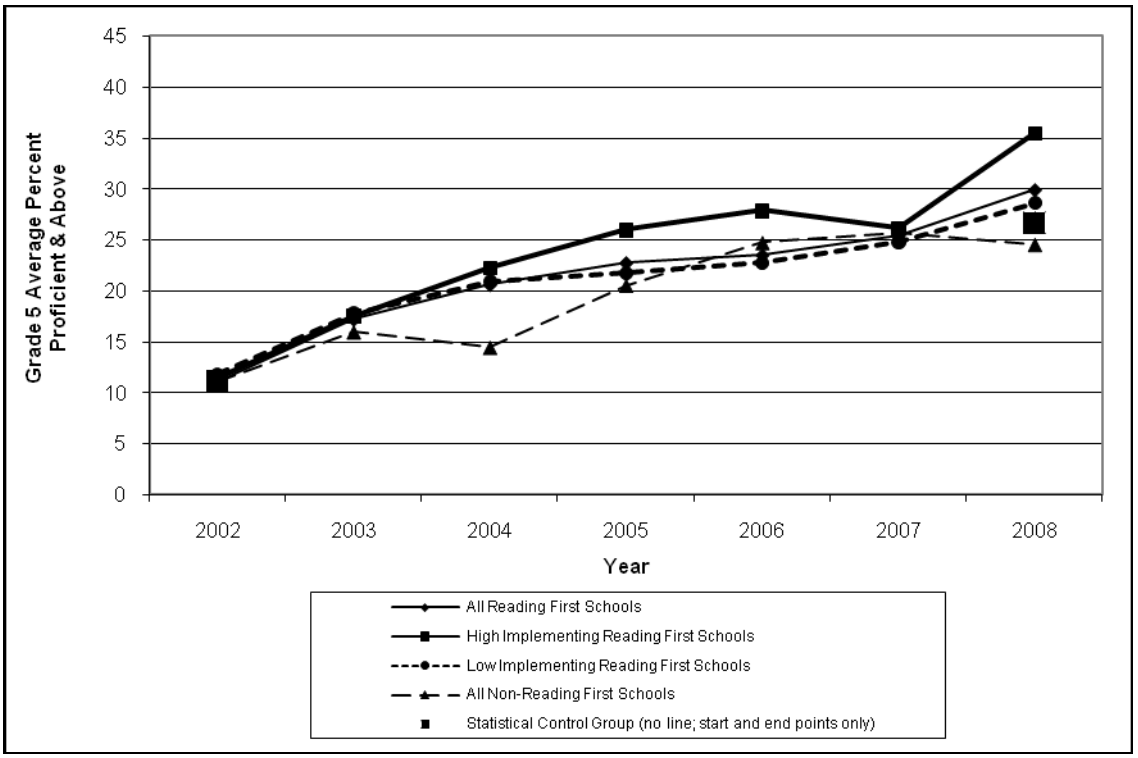


Figure ES.2.1: CST % Below Basic & Far Below Basic, YIP = 6, Grade = 5

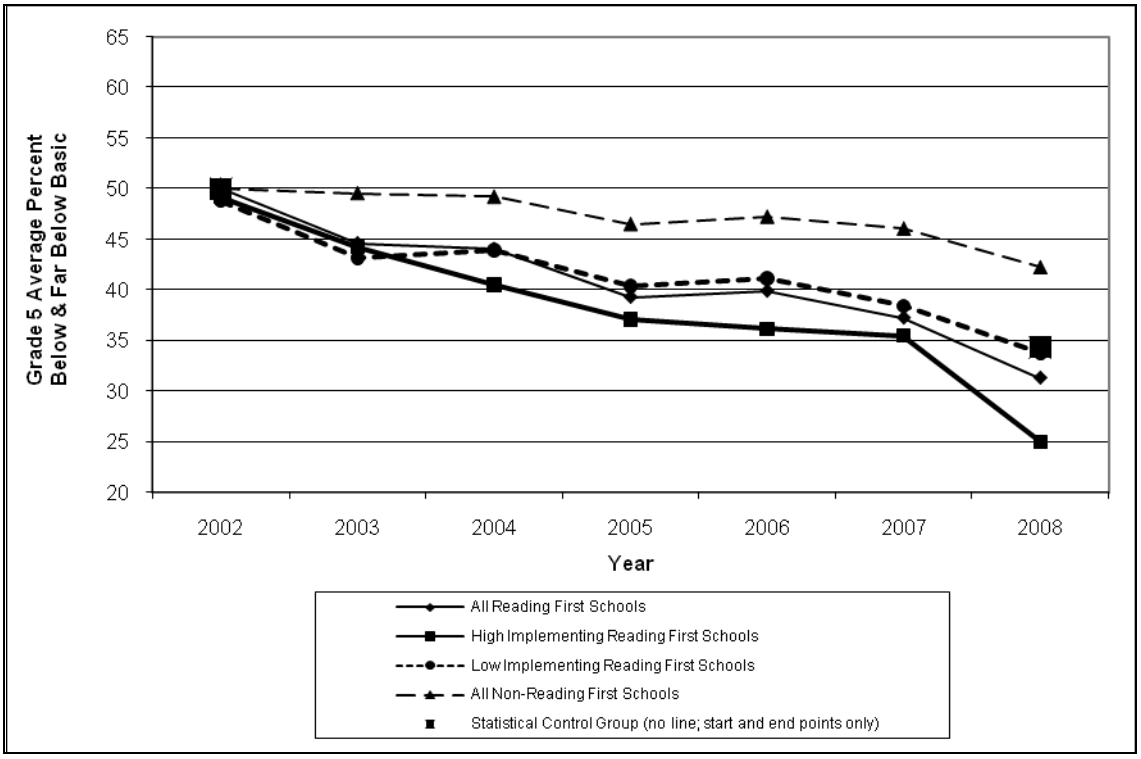
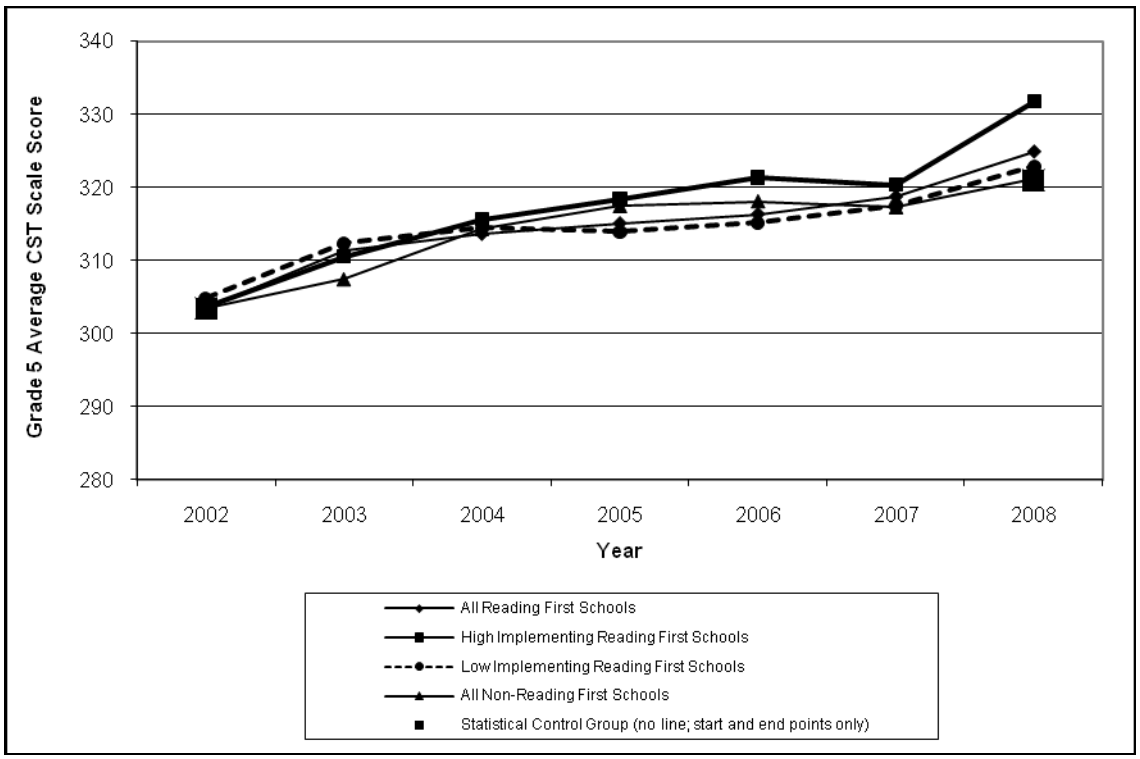


Figure ES.2.2: CST Mean Scale Score, YIP = 6, Grade = 5



Policy Recommendations

This and previous evaluation reports have provided consistent evidence that the Reading First program has had a significant impact on reading instruction and achievement in California. As part of NCLB, Reading First has been a politically charged and controversial program. However, our research shows that the Reading First program has improved reading achievement for English learners, students scoring Below or Far Below Basic on the CSTs, and students in the types of low-performing elementary schools that are eligible for Reading First. Teachers, coaches and principals have reported the importance of various aspects of the Reading First program. In addition, we have presented evidence that many non-Reading First schools have, at least in part, implemented program elements required in Reading First schools and they also have experienced reading gains.

The policy recommendations listed below are based on evidence gathered in six years of evaluation of the Reading First program. Since the Reading First federal funding will soon end and there is no clear indication that federal legislators will fund a similar reading initiative in the near future, these recommendations provide suggestions on how to sustain the benefits of the Reading First program in California in the coming years.

Fidelity of Implementation

The importance of fidelity of implementation of scientifically based reading instruction via a state-adopted curriculum is a strong and consistent finding of the six years of evaluation of Reading First in California. The Reading First assurances provided guidelines to ensure implementation and the Reading First funding has supported various elements that constitute a strong infrastructure to support implementation. Now that California has built an infrastructure of support, it is important to maintain that infrastructure and to apply lessons learned regarding how to ensure fidelity. The 2008 Reading/Language Arts Framework and textbook adoption outlines the key elements of a research-based reading program. Though this is an important step in maintaining excellence in Reading/Language Arts instruction, it is important for policy makers to consider how the Reading First assurances and support structures have provided the necessary guidance and support to achieve fidelity of implementation. Simply stated, a district's adoption of a new state-adopted Reading/Language Arts curriculum does not ensure fidelity of implementation. We strongly encourage California's policy makers to consider how to strengthen our educational infrastructure to continue to support fidelity of implementation. One important consideration in this regard is to develop a mechanism for measuring and monitoring implementation.

Maintain Support Structures

There are several mechanisms that may provide the support needed to maintain the positive effects of Reading First. Findings in the California Reading First evaluation reports have repeatedly highlighted the value and benefits of deep and ongoing professional development, a highly qualified coaching force, knowledgeable and involved site administrators, time for collaborative lesson planning, and the use of data to guide instruction. These elements that have been supported by Reading First funding must be sustained in order to maintain fidelity of implementation. It is our hope that budget cuts will not sidetrack the progress made in reading instruction through the Reading First program. Investing in maintaining the support structure that has been built with Reading First funding would be a wise investment in California's future.

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

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

Retain "Program Coherence" as a Guiding Principle

Much of the success of Reading First resides in its program coherence. It is adopted all as a piece or not at all. In the post-Reading First environment, California should resist the temptation to offer a menu of recommendations and resources from which schools can pick and choose at will. Care should be taken that when programs are adopted, they are adopted as a whole and are internally consistent, well-focused, and less prone to lead to fragmentation of time, energy, resources, and educational strategy.

Maintain a Focus on Improving Reading Outcomes for English Learners

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

Maintain Strong Policy

The Reading First program required states to adopt policy guidelines related to the Reading First assurances. These policies have thus far proven to be effective in raising the quality of reading instruction in California. We urge the state to maintain a strong policy to support continuation of the principles and practices promoted by Reading First.

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 six years of implementation from academic year 2002-03 to 2007-08.

The California Reading First Plan delineates the roles and operational procedures for personnel involved at the state and local levels. The State Board of Education (SBE), Office of the Secretary of Education (OSE), and the California Department of Education (CDE) direct the Reading First program in California. The Reading and Literacy Partnership Team, with membership broadly representing the interests of reading education in the state, serves an advisory role for Reading First. A subcommittee of the Partnership, the Evaluation Advisory Group (EAG), including designees of the members, advises the external evaluator. The California Technical Assistance Center (C-TAC) has responsibility for the statewide technical assistance program and oversight of the Regional Technical Assistance Centers (R-TACs) in providing regional and local support to LEAs. The C-TAC also coordinates the statewide network of professional development programs for teachers and site administrators through the Reading Implementation Centers (RICs).

The California Reading First Plan is based on a series of Assurances that are implemented by the LEAs. With these assurances, California's Reading First program is designed to ensure full implementation with fidelity to a comprehensive research-based reading program. Here, we briefly describe the assurances and program elements designed to address them.[3]

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

Vision Statement

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

Curriculum

Participant LEAs are required to use one of California's two state-adopted reading curricula: SRA/McGraw-Hill's *Open Court Reading 2000* or *2002* (OCR) or the Houghton Mifflin *Reading: A Legacy of Literacy 2003* (HM). The Reading First program has provided extensive support for LEAs in the implementation of the adopted curricula. In the 2004-05 school year, California's Reading First program began offering support for LEAs with "waivered" classrooms, that is, classrooms offering a bilingual instruction model using Spanish-language versions of the adopted curricula. California law (Proposition 227) mandates instruction in English for all students unless parents sign a waiver specifically requesting bilingual instruction. The two state-adopted Spanish-language reading programs are: SRA/McGraw Hill's *Foro abierto para la lectura* and Houghton Mifflin's *Lectura: Herencia y futuro.* Students receiving bilingual reading instruction in Spanish and English must transition from bilingual instruction to English instruction, and take the English Standardized Testing and Reporting (STAR) English Language Arts Content Standards Test (CST) at the end of grade 2 and grade 3. Regardless of the LEA's selected curriculum, each LEA is required to implement fully the district's state-adopted reading/language arts program for an uninterrupted 60 minutes per day in kindergarten and 150 minutes per day in Grades 1-3, according to a district-approved pacing plan that outlines when each daily lesson is taught at each grade level in an academic year. This plan not only assures that students will complete the grade-level curriculum but also that implementation occurs systematically in every Reading First school. Also, LEAs are beginning to plan and implement extensive intervention with those K-3 students who need an additional 30 minutes of instruction. The intervention materials are approved by the SBE as scientifically research-based.

Professional Development

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

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

Curriculum-Embedded Assessment

For program monitoring, LEAs are required (since 2005-06) to use curriculum-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).

Assessment is a key program element in Reading First, but this report will not treat assessment in chapter four with other program elements, instead reserving a separate chapter (Chapter 5) for a comprehensive analysis.

Collaborative Teacher Meetings

All Reading First schools are required to hold regular grade-level meetings twice a month to provide an opportunity for teachers to work together to refine their implementation of the program. School principals and reading coaches are encouraged to assist in facilitating and supporting these meetings.

District Commitment

Each LEA is required to conduct an internal evaluation on the effectiveness of its implementation of the Reading First program. This evaluation includes a district action plan for the subsequent year and each school's action plan for its first tri-semester based on student achievement data and principal, coach, and teacher recommendations. In addition, district personnel must assure that the Reading First program is well coordinated with other programs such as Title I, Language Acquisition, and Special Education. Each LEA must have a district Reading First Leadership Team that meets regularly to advise and support the program.

Coaching

LEAs may use Reading First funds to provide reading coaches, content experts, and coach coordinators and ensure that these experts are adequately trained. Coaches offer site-specific support for implementation of the LEA's adopted reading curriculum and effective instructional strategies. The C-TAC has provided these experts two Coach Institutes annually for in-depth training and a Leadership Program for selected experts in partnership with a California university. Additional training for new

coaches is provided by the RICs, and support for both coach and coach coordinators is offered by the R-TACs.

Site Leadership

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

Program Coherence

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

State Leadership

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

Technical Assistance

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

LEA Cohorts

California has now completed six 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 (342 schools) has been receiving funding

and implementing the program for approximately five and one-half years. LEAs in Cohort 2 (357 schools) were selected for funding in 2003-04. Cohort 3 (132 schools) was added in 2004-05. A small number of LEAs were added in 2006-07 to make a new cohort, Cohort 4 (19 schools). A total of 850 schools in 110 LEAs are included in this Reading First Year 6 report.

California Reading First Year 6 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[4]) 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 5. This current report represents the Year 6 evaluation report, and will include outcomes from the 2007-08 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?

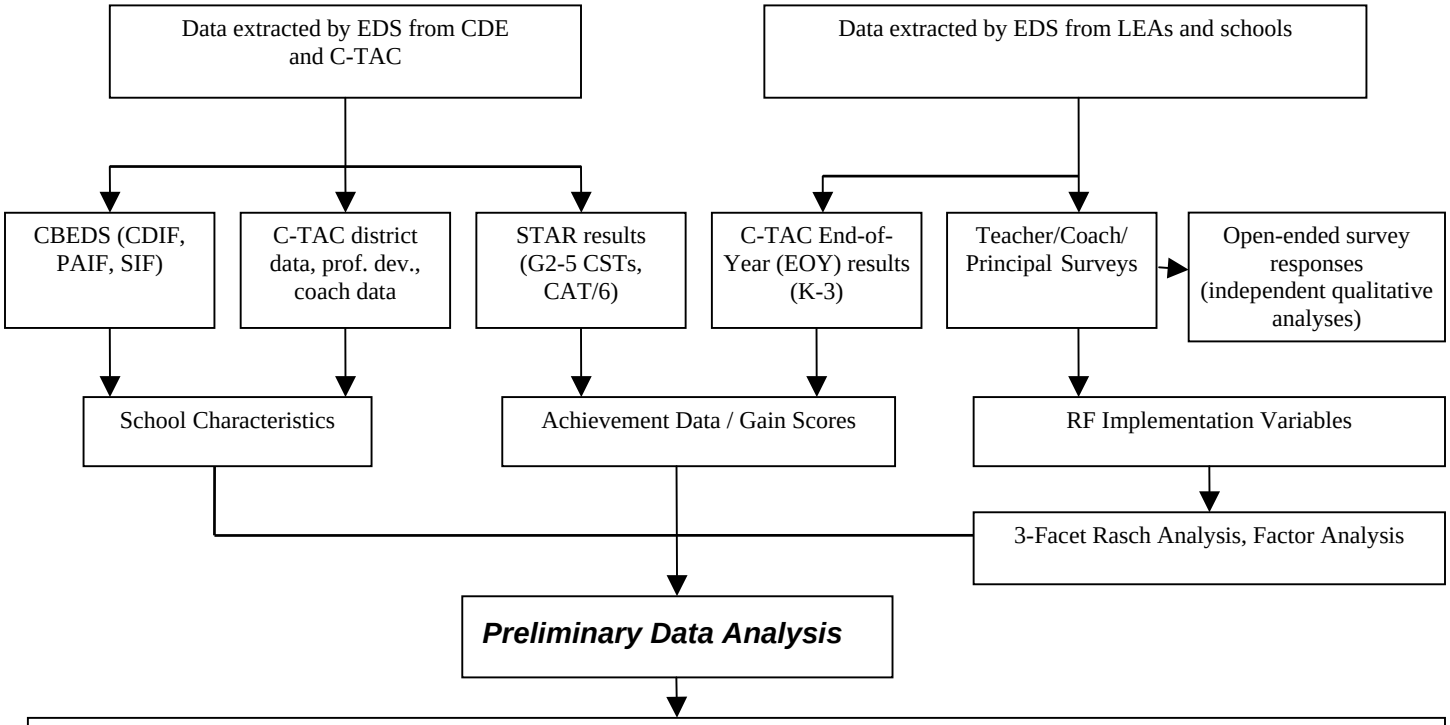
The conceptual framework below provides an overview of the evaluation study design. As described in the conceptual framework, the Reading First data can be organized into three types: a) school and district characteristics; b) achievement data; and c) implementation data. The school and district characteristics are described later in this chapter, with data drawn from state databases, including the California Basic Education Data System (CBEDS) file, the demographic sections of the California English Language

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

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 (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 the average school scale score. An additional analysis yields the Reading First Achievement Index (RFAI), which combines the STAR and EOY data. To examine implementation, a Many-Facet Rasch model is used to combine the teacher, coach, and principal surveys into a coherent measurement framework. The variables used and the analyses have been conducted in accordance with recommendations of the Reading First EAG. Finally, as noted in the executive summary, the Year 6 report examines the impact of Reading First on special education, particularly through an exploration of supporting professional development for special education teachers. This is a new dimension of the report added in Chapter 7.

Figure 1.1: Conceptual Framework – Year 6



Derived Variables of Study

Independent Variables

- Student subgroups: English language learners, special education, socio-economically disadvantaged, ethnic/racial subgroups
- School characteristics: Reading First or non-Reading First, 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 G2-5 CSTs and CAT/6 scores, End-of-Year results, Reading First Achievement Index (RFAI), school gain scores organized by years in program, STAR results

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)
- Qualitative Analysis (identification of themes and frequencies in open-ended data)

Reports

(By school, years in program, cohort, student subgroup)

- 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
- Themes and responses to open-ended questions

The Reading First National Study

In April, 2008, the U.S. Department of Education Institute of Education Sciences released a national Reading First study, titled *Reading First Impact Study: Interim Report* (RFIS).[5] While noting "positive, statistically significant impacts on the five essential components of reading instruction[6] promoted by the program," the study also stated, in an apparent contradiction, that "on average across the 18 study sites, Reading First did not have statistically significant impacts on reading comprehension test scores in grades 1-3." The contradiction is implied in that one of the "five essential components of reading instruction" referenced by the study as carrying a "positive, statistically significant" impact is reading comprehension—the exact component noted later in the national study as *not* impacted by Reading First with any statistical significance. Setting aside this contradiction within the study itself, however, we will note here that the California Reading First evaluation does not replicate the finding of no statistically significant effect; on the contrary, it finds that the Reading First effect has a high level of statistical significance. While a detailed comparison of the RFIS and the California evaluation is not attempted here, it is noted that the California evaluation has a substantially higher sample size than the national study, that it has been conducted over a longer period of time, and that, by defining its control group in terms of degree of school-level Reading First implementation it avoids contamination of the control group by Reading First treatment elements. We consider it possible, even likely, that the interim findings of the national study lead to an incorrect conclusion regarding the efficacy of the Reading First program, and that the California evaluation leads to a more correct conclusion, certainly with respect to California, and possibly with respect to the rest of the country as well.

That said, the California Reading First Plan contains unique elements which mitigate a unilateral transfer of either positive or negative conclusions to generalizing about statewide impact from the national study, and vice versa. California's pronounced emphasis on professional development based on implementation of State Board of Education-adopted reading programs, combined with advanced professional development and strong content coaching are elements which differ from many Reading First models in other states, although commonalities also exist due to federal requirements. While every educator and policy maker concerned with maximizing reading achievement in California should certainly study the above report, caution is advisable in assuming a strict application to California. The study should at a minimum be read in conjunction with this, and prior, statewide Reading First evaluation studies centered on California's students and Reading First programs.

[5] The text of the report is available online at http://ies.ed.gov/ncee/pubs/20084016/index.asp
[6] Phonemic awareness, phonics, vocabulary, fluency and comprehension.

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 starting point for their achievement as compared to Reading First schools. An additional difficulty with using comparison groups is the statewide effort to improve reading instruction in non-Reading First schools. It is likely that state-adopted curricula, state-funded professional development, and other elements of Reading First were present in non-Reading First schools, making it impossible to discern the true impact of the Reading First program. Indeed, a survey of LEAs eligible for but not participating in Reading First found that "Almost 60% of these LEAs use [Open Court and Houghton-Mifflin] programs exclusively in at least 67% of their schools[7]." Thus, the instructional materials and practices used in most eligible non-Reading First classrooms are likely to closely mirror those used in Reading First classrooms. This similarity in reading programs is matched by similar trends in student achievement, although Reading First schools have shown more substantial growth.

For a more complete discussion of the difficulties with constructing a valid comparison group of schools, the reader is referred to the Year 4 report. For this Year 6 report, no data are reported for non-Reading First comparison schools due to the inherent difficulties in establishing adequate comparisons. However, analyses are conducted using a statistically derived comparison group, as described in the Year 4 and Year 5 reports, and in Chapter 2 of this report.

Demographic Characteristics of Reading First Schools

California's Reading First program began in the 2002-03 academic year. During 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

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

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 863 in the 2007-08 academic year) from the cohort included in the current report:

(a) Cohort 1, first funded in 2002-03, with 13 LEAs (352 schools in current report); YIP 6
(b) Cohort 2, first funded in 2003-04, with 60 LEAs (358 schools in current report); YIP 5
(c) Cohort 3, first funded in 2004-05, with 27 LEAs (132 schools in current report); YIP 4
(d) Cohort 4, first funded in 2006-07, with 10 LEAs (21 schools in current report); YIP 2

The demographic data included in this chapter are extracted from the STAR research file published on the CDE Web site.[8] 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 for the starting year (varies by cohort) and for 2008. Table 1.1 also includes the 2003 and 2008 percentage of SED students in all elementary schools in the state. It is evident that Cohort 1 had the highest percentage of SED students compared to other cohorts, with 90.98% in 2008.

Cohort 2 had 85.2% and Cohort 3 had 84.42% SED students in 2008. Cohort 4 had the lowest percentage of SED students, 81.68%.

English Learners (ELs)

In 2008, 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 52.7%, 54.83% and 57.66% respectively. Cohort 4, with 33.21% ELs, more closely resembled the statewide figure of 30.09%.

Students with Disabilities

In 2008, the percentage of students with disabilities was reported as 7.47% for Cohort 1, 6.31% for Cohort 2, 6.27% for Cohort 3 and 7.42% for Cohort 4. This varies somewhat from the statewide

8The STAR research file used for the 2007-08 data was the version obtained by EDS on September 16, 2008, referred to as "P2."

percentage of 9.73%. It is interesting to note that all cohorts this year show a drop in percentage 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, 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 overrepresented in Cohort 1 compared to Cohorts 2, 3, and 4 and the All Elementary Schools category.

Table 1.1: Student Demographic Data, 2003-2008

	Reading First Schools								**All Elementary Schools**[1]	
	Cohort 1		**Cohort 2**		**Cohort 3**		**Cohort 4**			
	2003	**2008**	**2004**	**2008**	**2005**	**2008**	**2007**	**2008**	**2003**	**2008**
Number of Schools	*329*	*342*	*343*	*357*	*136*	*132*	*19*	*19*	*5823*	*6108*
SED (%)	91.26	90.98	82.69	85.20	85.15	84.42	73.37	81.68	51.00	54.12
EL (%)	58.50	52.70	52.97	54.83	57.50	57.66	31.21	33.21	27.15	30.09
Students with Disabilities (%)	7.52	7.47	8.02	6.31	7.05	6.27	7.89	7.42	9.77	9.73
African American (%)	17.20	13.30	8.85	7.29	6.56	6.52	14.05	14.21	7.81	7.48
American Indian (%)	.27	.30	.98	.72	.77	.95	7.68	8.37	1.33	1.30
Asian (%)	3.97	3.25	4.57	3.99	1.14	1.13	1.68	1.74	7.31	7.52
Filipino (%)	.96	1.01	1.66	1.55	1.26	.89	4.63	4.47	2.16	2.42
Hispanic (%)	71.53	76.72	72.02	75.13	77.12	77.20	50.53	51.42	40.22	44.34
Pacific Islander (%)	.50	.42	.81	.67	.54	.48	.63	.53	.62	.63
White (%)	3.71	2.63	9.57	7.27	11.20	9.87	19.32	16.58	36.51	31.18

[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 each Reading First cohort and for all cohorts combined. In this table, it is evident that most of the schools 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 suburb categories. Cohort 3 consisted mainly of suburbs of large cities and rural designations, resulting in high levels of migrant students. Cohort 4 was evenly split between urban and rural designations.

Table 1.2: Urban-Rural Distribution for Reading First Schools 2008

School Location	Cohort 1		Cohort 2		Cohort 3		Cohort 4		All cohorts	
	No. of Schools	% of schools[1]	No. of Schools	% of schools	No. of Schools	% of schools	No. of Schools	% of schools	No. of Schools	% of schools
City: Large	218	63.7	126	35.9	12	9.4	1	5.0	357	42.5
City: Midsize	6	1.8	55	15.7	13	10.2	6	30.0	80	9.5
City: Small	16	4.7	18	5.1	13	10.2	2	10.0	49	5.8
Suburb: Large	95	27.8	98	27.9	59	46.1	1	5.0	253	30.0
Suburb: Midsize	0	0	16	4.6	0	.0	1	5.0	17	2.0
Suburb: Small	0	0	2	.6	3	2.3	0	0	5	.6
Town: Fringe	2	.6	16	4.6	0	0	0	0	18	2.1
Town: Distant	1	.3	2	.6	13	10.2	3	15.0	19	2.3
Town: Remote	0	0	3	.9	0	0	0	.0	3	.4
Rural: Fringe	4	1.2	11	3.1	6	4.7	3	15.0	24	2.9
Rural: Distant	0	0	2	.6	6	4.7	2	10.0	10	1.2
Rural: Remote	0	0	2	.6	3	2.3	1	5.0	6	.7
Total	342	100.0	351	100.0	128	100.0	20	100.0	841	100.0

[1] The percent of the schools in that cohort in a particular type of location.
Data source: National Center for Education Statistics (NCES

Percent of Migrant Students

Table 1.3 presents the school-level average percent of migrant students broken out by "Years in Program" (YIP) instead of funding cohort. Table 1.3 shows that YIP 4 schools (Cohort 3) have a relatively high percentage (13%) of migrant students, more than twice the percentage of migrant students that are in YIP 6 (Cohort 1) schools and YIP 5 (Cohort2) schools. This has important implications for implementation and achievement because percent of migrant students is a strong negative predictor of achievement.

Table 1.3: Percent of Migrant Students in Reading First Schools, 2008

Years in Program	Number of Schools (with data available)	Percent of Migrant Students
YIP = 6, entered in 2003	255	4%
YIP = 5, entered in 2004	352	6%
YIP = 4, entered in 2005	**138**	**13%**
YIP = 3, entered in 2006	53	8%
YIP = 2, entered in 2007	34	8%
All YIPs	832	7%

Reading First in the Context of Program Improvement

Since Reading First is a federal initiative under NCLB, an increasingly important question is its relationship to supporting the goals of NCLB in reading/language arts. This question becomes more important as the measure of Adequate Yearly Progress (AYP) in NCLB has increased and more schools, and even numerous districts, are facing sanctions for failing to make AYP. All schools and LEAs that do not make AYP are identified for Program Improvement (PI) under the NCLB.

The accountability portion of NCLB can be briefly summarized as follows. The NCLB Act requires all states to implement statewide accountability systems based on challenging state standards in reading and mathematics, annual testing for all students in grades 3-8, and annual statewide progress objectives ensuring that all groups of students reach proficiency within 12 years. Assessment results are disaggregated by socioeconomic status, race, ethnicity, disability, and limited English proficiency to ensure that no group is left behind. LEAs and schools that fail to make AYP toward statewide proficiency goals are subject to improvement and corrective action measures.

In California, Program Improvement or PI is the formal designation for Title I-funded schools and LEAs that fail to make AYP for two consecutive years. A Title I school is identified for PI when, for each of two consecutive years, it fails to make AYP in the same content area (English-language arts or mathematics) school-wide or for any numerically significant subgroup, or on the same indicator (API or high school graduation rate) school-wide. Determinations are made using two years of data for schools and LEAs that receive Title I funds.

A PI school or LEA that makes AYP for one year will maintain the same PI status for one additional year and be required to continue implementing the applicable NCLB requirements. In order to exit PI, a school or LEA must make AYP for two consecutive years. Once in PI, a school or LEA that fails to make AYP will advance further in PI status with potential additional requirements and/or sanctions.

The goals of Reading First intersect with some of the requirements which have led schools to PI status, and this leads to questions regarding whether there is any discernable correlation between 1) years in program of a Reading First school and a PI school's exit of PI status, and 2) implementation of Reading First in a PI school and its exit from PI. Further analysis could be undertaken to review the impact of Reading First on the English learner and special needs populations in PI schools, since these two significant subgroups are frequent causes for a school's failure to make AYP and the corresponding assignment of PI status. A complete analysis of these issues is outside the scope of this report, but a preliminary examination of available data indicates that of the 863 Reading First schools, 518 are currently in PI status, while 123 have exited. Hence 641 Reading First schools have been or are in PI, or 74%. This high correlation is not unexpected, given the requirements for application to Reading First and

the achievement gaps necessitating PI status. Of the group of Reading First schools which have been in PI, the 123 schools which have exited PI constitute 19.2%. To put this figure into some context, of the 2,186 schools in California which have been or are currently in PI status, 571, or 26%, have exited.

Between the cohorts, a fluctuation in percentage exiting deserves further study. For example, Cohort 1 shows that 41 schools exited PI from a total of 255 in Program Improvement, leaving 214 in PI with 16% exiting. 62 exited PI from Cohort 2 from a pool of 276, leaving 214 in PI with 22% exiting. Of Cohort 3, 16 exited of 92 in PI, leaving 76 with 17% exiting, and in Cohort 4, 4 schools exited PI status from 17 total in PI, leaving 13 in PI with 23% exiting. While the size of the cohort and the number of years required to exit PI (two years making AYP are required) are factors which would have to be taken into account in gauging significance of these percentages, a study of the implementation of Reading First in those schools which exited would be needed to see if any significant impact of Reading First on PI status exists.

Teacher Qualifications in Reading First Schools

Table 1.4 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 bachelor's 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 95.29% in five years.

Table 1.4: Elementary Teacher Credentials and Experience 2003 – 2008

	Reading First Schools								All Elementary Schools[2]	
	Cohort 1		Cohort 2		Cohort 3		Cohort 4			
	2003	2008	2004	2008	2005	2008	2007	2008	2003	2008
Number of Schools	329	342	359	357	135	132	20	19	5647	6186
PhDs (%)	0.55	0.57	0.66	0.52	0.59	0.55	1.75	0.34	0.90	0.49
Masters plus 30 or more semester units (%)	9.39	18.62	13.73	13.37	16.28	13.08	14.06	16.03	14.00	14.48
Masters (%)	10.95	7.70	16.86	20.81	16.63	14.51	17.40	17.03	15.50	17.74
Bachelor's plus 30 or more semester units (%)	41.25	54.41	49.36	52.41	47.05	54.79	52.09	54.22	51.30	53.16
Total Advanced Degrees (%)	**62.14**	**81.29**	**80.61**	**87.12**	**80.56**	**82.94**	**85.30**	**87.63**	**81.70**	**85.87**
Bachelor's (%)	35.06	18.59	19.30	12.80	19.33	17.03	14.55	12.36	16.40	14.01
Less than Bachelor's (%)	.74	.09	.10	.07	.22	.02	.15	.00	.20	.09
Total Bachelor's or less (%)	**35.80**	**18.68**	**19.40**	**12.87**	**19.55**	**17.05**	**14.69**	**12.36**	**16.60**	**14.10**
Fully Credentialed Teachers (%)	77.80	95.29	93.73	98.26	92.05	95.14	98.03	95.34	90.90	97.14
Weighted Teacher Qualification[1]	2.01	2.27	2.26	2.36	2.31	2.25	2.36	2.37	2.20	2.34
Average years teaching	10.75	11.58	11.25	12.20	11.40	11.74	13.06	12.84	12.70	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 Bachelor's plus 30 or more semester units is given a weight of 2; and the percentage of teachers with Bachelor's 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.

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 1.4 shows that Cohort 1 Reading First schools had lower Weighted Teacher Qualification indices (2.01 to 2.27) than the other cohorts (ranging from 2.26 to 2.37) and the non-Reading First schools.

Conclusions

This chapter yields the following:

- For this Year 6 report, we have noted inherent difficulties in establishing adequate comparison groups; a statistically derived comparison group is used in the achievement analysis presented 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 90.98% in 2008 demographic files. Other cohorts ranged from 81.68% to 85.2%. The figure for All Elementary Schools was 54.12%.
- Reading First schools had higher percentages of ELs than the figure for All Elementary Schools (30.09%). Percentages of ELs in cohorts ranged from 33.21% to 57.66%.
- 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 suburban and rural designations. Cohort 4 had an even mix of urban and rural LEAs.
- Schools in the program 4 years (YIP = 4, roughly the same as Cohort 3) had a disproportionately large share of migrant students – 13% compared to 4% and 6% for YIP 6 (Cohort 1) and YIP 5 (Cohort2). This is important because percent of migrant students is a strong negative predictor of achievement and hinders Reading First implementation.
- 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 2007-08 CBEDS data, Cohort 3 Reading First schools had lower weighted teacher qualification indices than the other cohorts and the All Elementary Schools category.
- In 2008, 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 the reading achievement of 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? This chapter also continues an investigation into the degree to which Reading First influences achievement above grade 3. Last year's report looked at grade 4; the present report looks at grade 5. In addition, for the first time we attempt a "meta-analysis" of all possible Reading First effects that can be computed for the various achievement metrics from 2004 to 2008.[9] This allows us to calculate a single overall Reading First effect size that summarizes the total effect of the program to date with considerable precision. It also allows us to explore a possible "plateau effect" for schools that have been in the program longest, potential differences in how the various student cohorts have responded to the program, and Reading First effects for each grade.

The key findings in this chapter are:

- The Reading First Achievement Index (RFAI), a composite of K-3 achievement metrics for Reading First schools that ranges from 0 to 100, has risen an average of 3.1 points per year, equivalent to 18.6 points over 6 years relative to a starting year. This indicates a steady achievement gain in Reading First schools.
- 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 and Above" is matched or exceeded by a migration of students out of "Below or Far Below Basic." These migrations are confirmed by average student CST scale score gains on the order of 30 scale score points over a 6-year period, a strong and notable effect. Reading First continues to be very effective with low-performing students, in contrast with results in non-Reading First schools.
- Since 2002, Reading First schools have shown significantly more growth than either non-Reading First schools or the statistical control group across achievement metrics. High implementing schools show significantly higher gains than low implementing schools.
- Both the grade 4 and grade 5 achievement results show that K-3 participation in Reading First benefits students as they move into the upper grades. A meta-analysis of all grade-related effect sizes shows that the Reading First effect for grades 4 and 5 is on par with the effect for grades 2 and 3.

[9] Meta-analysis is a statistical technique for synthesizing the results of multiple studies.

- In line with predictions from previous reports, the growth rate of schools that have been in the program 6 years is not as great as in previous years, causing a flattening of achievement scores in high implementing schools and a narrowing of the difference between high and low implementation schools. This indicates a possible "plateau effect," a phenomenon observed in mature programs. However, while the YIP 6 schools may be encountering such an effect, the Reading First population as a whole is not. For the YIP 4 and 5 schools the effectiveness of Reading First has been increasing since 2006.
- The average Reading First (standardized beta) effect size in predicting all possible achievement outcomes since 2003, controlling for starting point and demographic factors, is 0.082 with a standard error of 0.004. This is approximately 16 standard errors higher than zero, where 2 standard errors above zero would be sufficient to claim a statistically significant effect with 95% confidence. This is a conservative estimate of the effect. If the definition of Reading First were to include Years in Program, the Reading First effect would nearly double in size.
- The Reading First effect is meaningful. Translated into real terms, the 0.082 Reading First effect means that a school that increases its RFII by 25 points for a given time period can expect, approximately, to double its achievement gains on a variety of grade 2-5 metrics over the same time period – a 100% increase in achievement gains. A school that increases its RFII by just 5 points can expect a 10-20% increase in its achievement gains.
- The Reading First standardized effect size is 63% as large as the average effect size for demographic variables widely thought of as important predictors of achievement. Translated into real terms, the decrease in grade 2 scale scores that would occur if the percentage of migrant students in a school increased by 10 percentage points would be almost exactly counterbalanced by increasing the school's RFII, its level of Reading First implementation, by 11 points. Thus, in this case, the negative effects of a percentage point increase in a demographic variable like Percent of Migrant Students are almost canceled by a one point increase in the RFII.

Achievement results for Reading First schools are presented in terms of the Standardized Testing and Reporting (STAR) Program assessments – the CST and the California Achievement Test (CAT/6) – and the Reading First End-of-Year (EOY) curriculum-embedded assessments. As of this report, grades 4 and 5 CST results are included to assess the sustained effects of Reading First. Achievement is compared in four ways:

- between years (gain scores)
- between Reading First and non-Reading First schools

- between Reading First schools and a statistical control group
- between high implementation and low implementation Reading First schools[10]

The objective of this evaluation is to determine whether or not, and to what degree, the Reading First program is effective as implemented in California. What is meant by "effective"? According to the federal guidelines for Reading First, the program is effective to the degree it ensures "that every student can read at grade level or above not later than the end of Grade 3" (U.S. Department of Education, 2002). There are several ways to examine the effect of Reading First on reading in California given the limitations of a non-experimental design:

- Measure the size of the achievement gains of the Reading First schools for grades K-3 and beyond
- Compare Reading First schools to comparable non-Reading First schools
- Compare Reading First schools to a "statistical control group" using statistical methods to profile how a school that is similar to Reading First schools would perform without access to the program
- Compare high implementation Reading First schools to low implementation Reading First schools

The first approach looks at the absolute size of the achievement gains of Reading First schools from the level of performance immediately preceding entry into Reading First (i.e., when implementation had not yet occurred) to the present, when the program has been in place and is presumably well implemented. A significant positive gain would suggest the 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 due to the lack of comparable non-Reading First schools as explained in Chapter 1 of this report, although some specifically limited comparisons were given in the March 2008 Reading First Supplemental Survey Report.[11] Given the constraints of the study, it is not possible to identify non-Reading First schools that are not to some degree employing the same program elements that are required of Reading First schools, making comparisons between them problematic.

[10] 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.

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

The statistical control group approach employed in the Years 4, 5 and 6 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 statistical control group gains are 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.

To these four approaches we add a fifth that becomes possible as data is accumulated through the years – synthesize all the effect sizes calculated using multiple regression to derive an overall effect size, and determine whether that effect size is significantly greater than zero.

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

Measures of Achievement Gains

School progress or growth, also called achievement gains, is measured using the CSTs, the CAT/6 Mean Percentile Ranks (called here "MeanPR"), the Reading First EOY tests, and the Reading First RFAI, which is a composite of the others and is used to make decisions about continued Reading First funding for LEAs. Each metric has unique characteristics described below.

The California Standards Test (CSTs). The CSTs are administered to all California students in grades 2 and above toward the end of the school year. We use the English language arts (ELA) component of the

CSTs for grades 2, 3, 4 and 5. The inclusion of grade 4 commenced with the Year 5 Report and is continued in the Year 6 Report because students in grade 4 can be expected to have experienced Reading First since kindergarten. Grade 5 is included for the first time in the Year 6 report. Students that were in kindergarten when Reading First was first implemented by the Cohort 1 LEA's moved into grade 5 in 2008.

We examine the achievement of students in ELA in three ways, described below. The first two are a simplification of the five CST performance categories (Advanced, Proficient, Basic, Below Basic, Far Below Basic). The third is a scale score derived from the ELA scores.

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 the No Child Left Behind (NCLB) Act.

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.[12] 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. This scale score is a number ranging from 200 to 500, which describes a student's performance on a test in a way that facilitates valid comparisons. Using scale scores (which have equal intervals and use information at all parts of the scale equally) to measure growth reduces anomalies caused by the nonlinearities present in percentage-based scales and particularly reduces anomalies caused when a given student distribution happens to fall near a relative cut point, for example, near the "Proficient" cut-point. We introduced scale scores in the Year 5 Report and include them in the Year 6 Report to avoid the 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

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

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 inclusion of this metric takes on greater importance relative to the April 2008 National Reading First impact study which was discussed in Chapter 1.

The CST gain score reported in the tables of this chapter is the 2008 percentage of students in a specified category minus the corresponding percentage in the year immediately *preceding* the first year of Reading First funding. The change in scale scores is calculated using the same time frame. The gain scores are averaged across a specified population of schools to produce the tabular statistics presented in this chapter.

CAT/6 MeanPR. 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 2008 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 (that is, classrooms in which instruction is conducted in Spanish, using State Board of Education-adopted Spanish translations of the adopted reading programs, by permission of a waiver). The EOY score for each grade within a school consists of the percentage of students that meet the benchmark established for that grade based on national norms recommended by Hasbrouck & Tindal (2005). The gain score for that grade is its 2008 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 5-6 years is 2004. For schools in the program 4 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, one cohort of students (YIP 6, see explanation below) has five years of RFAI data (2004, 2005, 2006, 2007, and 2008), as compared to seven 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 2008 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 prior reports, there are cases where LEAs that received funding starting in one year added schools to Reading First in a later year. For purpose of measuring program effects, it 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 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 6 is notably more urban than YIP 5 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 2008, we focus on those Reading First schools that have been in the program for 6 years (the longest), 5 years, 4 years, and 3 years (YIPs 6, 5, and 4, and 3). We have omitted schools in YIPs 1 and 2. They have relatively few schools (65 combined).

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. Additionally, the Spanish version of the EOY test became available in 2005.

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.[13]

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, 2007 and 2008 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 in two ways for the Year 5 Report and the current, Year 6, Report. The first change was to define a school's RFII for a

[13] 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.

given year to be a rolling 2-year average of the "preliminary RFIIs" calculated for each year from the survey results for that year. Thus, acting on advice of the EAG, when reporting an individual RFII for each school we average its preliminary RFII (computed from the 2008 surveys) and its preliminary RFII from 2007 on the theory that a rolling 2-year average is more stable and reliable than the RFII computed from a single year's data, but more responsive to school implementation efforts than an average of all the years' RFIIs.

The second change was to define a high implementation school as one whose RFII (the average of the preliminary RFIIs for the current year and previous year) is greater than 1 standard deviation above the original 36.0 cut-point -- approximately 41.4. A low implementation school is one whose RFII is less than 36.0.[14] 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 2008 is not comparable to that in 2006 or earlier.

Calculating Achievement for the Statistical Control Group

As discussed in prior reports, the statistical control group is defined using regression models to calculate the 2008 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 2008 achievement score and gain score for the statistical control group. As stated previously, the statistical control group is not a literal group of schools but an extrapolation based on a relationship between achievement and implementation derived statistically from the Reading First schools. (Non-Reading First schools could not be used to compute this relationship since they do not take the surveys and do not receive an RFII.) For additional background reference regarding the detailed procedure for computing the statistical control group achievement statistics, the reader is referred to Chapter 4 of the Year 4 Report.[15]

[14] 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.

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

Achievement Results

The following pages present a series of tables and trend-line charts showing starting scores, ending (2008) scores, and gains on each of 12 achievement metrics. They are the heart of the Year 6 Report and the basis of our finding that Reading First in California continues to be an effective program. Table 2.1 summarizes the gains of *all* Reading First schools taken as a whole, not broken out by 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 up to 2008" 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, 5 and 6.

Tables 2.3, 2.4, 2.5, and 2.6, with accompanying trend-line charts, show total gains on the CST and CAT/6 metrics for YIP 6, grades 2, 3, 4 and 5. Similar tables and charts for YIPs 3, 4 and 5 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:

- 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.
- 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 for each achievement metric, from the year they entered the program to 2008. As such, it summarizes the gains in the Year 6 report and is one way to address the question, "What has been the effect of Reading First on all schools currently in the program?" Because it combines all six 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.

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 5 number of schools reflects the fact that grade 5 data were collected only for YIP 6 schools. Grade 4 data were collected only for YIP 5 and YIP 6 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.

With this year's introduction of "meta-analysis" to summarize the Reading First effect in all its manifestations across the life of the program, the Summary Gains table is no longer the preferred way to draw general conclusions about the Reading First effect. It suffers at least one important conceptual shortcoming – its reliance on the current year (2008) achievement outcomes to define the regression models used to calculate the statistical control group. This makes it vulnerable to any specific issues or interaction effects involving the 2008 achievement metrics. Another limitation is its assignment of schools to the high implementation and low implementation categories based on their 2008 RFII (a rolling average of its 2007 and 2008 preliminary RFII statistics). This does not take into account the full implementation history of each school and overlooks relationships between implementation and achievement from previous years. It also leaves the analysis vulnerable to issues involving the 2008 RFII. Therefore, Summary Gains Table 2.1 should not be used to draw conclusions about Reading First that extend beyond the 2008 Year 6 Report, or seek to say anything about Reading First overall. That is a task best left to the meta-analysis provided at the end of this chapter.

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

All Schools, All Grades, 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=792	N=146	N=258	N/A	N=4057
% Proficient and Above	3.1[abc]	3.1[bc]	2.8[bc]	2.7	2.2
% Below and Far Below Basic	-3.5[abc]	-3.6[abc]	-3.5[bc]	-3.0	-1.6
Scale Score Metric	4.6[abc]	4.6[bc]	4.5[bc]	4.1	3.2
Grade 3, CSTs	794	147	259	N/A	4054
% Proficient and Above	1.4[abc]	1.8[abc]	1.1[bc]	0.9	0.3
% Below and Far Below Basic	-3.1[abc]	-3.8[abc]	-2.7[bc]	-2.5	-1.1
Scale Score Metric	3.0[abc]	3.7[abc]	2.4[bc]	2.2	0.8
Grade 3, CAT6, Mean Percentile Rank	794	147	259	N/A	4055
Reading, Mean PR Metric	0.9[abc]	1.2[abc]	0.7[bc]	0.5	0.0
Language, Mean PR Metric	1.2[abc]	1.5[abc]	0.9[bc]	0.9	0.4
Spelling, Mean PR Metric	2.5[abc]	2.9[abc]	2.1[bc]	1.9	1.1
Grade 4, CSTs	584	101	186	N/A	3996
% Proficient and Above	3.5[bc]	3.5[bc]	3.4[bc]	3.5	2.9
% Below and Far Below Basic	-3.3[bc]	-3.0[bc]	-3.1[bc]	-3.3	-1.5
Scale Score Metric	4.1[bc]	4.0[c]	3.9[c]	4.1	3.7
Grade 5, CSTs	238	26	87	N/A	3988
% Proficient & Above	3.1[abc]	4.0[abc]	2.8[ac]	2.3	2.8
% Below Basic & Far Below Basic	-3.1[abc]	-4.0[abc]	-2.5[bc]	-2.7	-1.3
Scale Score Metric	3.6[abc]	4.7[abc]	3.0[c]	2.8	3.0
Reading First Achievement Index	819	150	266	N/A	
RFAI metric	3.1[ac]	3.3[ac]	2.8[ac]	2.3	

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."

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

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

[1] The grade 5 sample includes only YIP 6 schools, the grade 4 sample YIPs 5 and 6. Hence, the smaller N's.

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 score as of 2008, 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 6 to get a 6-year expected gain relative to a starting year.

In the Year 6 Report the differences between the Low Implementation and High Implementation groups, and between the statistical control group and All Reading First schools, appear in some cases to be

smaller than they were in the Year 5 Report. This is true of grade 2 and particularly grade 4, where the Reading First effect appears to have vanished entirely. This pattern would be consistent with a diminishing of the Reading First effect with time, possibly a "plateau effect."

At the same time, the grade 3 differences between the implementation categories have sharpened relative to the Year 5 Report, both on the CST and the CAT/6 achievement metrics, a new pattern relative to previous years. We have also introduced grade 5 summary gains for the first time, and these show strong implementation effect differences.

It is by no means easy or straightforward to interpret the cross-year fluctuations in grades 2, 3 and 4, made more difficult by the afore-mentioned conceptual shortcomings embodied in this table. An attempt has been made to explore and dissect these fluctuations systematically through the meta-analysis presented later in this chapter. This meta-analysis finds that the Reading First effect has actually *increased* overall in Year 6 relative to Year 5, and that the apparent lack of a Reading First effect for grade 4 as manifest in Table 2.1 is actually a property of the particular student cohort that happened to pass through grade 4 in 2008.

Otherwise, the findings are generally in line with previous years' findings. 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.

RFAI Gains (Table 2.2)

Table 2.2 reports starting points, ending points, and total RFAI gains for YIP 3, 4, 5, and 6 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 4, 5, and 6

	Reading First Schools			
	All Reading First Schools	High Implementation Schools	Low Implementation Schools	Statistical Control Group
Year in Program: 6				
Number of Schools	255	28	97	N/A
2004	36.6	38.1	36.1	36.6
2008	47.7	49.9	47.1	45.8
RFAI Gain	**11.1**[ac]	**11.8**[ac]	**11.1**[ac]	**9.2**
Year in Program: 5				
Number of Schools	371	75	119	N/A
2004	34.5	35.7	33.2	34.6
2008	44.9	46.9	42.7	44.0
RFAI Gain	**10.4**[ac]	**11.2**[ac]	**9.5**[c]	**9.5**
Year in Program: 4				
Number of Schools	151	26	57	N/A
2005	34.4	37.0	31.0	34.9
2008	42.8	44.6	40.2	42.5
RFAI Gain	**8.4**[ac]	**7.5**[c]	**9.3**[ac]	**7.5**

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

The RFAI gains for YIPs 5 and 6 support the hypothesis that Reading First schools are growing, that they grow more quickly than the statistical control group, and that high implementing schools grow faster than low implementing schools. The schools in YIP 4 offer an exception to the pattern, however, as they did in 2007. 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. Also, the YIP 4 schools have a much higher migrant student population (13%) than the YIP 5 schools (6%) and the YIP 6 schools (4%). This is an as yet unexplored demographic difference that warrants attention, and explains why we have added "Percentage of Migrant Students" as a control variable in our regression equations.

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 six years.

Table 2.3: CST Metric, YIP = 6, Grade = 2

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

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."

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

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

The percent Proficient and Above has risen from 15.5 percentage points in 2002 to 35.8 percentage points in 2008, continuing a strong growth trend, although the gain is somewhat less from 2007 to 2008 than it was in previous years. The growth rate is equivalent to a gain of 30 scale score points on the grade 2 CST over six years, or 5 scale score points per year. 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 were to continue their current growth trajectory, they would 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 can interpret this rate of growth to mean that Reading First schools are now moving each new cohort of students 30 scale score points further up the scale from kindergarten to grade 2 than they were, with similar cohorts, six 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 6 alone, the pace of change is considerable by any standard.

Due to common elements in statewide reading instructional implementation, we see that the rest of the state's elementary schools have also shown significant growth, but they lag behind Reading First schools by 10.7 scale score points. We also see that lower performing students are moving out of the bottom performance levels at an even higher rate than 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 more than half the rate that students enter the top categories. This remains 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 3.6 scale score points higher than for Reading First schools as a whole, more than 4 scale score points higher than for low implementing schools. This demonstrates that fidelity or depth of implementation makes a measurable improvement in achievement.

On the other hand, as Figures 2.3a, 2.3b, and 2.3c show, the difference between high and low implementing schools is smaller than it was in 2007 reflecting a possible diminution in the Reading First effect for YIP 6 schools. This is especially evident in the flattening achievement curves of the high implementing schools for the Proficient and Above metric. While the low implementing schools and the Reading First population as a whole continue to grow, the high implementing schools appear to have hit a ceiling, as have the non-Reading First schools. But we need to treat this flattening with caution. First, without knowing its cause it is not valid to assume a "ceiling." Second, it is not apparent in the scale score curves shown in Figure 2.3c. It may even be a statistical artifact of the Proficient and Above metric, which can be highly nonlinear and sample dependent. These patterns are addressed more rigorously in the meta-analysis section of this chapter under the heading of "plateau effects."

In the figures below, 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 = 6, Grade = 2

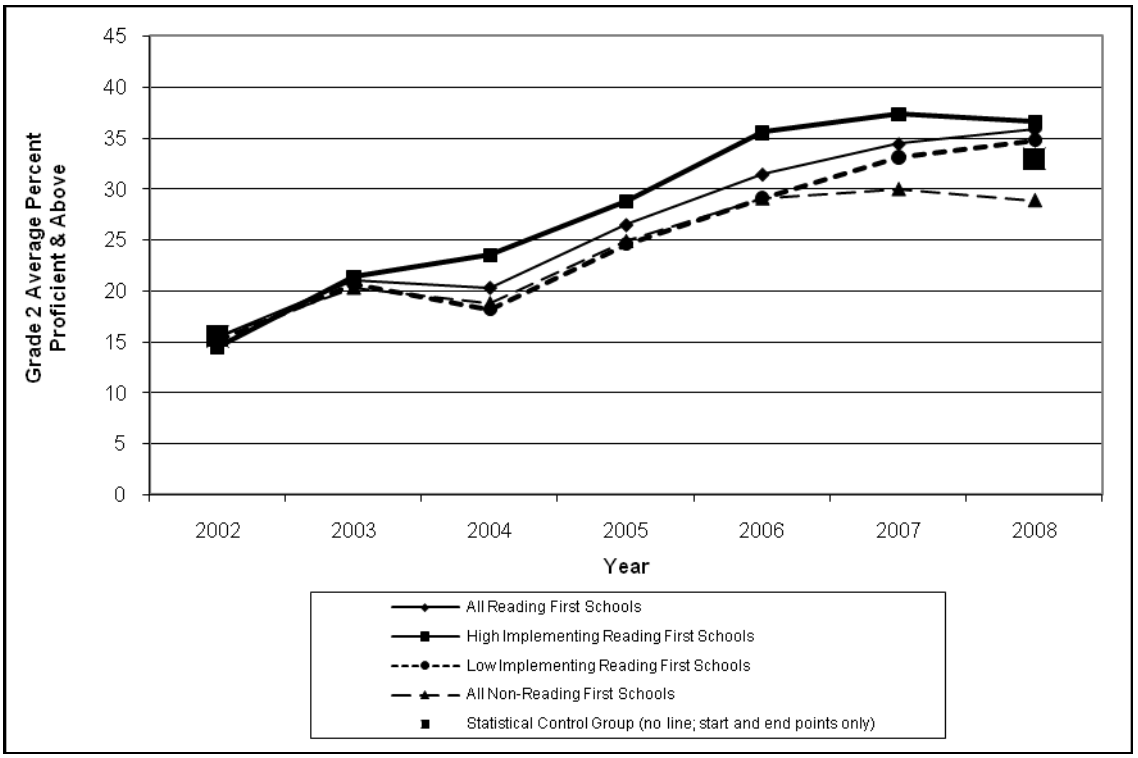


Figure 2.3b: CST % Below and Far Below Basic, YIP = 6, Grade = 2

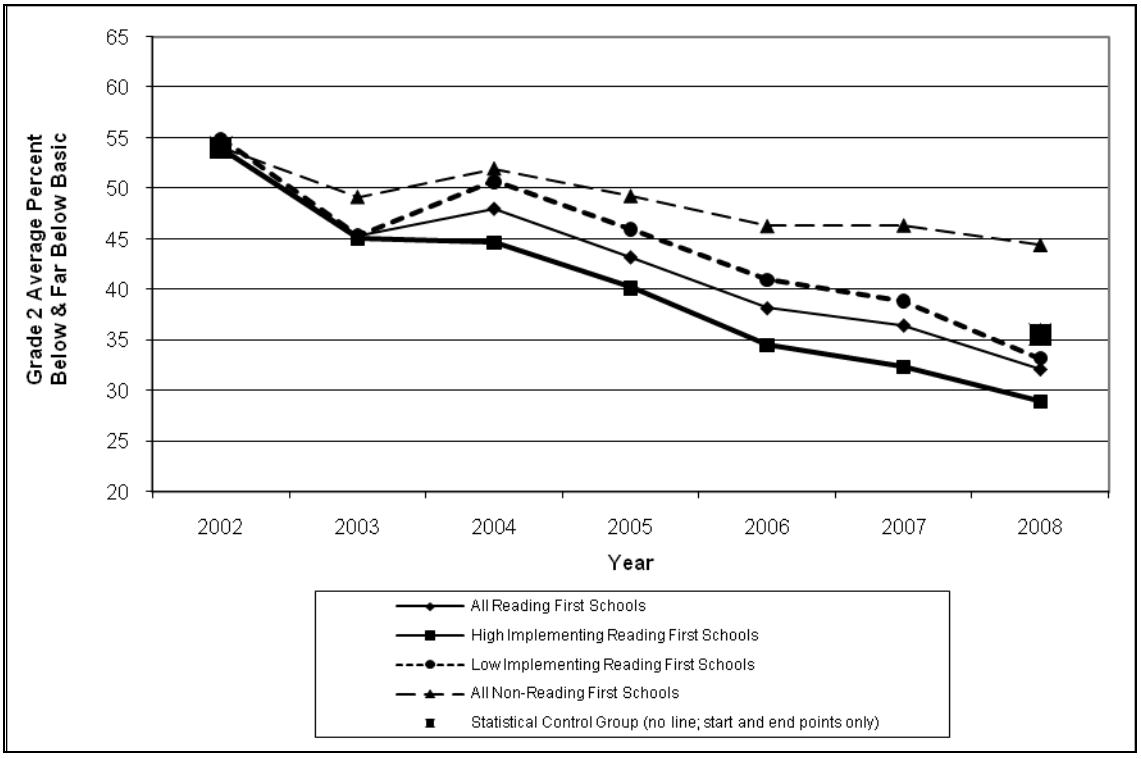
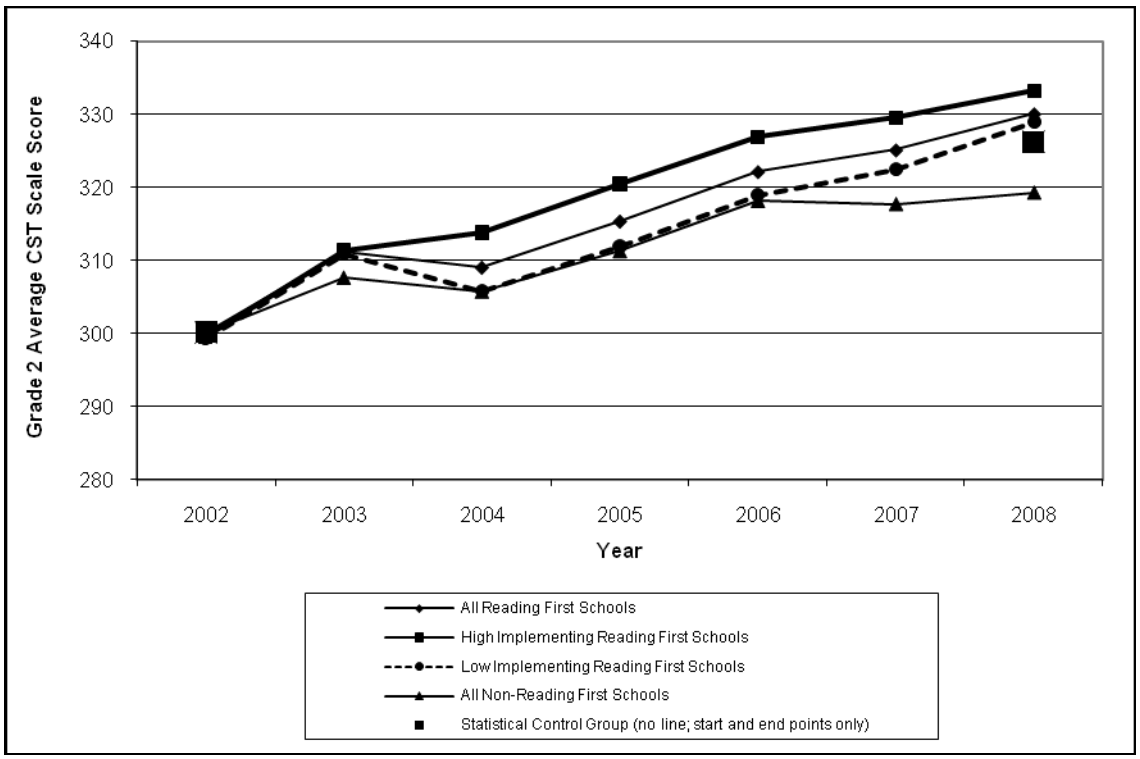


Figure 2.3c: CST Mean Scale Score, YIP = 6, Grade = 2



In addition to the patterns discussed above, we see that growth on the grade 2 scale score metric has, with the exception of 2004, been fairly steady. We see that the high and low implementation schools started at approximately the same location on the scale and fanned out according to their level of implementation. 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.

However, also notice that the growth from 2007 to 2008 shows a *narrowing* between the high and low implementing Reading First schools. This, along with the flattening of some of the other curves, revives the hypothesis of a possible "plateau" effect which we have predicted in previous reports and which is explored in detail later in this chapter.

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 = 6, Grade = 3

Years in Program: 6 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	253	28	96	N/A	4,054
% Proficient and Above					
2002	14.9	13.0	15.0	14.9	39.9
2008	21.7	23.7	21.7	19.3	41.6
Change Since Starting Year	**6.8[abc]**	**10.7[abc]**	**6.7[abc]**	**4.4**	**1.7**
% Below or Far Below Basic					
2002	57.6	58.6	57.2	57.6	31.4
2008	41.6	37.9	42.4	44.3	24.8
Change Since Starting Year	**-16.0[abc]**	**-20.6[abc]**	**-14.8[bc]**	**-13.3**	**-6.6**
Mean Scale Score Per Student					
2002	294.9	293.3	295.1	294.9	333.6
2008	310.1	313.5	309.8	306.8	338.2
Change Since Starting Year	**15.2[abc]**	**20.2[abc]**	**14.7[bc]**	**11.9**	**4.6**
CAT/6 Mean Percentile Rank Reading					
2002	22.7	22.0	22.4	22.7	45.6
2008	27.6	28.8	27.5	25.9	45.9
Change Since Starting Year	**5.0[abc]**	**6.8[abc]**	**5.2[abc]**	**3.3**	**0.2**
CAT/6 Mean Percentile Rank Language					
2002	24.9	24.2	25.3	24.9	44.5
2008	31.3	32.5	31.6	30.1	47.1
Change Since Starting Year	**6.3[abc]**	**8.4[bc]**	**6.3[bc]**	**5.1**	**2.7**
CAT/6 Mean Percentile Rank Spelling					
2002	36.7	35.1	36.9	36.7	52.0
2008	52.1	53.2	51.4	48.6	58.4
Change Since Starting Year	**15.3[abc]**	**18.1[abc]**	**14.4[abc]**	**11.8**	**6.3**

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."

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

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

While absolute gains in the grade 3 metrics tend to be smaller than those for other grades, reflecting a state-wide pattern, the Reading First gains are quite large relative to those for non-Reading First schools. Movement out of the bottom categories is particularly impressive. Figures 2.4a – 2.4f reveal that grade 3 has a complexity not shared by the other grades, both in terms of the long-term growth trajectory (largely

an artifact of the Grade 3 CST) and in terms of an apparent "plateau effect" among the high-implementing schools.

Figure 2.4a: CST % Proficient & Above, YIP = 6, Grade = 3

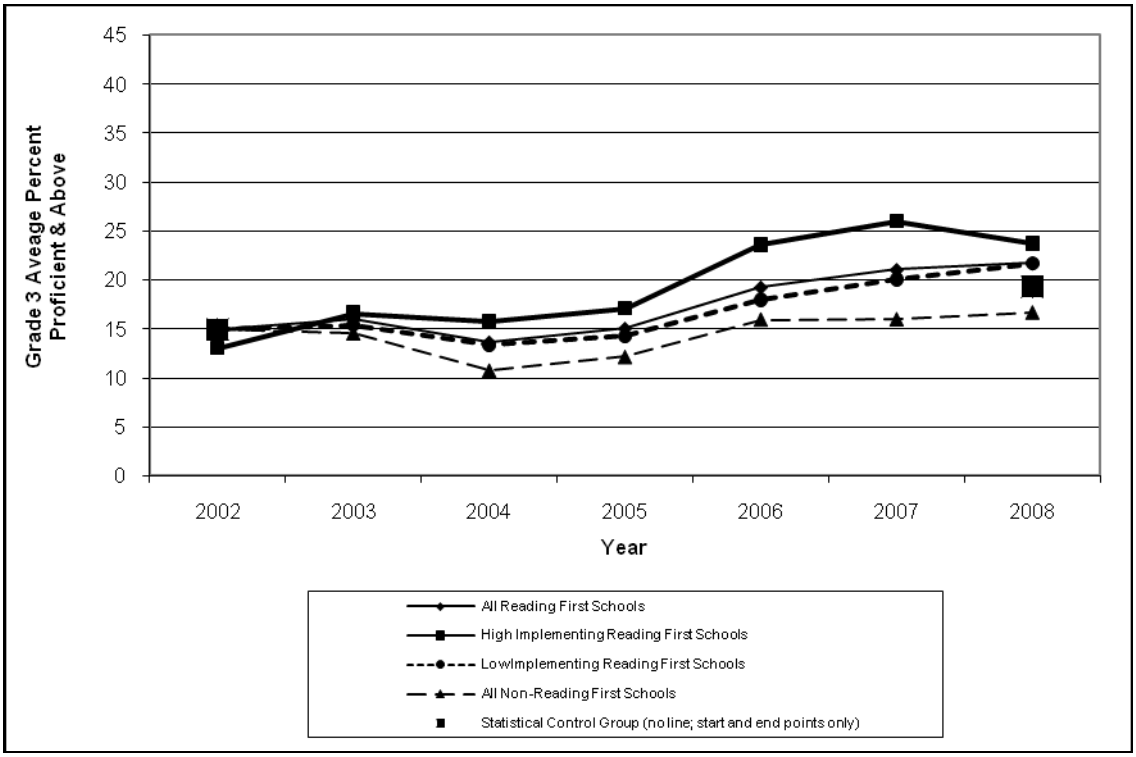


Figure 2.4b: CST % Below and Far Below Basic, YIP = 6, Grade = 3

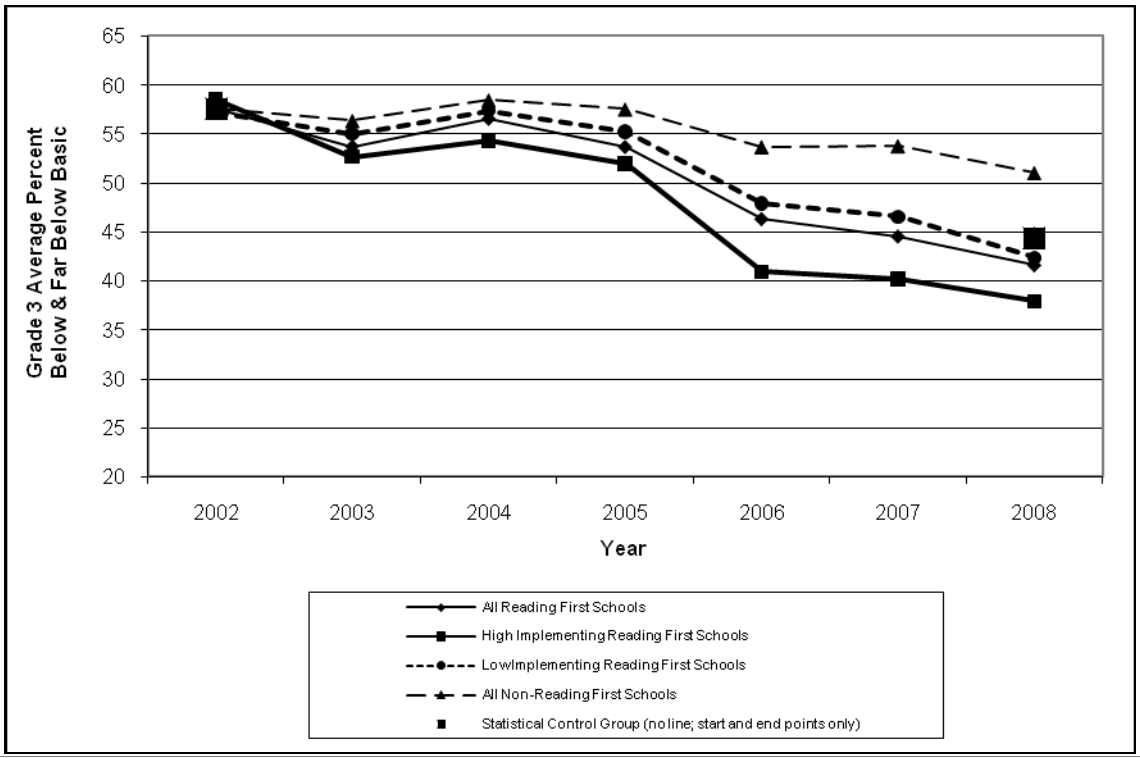


Figure 2.4c: CST Mean Scale Score Per Student, YIP = 6, Grade = 3

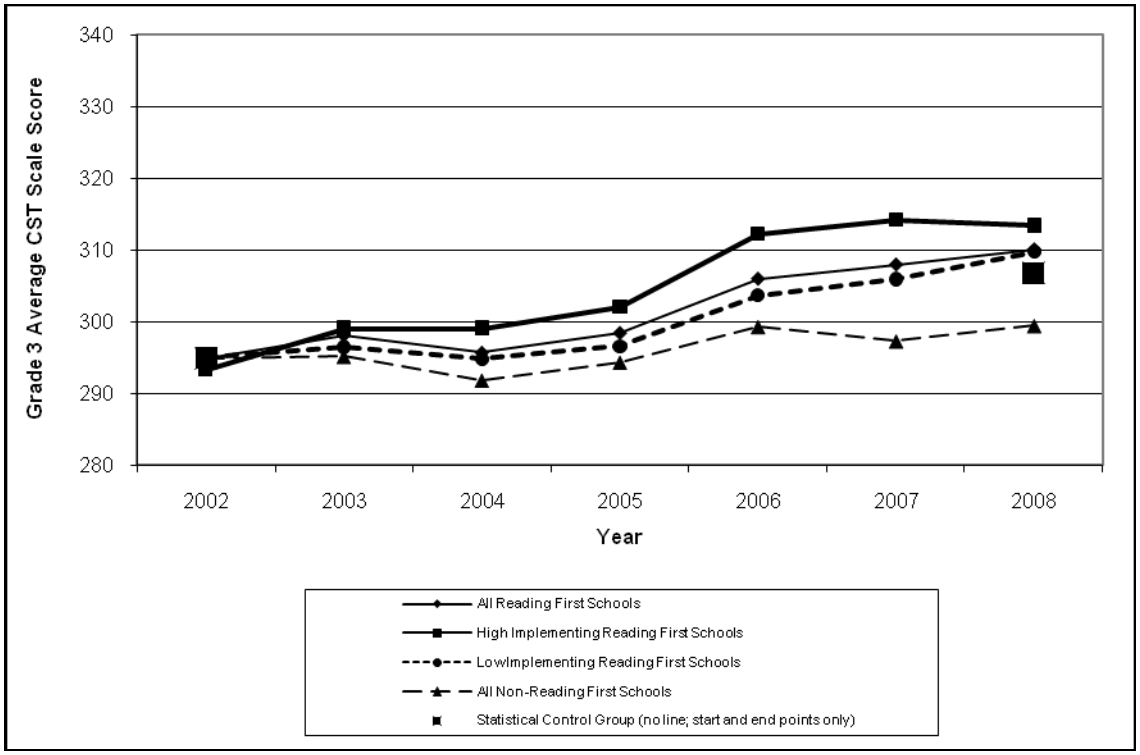


Figure 2.4d: CAT/6 Reading, Mean Percentile Rank, YIP = 6, Grade = 3

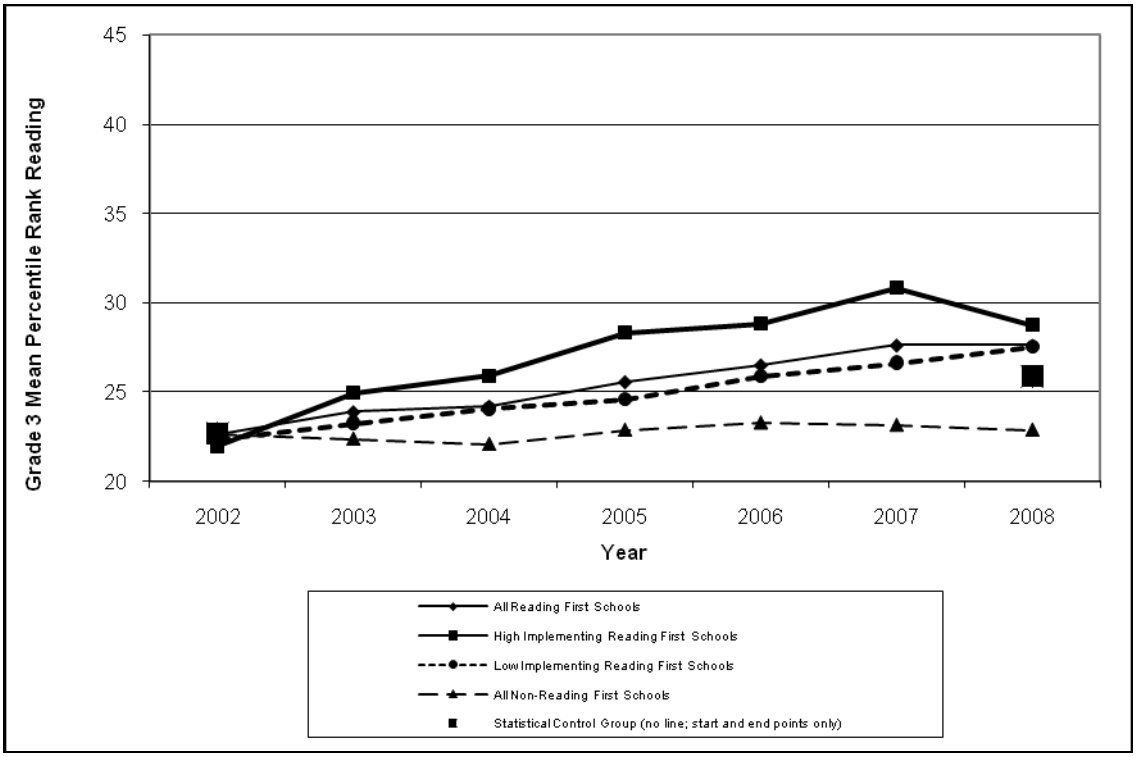


Figure 2.4e: CAT/6 Language, Mean Percentile Rank, YIP = 6, Grade = 3

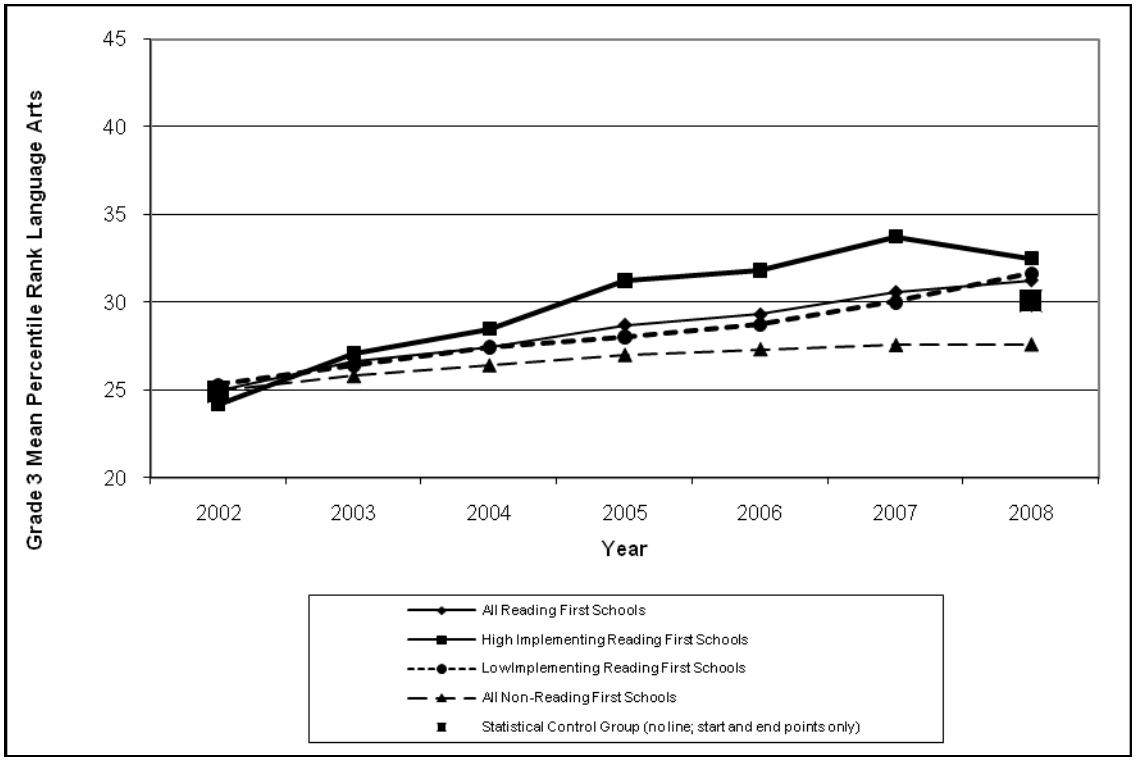
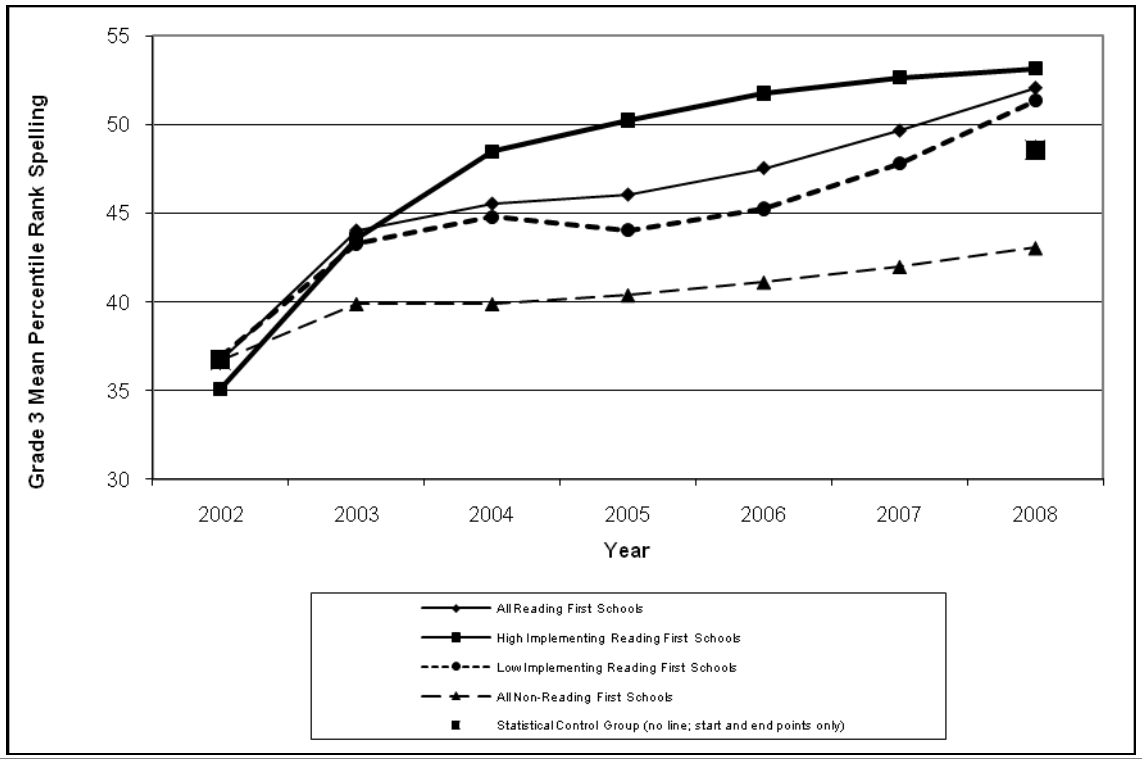


Figure 2.4f: CAT/6 Spelling, Mean Percentile Rank, YIP = 6, Grade = 3



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 flattening from 2007 to 2008 for the high implementing schools.

It is this flattening effect, again, together with the narrowing of the Reading First curves, which is most striking in 2008. Unlike grade 2, in grade 3 it appears in the scale score metric as well as most of the other achievement metrics, including Reading and Language in the CAT/6. This suggests that it is not a statistical artifact but something more fundamental.

Another 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.

Another 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, but that it is diminishing for YIP 6. This finding is explored further in the meta-analysis section.

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 6 schools. (The grade 2 and grade 3 results for YIPs 3, 4 and 5 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 it moves into the upper grades. This continues to prove a telling indicator of

Reading First effects since it supports the hypothesis that students in Reading First classrooms are learning skills and mastering concepts that generalize beyond the course content and test material of the first few grades in elementary school, showing replicable and sustainable patterns of achievement.

Table 2.5: CSTs, YIP = 6, Grade = 4

Years in Program: 6 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	249	26	96	N/A	3,996
% Proficient and Above					
2002	15.4	15.0	16.2	15.4	41.9
2008	37.2	41.2	36.1	36.3	59.1
Change Since Starting Year	**21.7[bc]**	**26.2[abc]**	**19.8[bc]**	**20.8**	**17.1**
% Below or Far Below Basic					
2002	47.4	47.9	46.1	47.4	23.4
2008	25.8	24.3	26.8	27.5	14.1
Change Since Starting Year	**-21.6[bc]**	**-23.6[bc]**	**-19.3[bc]**	**-20.0**	**-9.2**
Mean Scale Score Per Student					
2002	307.1	305.7	308.2	307.1	340.6
2008	334.1	337.0	333.2	332.4	362.9
Change Since Starting Year	**26.9[bc]**	**31.3[abc]**	**25.0[bc]**	**25.3**	**22.3**

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."

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

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

Grade 4 reinforces the growth picture presented by the grade 2 trend-lines, though it is less pronounced in 2008 than it was in 2007. We hypothesize later in the chapter that this is a student cohort effect. Nonetheless, Reading First schools continue to grow faster than the control group. High implementing schools grow faster than low implementing schools. Movement out of the bottom two categories matches movement into the top categories, unlike non-Reading First schools. The average scale score growth is 26.9 points over six years, not far shy of the 30 points seen in grade 2. What makes this table highly significant is that Reading First is only administered in grades K-3. There is no grade 4 Reading First program: yet as we saw last year, and continue to see this year, the CST scores are almost what one would expect if Reading First extended to grade 4. This continues to demonstrate that Reading First students have been able to carry with them the skills, reading habits and conceptual understanding that they developed in the earlier grades, and that rigorous instruction in the lower grades lays the groundwork for large gains in the higher grades.

Non-Reading First schools also show substantial gains over this period, but the gains are smaller and less uniform across the population as can be seen in Figures 2.5a – 2.5c.

Figure 2.5a: CST % Proficient & Above, YIP = 6, Grade = 4

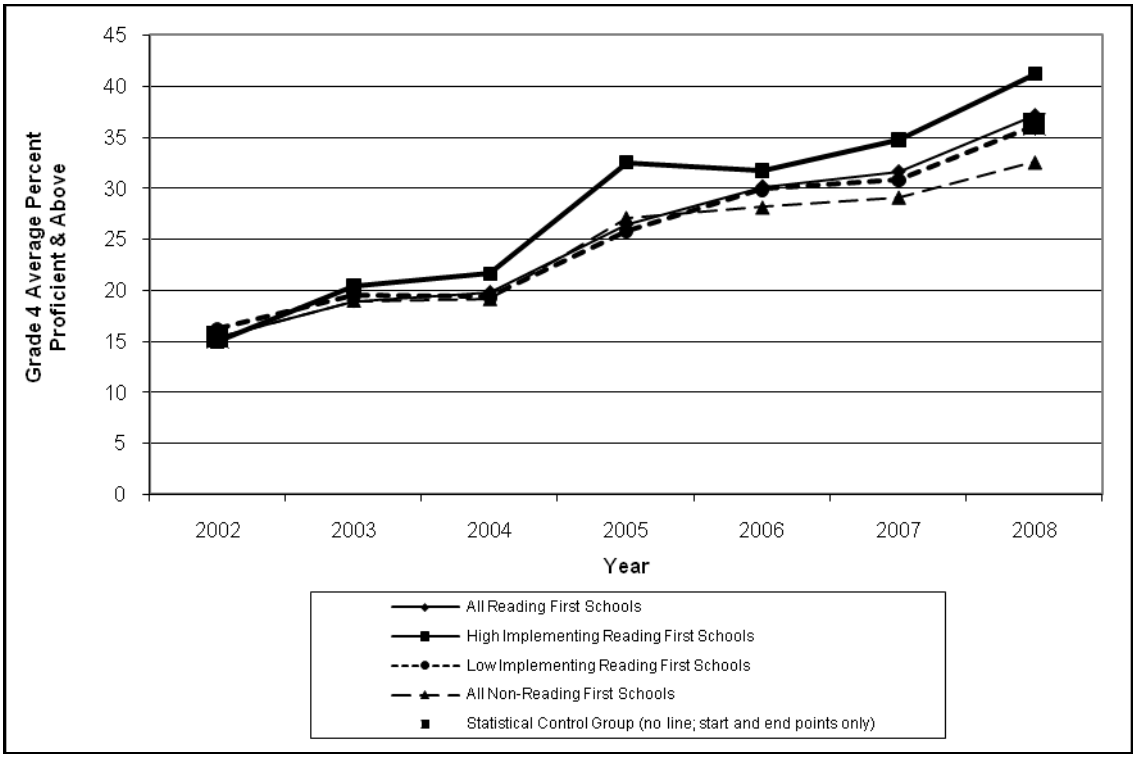


Figure 2.5b: CST % Below and Far Below Basic, YIP = 6, Grade = 4

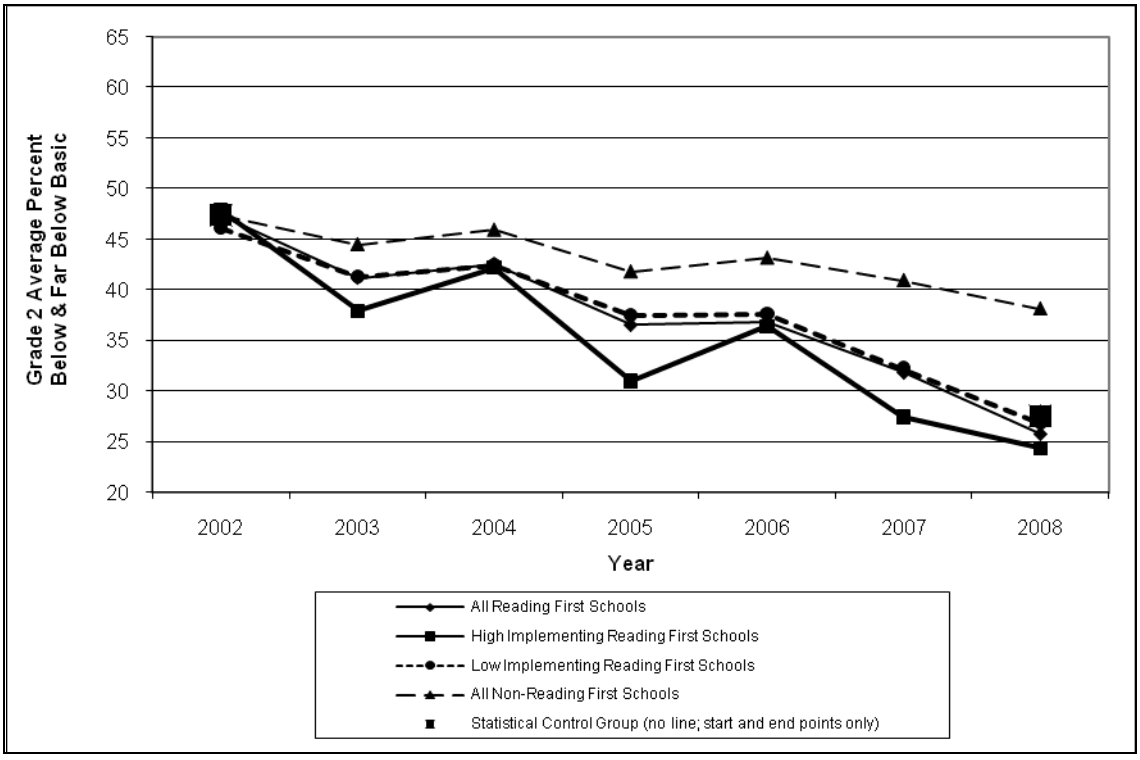
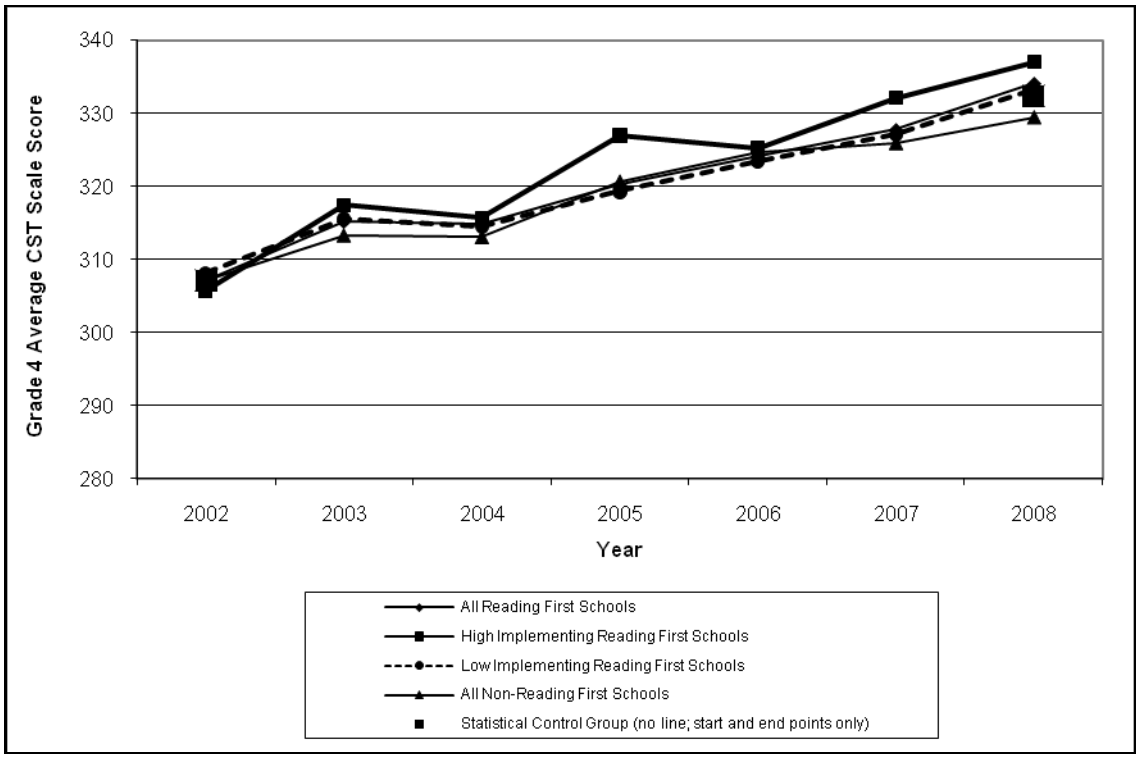


Figure 2.5c: CST Mean Scale Score, YIP = 6, Grade = 4



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. However, Figure 2.5a does show a widening gap separating high implementing schools from the rest. In Figure 2.5c, all the curves show robust growth, including the non-Reading First schools, but the differential effect of Reading First is less apparent.

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 continue to 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 enjoy a decisive advantage and possess the skills to keep up with their peers.

CST Results for Grade 5 (Table 2.6 and Figures 2.6a – 2.6c)

Table 2.6 reports the CST results for grade 5 which have been collected only for YIP 6 schools. 2008 was the first year that students who have been in Reading First since kindergarten moved into grade 5. Note that they are the same students who were in grade 4 last year and provided evidence of a strong post-grade 3 Reading First effect.

Table 2.6: CSTs, YIP = 6, Grade = 5

Years in Program: 6 **Grade: 5**	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	238	26	87	N/A	3,988
% Proficient and Above					
2002	11.1	11.5	11.8	11.1	35.8
2008	29.9	35.5	28.6	26.7	52.3
Change Since Starting Year	**18.8**[abc]	**24.0**[abc]	**16.8**[c]	**15.7**	**16.5**
% Below or Far Below Basic					
2002	50.0	49.1	48.7	50.0	24.3
2008	31.3	25.0	33.8	34.3	16.6
Change Since Starting Year	**-18.7**[abc]	**-24.1**[abc]	**-15.0**[bc]	**-15.7**	**-7.7**
Mean Scale Score Per Student					
2002	303.4	303.7	304.8	303.4	334.5
2008	324.9	331.7	322.9	321.0	352.3
Change Since Starting Year	**21.5**[abc]	**28.0**[abc]	**18.1**[c]	**17.6**	**17.9**

[a] Significantly different ($p < 0.05$) relative to the "Statistical Control Group."

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

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

Grade 5 is striking in the way grade 4 was striking in 2007, perhaps due to having the same student cohort (see meta-analysis later in the chapter). The overall size of the gains is less in grade 5 than in grades 4 and 2, and the Reading First schools are more similar to the non-Reading First schools. There is no special program being implemented in grade 5 that would differentiate it from the rest of California schools. But the Reading First effect is nonetheless quite pronounced. There is a 10 point scale score difference between high and low implementing schools for grade 5 compared to a 6 point difference for grades 3 and 4, and a 4 point difference for grade 2. While the gains are similar between non-Reading First schools and the low implementing schools on the Proficient and Above and Mean Scale Score metrics, the movement out of the "Below or Far Below Basic" categories is twice as large in the Reading First schools, three times as large for high implementing schools. Movement into the top categories is balanced by movement out of the bottom categories.

Figures 2.6a – 2.6c show that as students who entered Reading First classrooms in 2003 moved into grade 5, high implementing schools start showing substantially higher grade 5 CST scores. The effect disappears in 2007, but breaks out dramatically in 2008.

Figure 2.6a: CST % Proficient & Above, YIP = 6, Grade = 5

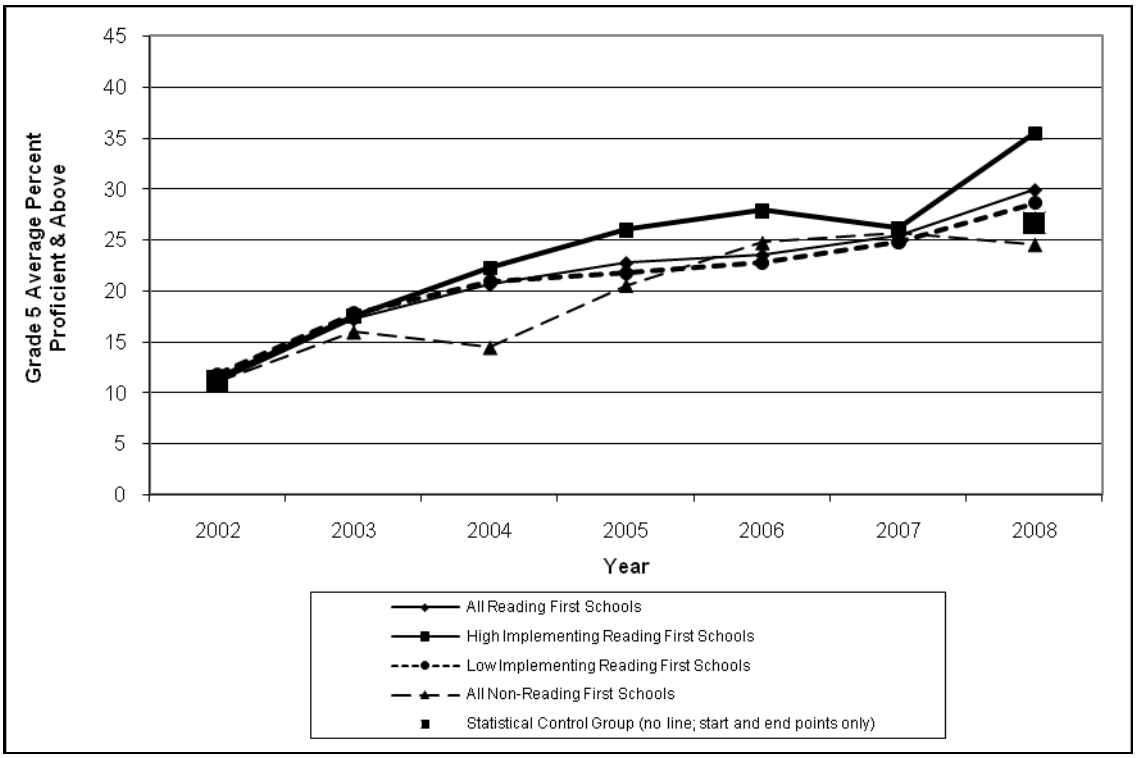


Figure 2.6b: CST % Below and Far Below Basic, YIP = 6, Grade = 5

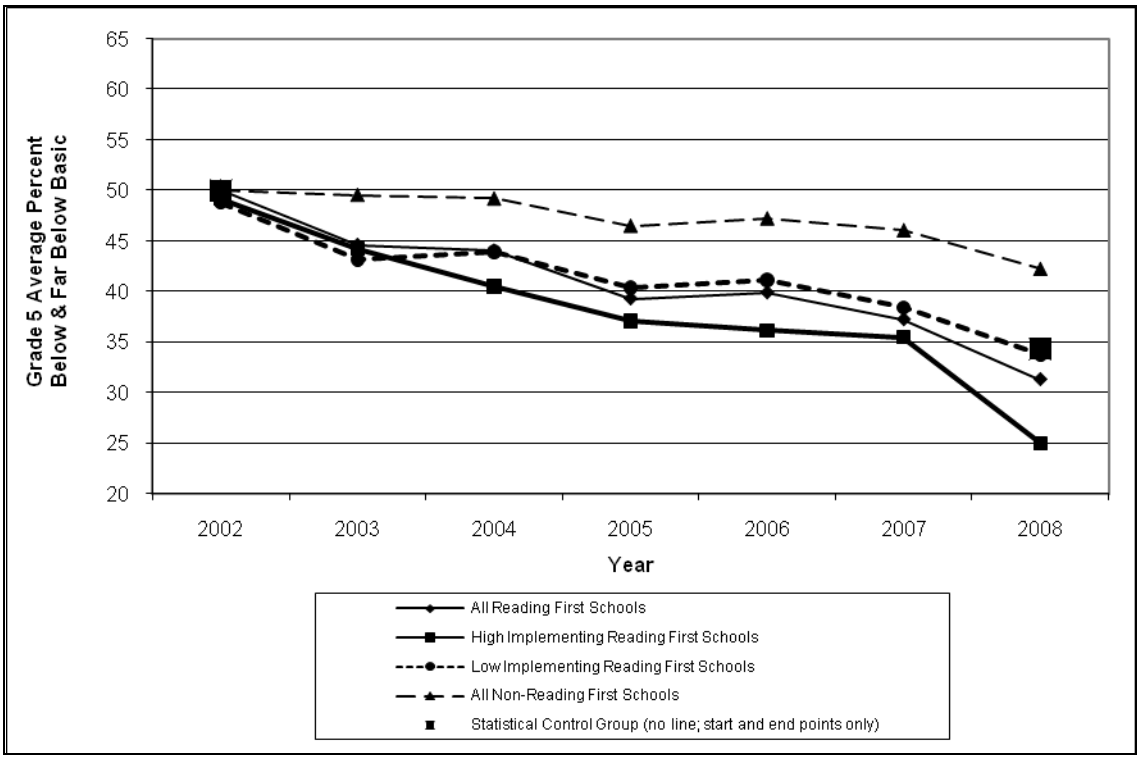
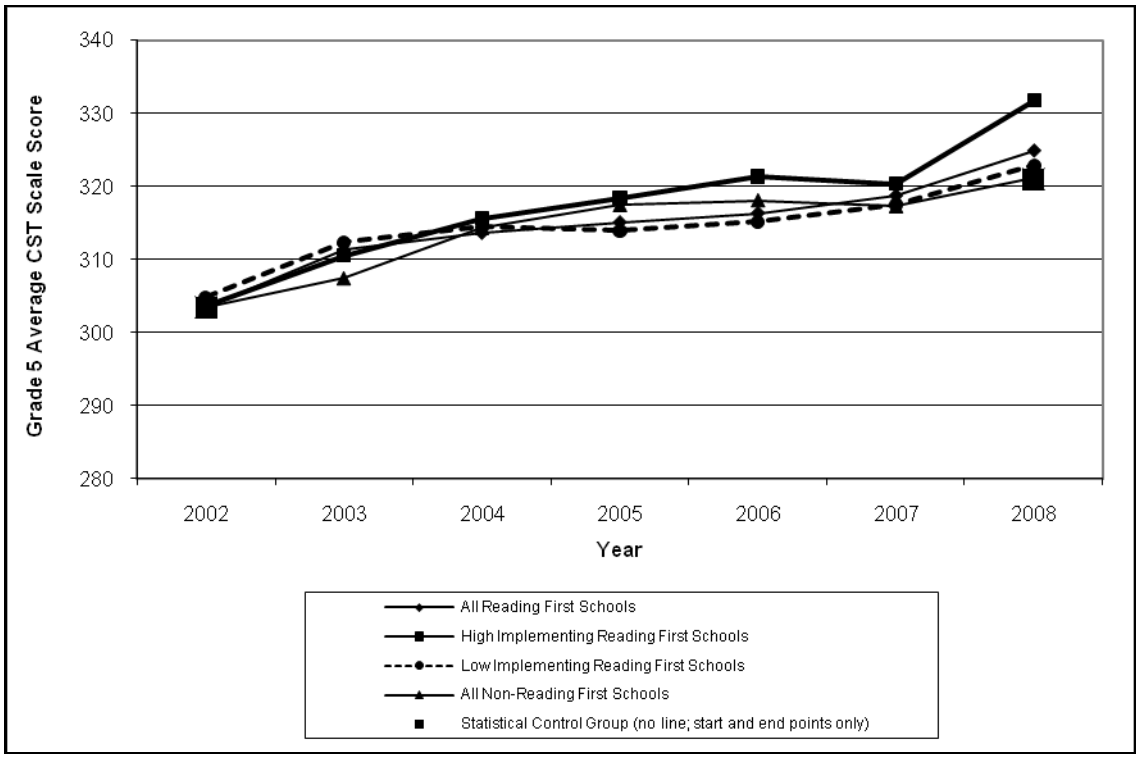


Figure 2.6c: CST Mean Scale Score, YIP = 6, Grade = 5



We continue to see in grades 4 and 5 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, coincident with the arrival of students taught in Reading First classrooms. While the grade 4 high implementing trend-lines tend to be somewhat jagged, that is likely 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, perhaps, that there is some threshold of implementation above which schools experience a qualitatively higher level of achievement and sustainability. Or, as discussed below, it could be a function of a uniquely responsive cohort of students.

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

Student Cohort Trend-Line (Table 2.7 and Figure 2.7)

The charts and trend-lines presented so far only describe trends for a given grade and cohort of schools as they have changed over the years. They do not provide a true longitudinal view tracking the progress of a given cohort of students through the years. Such an analysis requires: a) an ability to track individual students over time; b) an equated CST-based vertical scale that can measure cross-grade student growth. Neither requirement is currently available. Nonetheless, it may be of interest to get at least a top-level view of how a given cohort of students has scored since their first exposure to the program in grade 2. Table and Figure 2.7 report ongoing CST scores for the student cohort that was in Grade 2 in 2005. They are drawn from the YIP 6 schools and from all non-Reading First schools.

Table 2.7: Cross-Grade CST Scale Scores for Students that Were in Grade 2 in 2005

Student Cohort that was in Grade 2 in 2005	Reading First Schools (YIP 6 Only)			All Non-Reading First Elementary Schools
	All Reading First Schools	High Implementation Schools (Avg. RFII > 41.4)	Low Implementation Schools (Avg. RFII < 36.0)	
CST Mean Scale Score Per Student				
Grade 2 in 2005	315.1	319.0	312.8	343.9
Grade 3 in 2006	305.8	313.7	302.8	338.0
Grade 4 in 2007	327.3	331.3	325.8	359.5
Grade 5 in 2008	324.9	331.7	322.9	352.3

Figure 2.7: Cross-Grade CST Scores for Students that were in Grade 2 in 2005

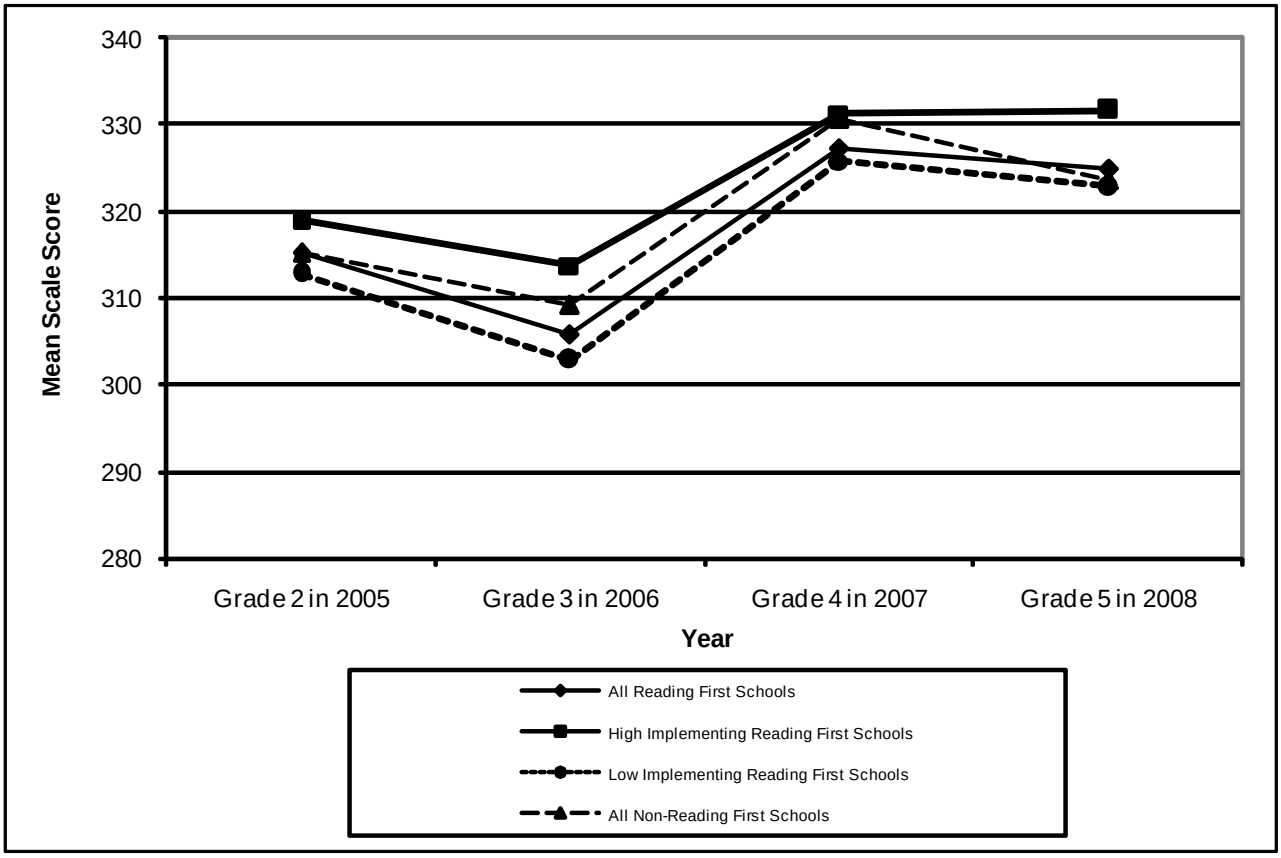


Figure 2.7 should not be interpreted as a growth chart. Because the CSTs are not equated across grades, they do not measure cross-grade growth. They reflect rather the relative difficulty of the exams. The grade 3 CST is the most difficult. The grade 4 and grade 5 CSTs are easier. Bear in mind that the "All Non-Reading First Schools" trend-line has been adjusted downward to have the same starting point as the Reading First schools.

Despite its limitations, Figure 2.7 does show that students in high implementing schools enjoy a considerable advantage over their peers in low implementing schools, as much as 10 scale score points in grade 3. In addition, the difference is fairly consistent for each grade.

The Reading First Effect: Regressions for the RFAI and Grade 5

The tables and trend-line charts presented so far in this chapter – in particular the comparisons between All Reading First Schools and the Statistical Control Group – are enough to establish that Reading First is effective for YIP 6 schools across 16 achievement metrics. Comparisons with the Statistical Control Group remove any confounding effects caused by variation in school demographic characteristics. However, these tables do not address YIP 5 or YIP 4 schools (which are addressed in Appendix E). Nor do they yield one overall statistic that can be identified as "the Reading First effect." The next two sections of Chapter 2 address this deficiency by explicitly answering the question, "What is the overall effect of Reading First across all YIPs, all achievement metrics, and all implementation years after controlling for school-level variation in demographic characteristics?"

The tool used to answer that question is the same as that used to calculate gains for the Statistical Control Group – multiple regression. For readers unfamiliar with multiple regression, the next three sections may feel overly technical and can safely be skipped; the essential story is already told in the preceding tables and trend-lines. However, readers interested in what the Reading First effect looks like in the *abstract*, disentangled from the effects caused by demographics, YIPs, achievement metrics, and so on, will find benefit in this discussion once the technical terminology has been absorbed.

What does it mean to "control for demographics"?

In tables such as 2.1 and 2.2 in this chapter, we report gain scores associated with high and low implementing schools and generally find that high implementing schools have higher gain scores. However, it is possible that high implementing schools show higher gains for reasons that have little to do with implementation of their reading program. For instance, if all high implementing schools happened also to be schools with mostly well-educated parents, one could argue that it was parent education causing higher gains, not Reading First.[16]

To address this issue, each year we perform a series of multiple regression analyses to control for non-Reading First factors that might impact achievement. The fruit of these analyses is the statistical control group statistic reported in Tables 2.1 to 2.6 – the expected gain for a school that is posited to be the same as the Reading First schools in a given YIP except that it has a minimal Reading First implementation statistic of 25. Each regression analysis yields a regression equation that relates Reading First implementation and school demographic variables to some specified achievement outcome. Without going into detail, we simply import averages of Reading First demographic variables, plus an RFII statistic of 25 (defined to be minimal implementation), into each regression equation to come up with a

[16] Indeed, Parent Education is a significant predictor of achievement gains as noted below, but the Reading First effect is independent of the Parent Education effect.

prediction for how a typical Reading First school *would* perform given minimal Reading First implementation on that achievement metric. The prediction is the basis for the statistical control group statistics.

But instead of focusing on the statistical control group statistics, let us focus on the regression analyses that lie behind them.

Tables 2.8 and 2.9 present results of two regression analyses (there are many more). The first has the 2008 RFAI score as an outcome variable. The second has the 2008 grade 5 CST percent Proficient and Above score. 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 grade 5 percent Proficient and Above as an outcome variable because like the grade 4 achievement scores it measures the degree to which participation in the K-3 Reading First program influences performance in higher grades.

Regression analysis involves identifying a number of potential "predictor" variables that contribute information regarding a "dependent" or outcome variable, but that are as independent as possible of the other predictor variables. In previous years the predictor variables have been: Average Yearly RFII, Years in Program, Percent of Socio-Economically Disadvantaged, Percent of English-language Learners, and a composite variable called "Yearly RFII*Years in Program." In Year 6, we have added four new predictor variables: Percent of Black Students, Percent of Migrant Students, Number of Students in School, and Student/Teacher Ratio. Parent Education is also a significant predictor of achievement but was dropped due to the number of schools lacking data on this variable. The inclusion of new predictor variables had only a slight effect on the size and significance of the "Average Yearly RFII" effect size, indicating that the Reading First effect is robust and independent of variation in these other demographic variables.

A large number of other variables were considered as predictors, mainly relating to ethnicity, but their effect sizes were not consistently significant.

One other note: "Average Yearly RFII" means the average Reading First Implementation Index (RFII) statistic over the previous four years for a given school, four being the maximum number of years over which a grade 3 student can have received Reading First-based instruction.

Table 2.8: Effect Size of Variables Predicting Percent of Students Proficient & Above on Grade 5 CSTs in 2008 ($R^2 = 0.31$)[1]

Predictor Variable (Predicting 2008 Grade 5 % 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. 5 % Proficient & Above	0.39	10.1	0.0000
Number of Years in Program	0.20	5.3	0.0000
Average Yearly RFII[5]	**0.09**	**2.8**	**0.0049**
Yearly RFII * Years in Program[6]	**0.22**	**6.1**	**0.0000**
Percent of SEDs in School	-0.13	-5.1	0.0000
Percent of ELs in School	-0.09	-2.2	0.0291
Percent of Migrant Students in School	-0.13	-4.0	0.0001
Percent of Black Students in School	-0.25	-6.7	0.0000
Number of Students in School	-0.12	-3.4	0.0007
Student/Teacher Ratio	0.06[7]	1.7	0.0981

Table 2.9: Effect Size of Variables Predicting the 2008 RFAI ($R^2 = 0.49$)[1]

Predictor Variable (Predicting the 2008 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	0.55	18.7	0.0000
Number of Years in Program	0.18	6.7	0.0000
Average Yearly RFII[5]	**0.11**	**4.3**	**0.0000**
Yearly RFII * Years in Program[6]	**0.20**	**7.8**	**0.0000**
Percent of SEDs in School	0.02	0.6	0.5545
Percent of ELs in School	-0.08	-2.3	0.0237
Percent of Migrant Students in School	-0.19	-6.7	0.0000
Percent of Black Students in School	-0.26	-8.7	0.0000
Number of Students in School	-0.11	-3.8	0.0001
Student/Teacher Ratio	0.11[7]	3.7	0.0002

[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 dependent variable (e.g., CST "% Proficient & Above") increases for every one standard deviation increase of a given 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 "Average Yearly RFII" is the average RFII statistic over the previous four years, 2005-08.
[6]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." To avoid colinearity, its effect size is computed in a separate regression run in which "Average Yearly RFII" and "Years in Program" are removed.
[7]Some may be surprised that the Student/Teacher Ratio effect is positive – the more students per classroom, the higher the achievement. But this predictor variable is notoriously non-linear, hence sample-dependent. In other regression equations it may come out negative.

How to interpret the "standardized beta coefficient" effect size

The predictor variables we are interested in are Number of Years in Program and Average Yearly RFII[17] and the variable that is obtained by multiplying them together -- Yearly RFII*Years in Program. They are in bold. The latter can be thought of as a school's total amount of Reading First implementation over time as it would impact a set of students moving up through the grades to grade 3. For the moment, we focus on the Average Yearly RFII effect. The remaining predictor variables improve the overall accuracy of the model and in so doing make the Average Yearly RFII effect more accurate as well.

The column of greatest interest is the "Standardized Beta Coefficient Effect." It tells us the positive or negative effect of the variable in question in standard deviation units. A standard deviation unit is the root mean squared difference of each value from the average for a given variable. It is analogous to one quartile up or down from the median of a set of scores. The standardized beta coefficient is defined as the number of standard deviations that the outcome variable changes given a one standard deviation increase in the predictor variable. So, looking at Table 2.9, we see that for every one standard deviation increase in the Average Yearly RFII, there is a 0.11 standard deviation increase in the 2008 RFAI achievement score. For every one standard deviation increase in the Percent of Migrant Students, there is a -0.19 standard deviation *decrease* in the 2008 RFAI achievement score. It is not hard to convert these standard deviation increases and decreases into real numbers (as is done for the statistical control group), but the point of using an abstract standardized beta coefficient metric is to facilitate comparisons across different predictor variables and achievement metrics. It puts them on the same scale.

Note that the standardized beta effect (for continuous variables) is *not* the same as other effect size metrics such as "Cohen's *d*." There is no accepted definition of a "large" standardized beta. Rather, one looks at: a) whether the effect is statistically significant, i.e., would probably not have occurred by chance; and b) whether the effect is "meaningful." The statistical significance is given in the two last columns. If a t-test yields a value greater than 1.96 or less than -1.96, the probability that the effect could have occurred by chance is less than 0.05, the widely used criterion for deciding whether or not an effect is "statistically significant." In predicting the 2008 RFAI, we see that the t-test for the "Average Yearly RFII" yields a value of 4.3, and the probability the effect would have occurred by chance is smaller than 0.0000. So it is definitely statistically significant.

However, it is quite possible for an effect to be statistically significant without being "meaningful". That is, the effect probably would not have occurred by chance but is too small to bother with. There is no

[17] Average Yearly RFII is the average RFII over the four previous years, e.g., 2005-2008. We have opted not to include a fifth year in the average since the cumulative effect of Reading First implementation at the student level cannot exceed the four years ranging from K-3. The fifth year adds little at the student level.

widely accepted test for deciding whether an effect is meaningful, but there are several ways to approach the question. First, we can translate the standardized betas into real gain scores and see if these gain scores are large enough to be "meaningful" to education professionals. This has already been done with the statistical control group statistics. When the grade 5 Average Yearly RFII standardized beta of 0.09 standard deviations is converted into real numbers, we get the values shown in Table 2.6, for instance. Over 6 years, a school defined to be non-implementing with an Average Yearly RFII of 25 will grow 17.6 scale score points in grade 5, whereas a typical Reading First school will grow 21.5 scale score points, and a high implementing school will grow 28.0 points over the same period. Are these differences educationally meaningful?

Table 2.11 in the next section assesses "meaningfulness" by translating a standardized beta effect size averaged across all possible regression analyses into predicted 6-year gain scores for each achievement metric. It shows that increasing the level of Reading First implementation (RFII) at a school by 25 points effectively doubles achievement gains. Are these "meaningful" effects? We certainly think so, but it is a subjective, contextual determination.

Another way to assess "meaningfulness" is to compare it to other standardized beta effects that we have reason to think are meaningful. Social scientists and educational researchers are accustomed to finding that demographic variables are frequently the strongest predictors of educational outcomes, and thus are "meaningful." Tables 2.8 and 2.9 provide four demographic variables and two institutional variables – Percent of SED Students, Percent of EL Students, Percent of Migrant Students, Percent of Black Students, Total Number of Students, and Student/Teacher Ratio. The demographic variables have negative standardized beta effects (meaning that as they increase, the achievement decreases) ranging from -0.08 to -0.26. Average Yearly RFII effect is a positive 0.09 (Table 2.8) and 0.11 (Table 2.9). So the RFII effect is at least in the same range, more or less, as the demographic variables, but in a positive direction. In fact, it is similar in size to the negative effect of Percent of EL Students. Therefore, the deficit in achievement experienced by increasing the Percent of EL Students in a school can be counterbalanced by increasing the level of Reading First implementation by the same number of standard deviations. In this sense, Average Yearly RFII is just as "meaningful" as Percent of EL Students in predicting achievement.

One reason why it does not make sense to attach artificial labels of "large" or "small" on standardized beta effect sizes is that they are greatly influenced by measurement error, as has been discussed in previous reports. The larger the measurement error of either the achievement outcome variable or the Reading First implementation predictor variable, or any other predictor variable, the smaller the effect size will be. Therefore, the true effect sizes for Average Yearly RFII and Yearly RFII * Years in Program

are certain to be higher than reported here. Unfortunately, there is no easy or obvious way to compensate for the "shrinkage" effects of measurement error.

The non-RFII predictor variables

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 achievement level. Percent of SEDs is not a variable of primary interest, but its role is to remove confounding influences that socio-economic status might have on the implementation effect. The same is true of Percent of EL Students. The new predictor variables – Percent of Black Students, Percent of Migrant Students, Number of Students, Student/Teacher Ratio -- control for other factors that might otherwise distort the RFII effect. It is hypothesized that Percent of Black Students is a significant negative predictor because it flags inner city schools which may have high drop-out rates and low attendance, which would lower achievement. Similarly, Percent of Migrant Students may flag rural schools with high percentages of students that "drop out" because of movement to other schools and districts. It is reasonable that any factor that affects student attendance will impact the effectiveness of programs at that school.

How to interpret the "Average Yearly RFII * Years in Program" effect

We must consider the thorny question of how to measure the total amount of implementation that has occurred in a school since its entry into the program, and to relate that to total achievement gains. A school that has an Average Yearly RFII of 39 over *six* years would appear to have three times as much implementation as a school that has an Average Yearly RFII of 39 over *two* years. If so, "Years in Program" should be folded into the total implementation measure, i.e., the two variables should be multiplied together to create a composite predictor variable.

Table 2.8 and Table 2.9 show that this substantially increases the apparent size of the Reading First effect relative to Average Yearly RFII alone, to the 0.20 range as has been reported in previous Reports. This is due to the fact that Reading First schools have been trending upward since 2002, making Years in Program a strong positive predictor.

The problem is that we cannot discount the possibility that Reading First schools have been trending upward for reasons that are unrelated to Reading First, a possibility that seems more likely given that non-Reading First schools have also been trending upward over the same period. If so, then including Years in Program artificially inflates the Reading First effect. But if the Reading First effect is real – and it is – then we would certainly expect that it contributed *something* to the upward trend.

We are left with an awkward dilemma: we either understate the Reading First effect by not combining it with Years in Program, or overstate it by combining it with Years in Program. For reasons of simplicity

and conservatism, we have chosen to *understate* it. All of the summary Reading First standardized beta effect sizes cited in this and previous Reports are based on Average Yearly RFII only. The Years in Program effect is not treated as a Reading First effect.

Nonetheless, it is important to explore the implications of treating Years in Program as a Reading First effect. Doing so requires us to assume that the positive impact of Years in Program is due primarily to the school's implementation of Reading First and that it is independent of any 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 the important question raised in previous Reports: If the achievement gains experienced by Reading First schools over a six-year period are primarily a Reading First effect, as we argue, 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 and State Board of Education directives, as well as responses to the March 2008 Supplemental Survey Report regarding use of Reading First program elements in eligible non-Reading First schools, firmly 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 430 principal professional development, and the 6-8 week skills assessments. We found in *The Reading First Supplemental Survey Report, March 2008* that a sample of 50-60% of eligible non-Reading First schools use the same Houghton-Mifflin reading program that is one

of the two required by Reading First.[18] 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 differences are that the non-Reading First schools must use other funding sources to hire reading coaches and provide professional development, and that Reading First schools tend to implement the 80 hours follow-up practicum in SB 472 training within the first year. Also, Reading First LEAs agree, as a condition of eligibility, to a set of "Assurances" about how they will implement the program.

Therefore, we believe that the statewide trend is fundamentally an effect of the same educational practices and program elements that are more stringently required and better supported by the Reading First (California) program and accompanying funding.

The Overall Reading First Effect: Generalizations Based on Meta-Analysis

Tables 2.8 and 2.9 provide estimates of the Reading First implementation effect for two achievement outcome variables – grade 5 percent Proficient and Above for 2008 and the RFAI for 2008. They provide examples of two "regressions".[19] Although these regression-based (standardized beta) effect sizes are more general than the individual statistical control group comparisons reported earlier in the chapter, since they combine all the YIPs in one analysis, we still do not have a single statistic in answer to the question, "What is the overall Reading First effect?"

A Brief Technical Digression into our Meta-Analysis Methodology (which may be skipped)

The answer to the question, "What is the overall Reading First effect?" may be approached by averaging the standardized beta coefficients (effect sizes) for Average Yearly RFII across the 16 achievement metrics. The 16 achievement metrics consist of the three CST achievement metrics (Proficient and Above, Below and Far Below Basic, Scale Score), across four grades (grades 2, 3, 4, and 5), plus the three CAT/6 grade 3 achievement metrics, plus the RFAI. The average effect size is our best estimate of the "true" effect size. This type of analysis, where we average the effects across a variety of studies and regression equations, is an example of "Meta-analysis."[20] Again, note in this context that a standardized beta effect size is *not* comparable to a Cohen's *d* effect size, another type of effect size used widely in meta-analysis.

[18] See *The Reading First Supplemental Survey Report, March 2008* at *www.eddata.com/resources/publications/*.
[19] For purposes of discussion in the meta-analysis section, the term "regression" refers to the process of computing effect sizes for several predictor variables used to predict one achievement outcome variable. Thus, there are as many "regressions" as there are possible outcome variables. Tables 2.8 and 2.8 give results of two such "regressions."
[20] Cooper & Hedges, Eds. (1994) *The Handbook of Research Synthesis.*

In this approach, each regression equation looks at school starting point (a school's achievement score when it entered the program) as well as the cumulative implementation over subsequent years, and it combines all the YIPs together for each analysis. This method of calculating Reading First effect sizes may be called the "Cumulative Gain/RFII Method." Tables 2.8 and 2.9 are outputs of this method. It is a desirable method for calculating the "overall Reading First effect" because, among other things, it uses the maximum number of schools (863) in each regression equation, which leads to greater precision and stability in calculating the RFII effect.

Unfortunately, all 16 of the regressions used in the "Cumulative Gain/RFII Method" predict achievement outcomes for just one year – 2008. The procedure leaves out the regressions used to predict achievement for 2007, 2006, 2005, 2004 and 2003, which can lead to misleading findings in some cases (e.g., the small grade 4 Reading First effect in 2008). Also, the "Cumulative" method is awkward to implement for exploratory analyses where we want to examine Reading First effects for particular YIPs, grades, student cohorts, and years separately, to disentangle their effects.

That led us to construct a slightly different type of regression equation based on what might be called the "Yearly Gain/RFII Method". The "Yearly" method focuses only on the gain that occurs within one year and the RFII calculated for just that year, and it performs separate analyses for each YIP. When broken out this way, there turn out to be 221 independent regression equations that can be calculated. However, because each regression equation is based on a smaller sample of schools (only those of the same YIP) the error is greater and the standardized beta effect size is smaller. (The total statistical significance across all 221 regression equations is not diminished, however, and is in fact somewhat greater than the "Cumulative" method.) Thus, the two species of regression models each have their deficiencies. Our "Cumulative" approach leaves out regressions that predict years prior to 2008. Our "Yearly" approach underestimates the "true" effect size.

That is why, for purposes of this report, we amalgamate the two regression methodologies. We do the bulk of analysis using the "Yearly Gain/RFII Method" but convert its standardized beta effect sizes into a "Cumulative Gain/RFII" metric, which inflates them to approximately what we would obtain if each regression equation included all 863 schools. We call this the "Adjusted Yearly Gain/RFII Method". (All significance tests – and other statistics where appropriate -- are performed with the *unadjusted* values, however.)

The Overall Reading First Effect

We have asked the question, "What is the overall Reading First effect?" The answer, according to Table 2.10, is 0.082.

Table 2.10: Average Effect Size

	"Adjusted Yearly Gain/RFII Method"
Count of Effect Sizes from Different Regressions	221
Average RFII Effect Size, Adjusted to "Cumulative" metric	**0.082**
Standard Deviation	0.024
Minimum	0.022
Maximum	0.148
Standard Error	0.004
t-test (t > 1.96 implies significance with 95% confidence)	16.5
Probability the Effect is by Chance	0.0000
Mean Absolute Demographic Effect Size	0.130
RFII Effect/Mean Abs. Demographic Effect Ratio	0.634

Table 2.10 reveals that across 221 separate regressions the RFII effect on achievement ranges from a minimum standardized beta coefficient of 0.022 to a maximum of 0.148, with a mean standardized beta of 0.082. We now ask two questions regarding this 0.082 standardized beta effect size:

- Is 0.082 significantly greater than zero at the 95% confidence level?
- Is it meaningful?

The statistical significance is found by calculating the standard error around the 0.082 effect size statistic. Because we are averaging many effect sizes (221), the standard error is small – only 0.004. Thus, the Adjusted Average RFII Effect is 0.082 plus or minus 0.004 (to achieve 67% confidence), or plus or minus 0.008 (to achieve 95% confidence). This makes it clear that the Reading First effect is very significant indeed – approximately 16 standard errors above zero (as shown by the t-test), where 1.96 standard errors would be enough to establish "significance" at the 95% confidence level. This effect ensures that while individual studies may conceivably find a negative or non-existent Reading First effect, they are outliers and not representative of the whole.

Note, however, that a "statistically significant" difference from zero by no means guarantees a "meaningful" or "practical" difference from zero, just a difference that is unlikely to have occurred by chance. To assess whether 0.082 is "meaningful" we ask two questions, as discussed above.

- How big is the Reading First effect relative to demographic effects that are widely considered to be "meaningful"?

- What does the 0.082 standardized beta effect look like when we convert it from standard deviation units into well-understood achievement units (e.g., scale scores or percent proficient)?

Table 2.10 tells us that the "Mean Absolute Demographic Effect Size" is 0.130.[21] The Reading First effect is therefore 0.634 the size of the mean demographic effect. Thus Reading First is around 63% as strong in a positive direction as the demographic effects are in a negative direction. On its own, this is evidence that if the demographic effects are "meaningful" so is the Reading First effect.

But what does the 63% mean in real terms? Here is one example. The decrease in grade 2 scale scores that would occur if the percentage of migrant students in a school increased by 10 percentage points would be almost exactly counterbalanced by increasing the school's RFII, its level of Reading First implementation, by 11 points. Thus, in this case, the negative effects of a percentage point increase in a demographic variable like Percent of Migrant Students are almost canceled by a one point increase in the RFII.

Table 2.11 addresses the "meaningfulness" question in another way -- by listing the 6-year gain scores that can be expected for a Reading First school implementing at different levels (RFIIs) across all the achievement metrics. An RFII of 25 is around the minimum that a Reading First school implements, not very different from what is done in many non-Reading First schools. An RFII above 41 is considered "high implementation" based on a convention established in previous reports. An RFII of 39 is typical in 2008.

[21] "The demographic variables are: Percent SED, Percent EL, Percent Migrant, and Percent Black. Number of students in the school is also included, though it is an institutional variable. Student/Teacher Ratio is not included.

Table 2.11: Predicted 6-Year Gain Across Achievement Metrics for Different Levels of RF Implementation

Based on data from all RF schools and all report years, since 2003	Predicted 6-Year Gain						Ratio (RFII = 50) / (RFII = 25)
	RFII = 25	RFII = 30	RFII = 35	RFII = 40	RFII = 45	RFII = 50	
Grade 2, CSTs							
% Proficient and Above	5.4	6.3	7.3	8.3	9.3	10.2	1.91
% Below and Far Below Basic	-4.4	-5.8	-7.2	-8.5	-9.9	-11.3	2.53
Mean Scale Score	6.0	7.5	9.1	10.6	12.2	13.7	2.29
Grade 3, CSTs							
% Proficient and Above	2.4	2.8	3.2	3.6	3.9	4.3	1.78
% Below and Far Below Basic	-5.2	-6.0	-6.7	-7.5	-8.2	-9.0	1.72
Mean Scale Score	4.8	5.6	6.4	7.2	8.0	8.8	1.85
Grade 3, CAT6, Mean Percentile Rank							
Reading, Mean PR metric	1.2	1.7	2.1	2.6	3.1	3.6	3.06
Language, Mean PR metric	2.1	2.5	2.9	3.2	3.6	4.0	1.94
Spelling, Mean PR metric	3.4	4.2	5.0	5.7	6.5	7.2	2.11
Grade 4, CSTs							
% Proficient and Above	7.2	7.8	8.4	9.0	9.5	10.1	1.41
% Below and Far Below Basic	-6.6	-7.1	-7.7	-8.3	-8.9	-9.4	1.44
Mean Scale Score	7.1	7.9	8.7	9.5	10.3	11.2	1.57
Grade 5, CSTs							
% Proficient and Above	5.4	6.1	6.7	7.3	8.0	8.6	1.60
% Below and Far Below Basic	-5.2	-6.2	-7.2	-8.2	-9.1	-10.1	1.93
Mean Scale Score	5.4	6.3	7.2	8.1	8.9	9.8	1.83
Reading First Achievement Index							
RFAI metric	4.0	4.5	5.1	5.7	6.3	6.9	1.74
						Average Ratio	**1.92**

We see that a Reading First school with an RFII of 25 is expected to grow 6.0 scale score points on the grade 2 CSTs over six years. That same school would grow 13.7 scale score points over the same period at an RFII implementation level of 50 – more than twice as much. Similar patterns exist across the achievement metrics and across grades 2-5. A school's grade 3 CAT/6 Reading mean percentile rank is expected to *triple* if the school moves from an RFII of 25 to 50.

The column to the right, "Ratio (RFII = 50) / (RFII = 25)", compares the growth that can be expected with an RFII of 50 with the growth that can be expected with an RFII of 25. Averaging these ratios across the various achievement metrics, we are able to derive a rule of thumb:

"A school that increases its RFII by 25 points for a given time period can expect, approximately, to double its achievement gains on a variety of grade 2-5 metrics over the same time period – a 100% increase in achievement gains. A school that increases its RFII by just 5 points can expect a 10-20% increase in its achievement gains."

The Reading First Effect: Six Analyses for Specified Sub-Groups

We now explore the Reading First effects for individual groups of schools, including individual YIPs and individual outcome years. All Reading First effect sizes are computed using the "Adjusted Yearly Gain/RFII method" in order to yield the most accurate effect sizes. Note that the charts below do not track achievement; they track *change* in achievement specifically attributable to Reading First.

Analysis #1: Which YIP has the highest RFII effect?

Figure and Table 2.12 break out schools by YIP, i.e., the year they entered the Reading First program, and present their average effect sizes.

Figure 2.12: Average RFII Effect by YIP

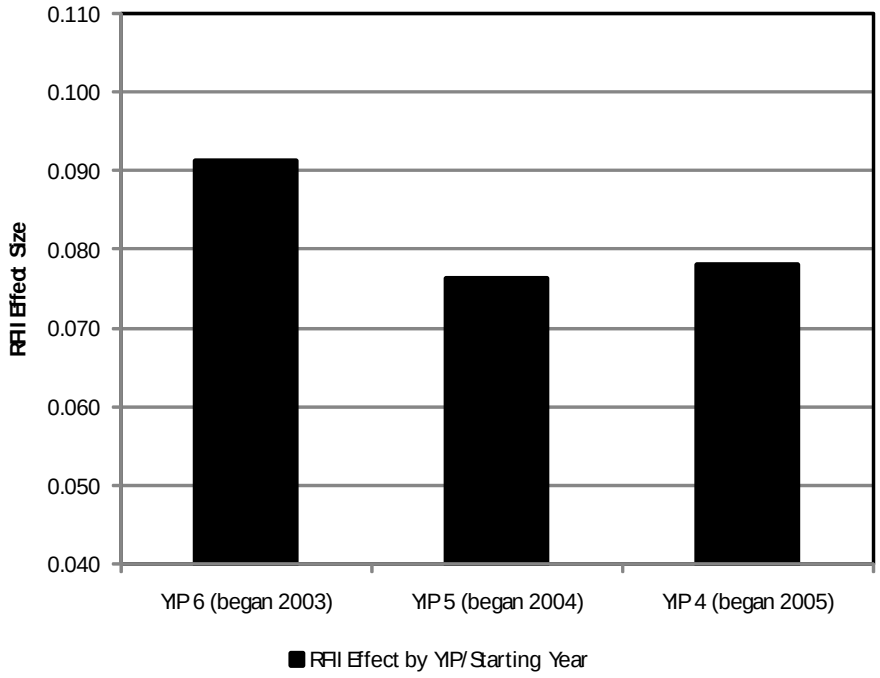


Table 2.12: Average RFII Effect by YIP

	Average RFII Effect (SD Units) by YIP
YIP 6 Schools (began in 2003)	0.091
YIP 5 Schools (began in 2004)	0.076
YIP 4 Schools (began in 2005)	0.078

We see that Reading First has been most effective with the YIP 6 schools (entering the program in 2003) and that the YIP 4 and 5 schools were less effective. This highlights the finding in previous evaluation reports that the LEAs from different funding cohorts face very different demographic challenges. For one thing, we know from Chapter 1 that the YIP 6 schools are predominantly urban and include schools within the Los Angeles Unified School District (LAUSD) whereas the YIP 5 schools are suburban and the YIP 4 schools predominantly rural, with nearly twice as many migrant students. Urban schools may have an advantage in program implementation because they are nearer to their R-TACs, nearer to their students, have large and well-staffed central offices, and have fewer migrant students.

Analysis #2: Which grade has the highest RFII effect? Addressing a question raised earlier, Is there a Reading First effect in grade 4?

Figure and Table 2.13 break out the RFII effect by grade level. Grades K and 1 are not included because they do not have an appropriate achievement metric.

Figure 2.13: Average RFII Effect by Grade

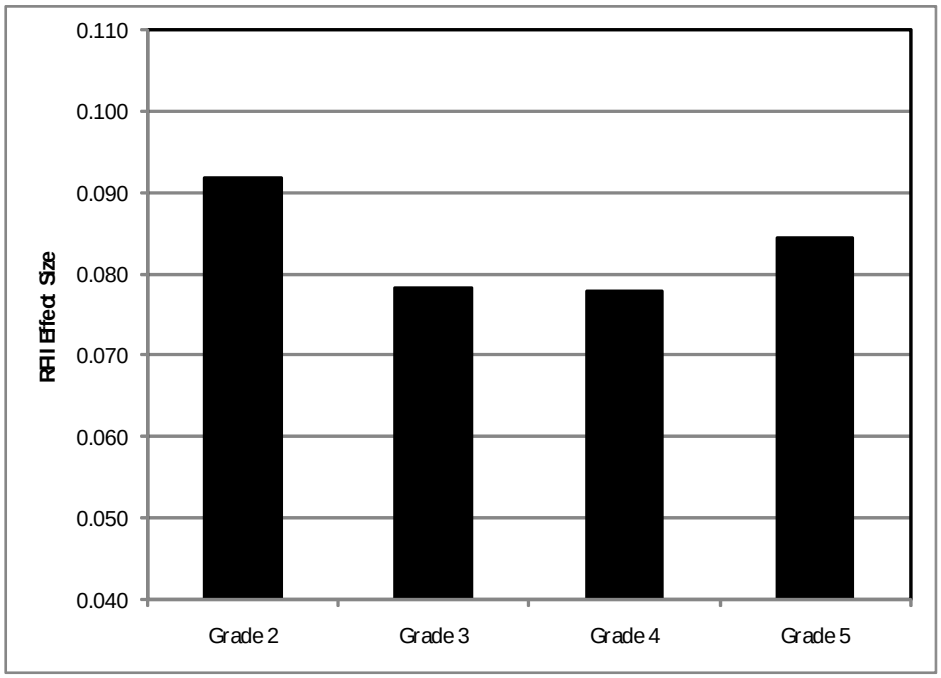


Table 2.13: Average RFII Effect by Grade

	Average RFII Effect (SD Units) by Grade
Grade 2	0.092
Grade 3	0.078
Grade 4	0.078
Grade 5	0.085

Grade 2 and grade 5 register higher Reading First effects than grades 3 and 4. It has been suggested that this may be because the Open Court and Houghton-Mifflin reading programs are more targeted on the types of skills that are tested in the grade 2 CSTs than on those tested in grades 3 and 4 (such as higher-level reading comprehension).

It is important to note that, contrary to what one might expect from Summary Gains Table 2.1, which shows no Reading First effect for grade 4, a complete meta-analysis using outcomes from previous years shows that the grade 4 Reading First effect is on a par with grade 3. It is only when the outcome variable consists of 2008 grade 4 scores that the Reading First effect appears to vanish. A possible explanation is suggested below.

Analysis #3: Which type of achievement metric shows the highest RFII effect?

Figure and Table 2.14 displays average RFII effects for seven types of achievement metrics.

Figure 2.14: Average RFII effect by Type of Achievement Metric

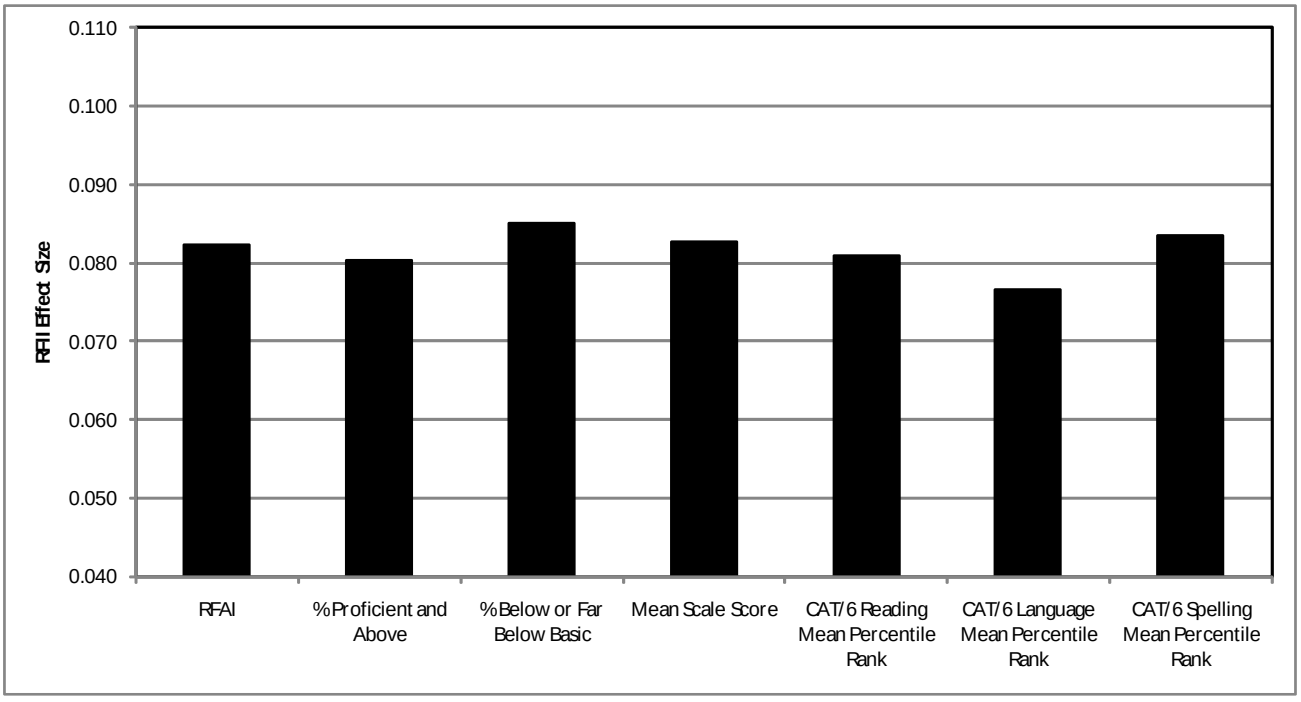


Table 2.14: Average RFII effect by Type of Achievement Metric

	Average RFII Effect (SD Units) by Type of Achievement Metric
RFAI	0.082
% Proficient and Above	0.080
% Below or Far Below Basic	0.085
Mean Scale Score	0.083
CAT/6 Reading Mean Percentile Rank	0.081
CAT/6 Language Mean Percentile Rank	0.077
CAT/6 Spelling Mean Percentile Rank	0.084

Table 2.14 confirms that Reading First is most effective with students in the "Percent Below or Far Below Basic" performance categories – a fulfillment of its NCLB charter. As a "reading" program, it is less effective with "language" skills, as shown in the CAT/6 Language metric, and it is very effective with more technical basic skills as shown by the CAT/6 Spelling metric.

Analysis #4: Which student cohort has responded most to Reading First implementation?

There is a tendency in evaluation studies to assume that student cohorts are interchangeable, that grade 2 students in 2003 are like grade 2 students in 2004, and that any change in performance is due to some external factor. Table 2.15 and the accompanying figure show that this is not a valid assumption. Four sequential student cohorts in our study have responded to the program in very different ways. Each trend-line tracks the progress of a student cohort up to 2008, starting from the year it was in grade 2.

Figure 2.15: RFII Effect by Student Cohort and Year

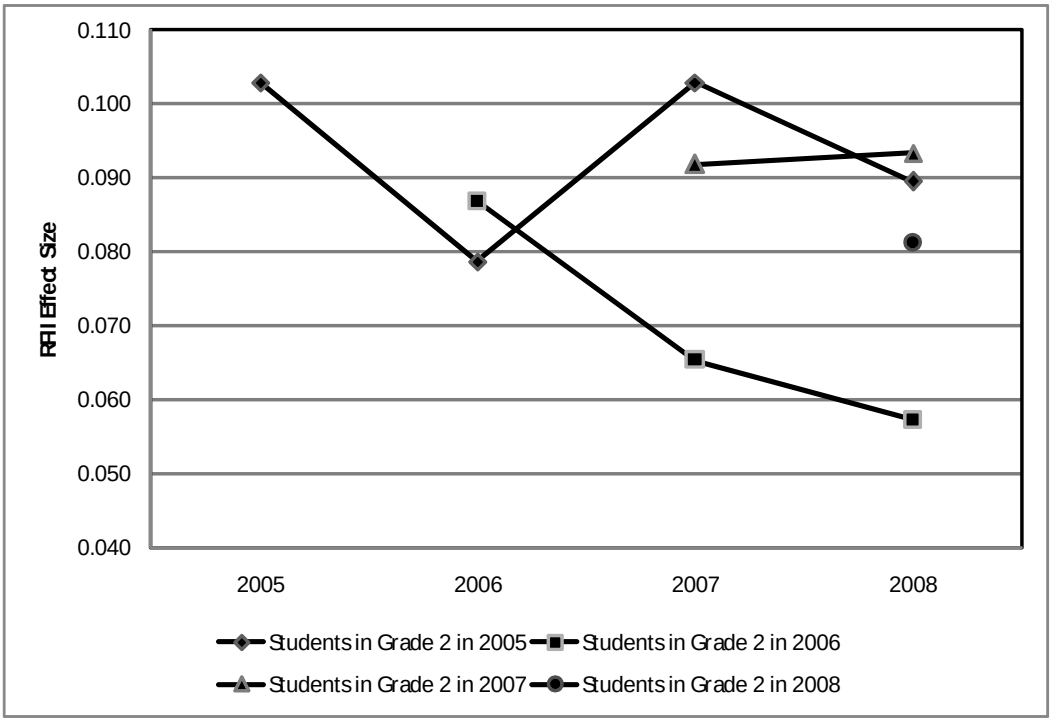


Table 2.15: Average RFII effect by Student Cohort and Year

	2005	2006	2007	2008
Students in Grade 2 in 2005	0.103	0.079	0.103	0.089
Students in Grade 2 in 2006		0.087	0.065	0.057
Students in Grade 2 in 2007			0.092	0.093
Students in Grade 2 in 2008				0.081

The first student cohort (grade 2 in 2005) has in general experienced a strong and sustained Reading First effect, except for a dip in 2006. The second cohort started off with a medium effect which dwindled markedly in 2007 and 2008. The third cohort had a medium-to-strong effect in both 2007 and 2008. The most recent cohort has begun with a medium-to-low RFII effect.

There are several important messages to take from this analysis:

- We cannot assume that successive student cohorts will be similar to each other in how they respond to a program. This finding is relevant to all evaluation studies that are not based on a student-level longitudinal design, such as the present study.
- The first student cohort (grade 2 in 2005) has been by far the most responsive to Reading First instruction. It would be helpful to know why. What is different about these students?
- The responsiveness of the second student cohort (grade 2 in 2006) not only is lower than the first, but it has declined steadily over time. Again, it would be helpful to know what has happened to cause these students to become less responsive to Reading First implementation.
- The Reading First effect does not necessarily decline with each successive cohort. While the first cohort has been most responsive, the next most responsive has been the third cohort.
- We now have a theory for what happened with grade 4. Grade 4's apparent lack of responsiveness to Reading First in Summary Gains Table 2.1 seems to be due to the fact that the most poorly responding student cohort (grade 2 in 2006) happened to be in grade 4 in 2008. That suggests, as was noted above, that the problem lies not with grade 4 *per se* but with the cohort of students that passed through grade 4 in 2008. If this is true, we would expect a similar dip in the Reading First effect size for grade 5 in 2009.

Analysis #5: Is there evidence of a "plateau effect"?

It has been stated in previous evaluation reports that it is not realistic to expect Reading First to cause student achievement to increase indefinitely. Program implementation in education tends to reach a "saturation point" beyond which other variables outside the program may be necessary to impact achievement, variables which are difficult to measure or predict. Predicting and analyzing a plateau effect allows us to gauge expectations of Reading First and temper unrealistic or idealistic goals. If a patient has a headache, the first and second aspirin are effective in reducing the pain. The third aspirin has less effect. The fourth, fifth, and sixth aspirin appear to have no effect at all. One would not thereby conclude that aspirin is ineffective for headaches. Nor should we conclude that an educational program is ineffective if it does not continue to produce additional gains beyond a certain level of implementation.

Table and Figure 2.16 look at the average RFII effect by year to assess whether Reading First is experiencing a plateau effect. If Reading First is approaching a point of "saturation" beyond which further growth is increasingly difficult given institutional and demographic constraints, we would expect the Reading First effect to decline steadily.

Figure 2.16: Average RFII Effect by Year

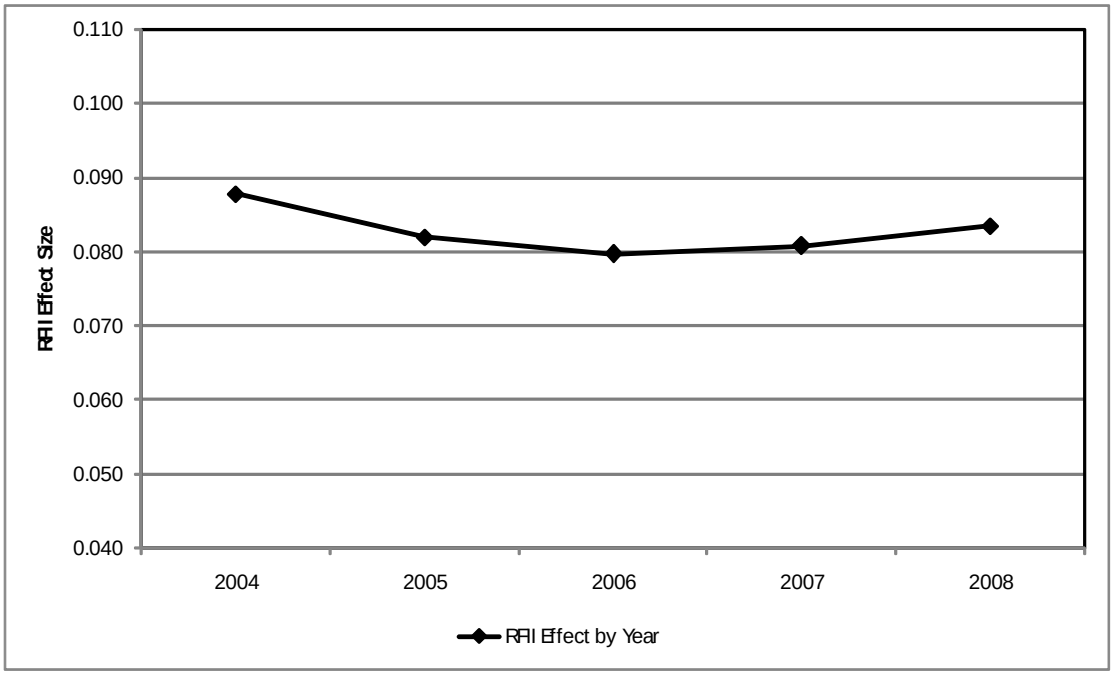


Table 2.16: Average RFII Effect by Year

	Average RFII Effect (SD Units) by Year
2004	0.088
2005	0.082
2006	0.080
2007	0.081
2008	0.083

Let us restate that these curves do not represent achievement. They represent *change* in achievement specifically attributable to Reading First. A declining Reading First effect curve in Figure 2.16 corresponds to a rising but flattening achievement curve, a "plateau", in trend-line charts such as those presented earlier in this chapter (Figures 2.4a-f provide good examples). This is especially true when accompanied by a narrowing of the difference between high and low implementing schools.

Figure 2.16 shows that there was a strong Reading First effect in 2004 when the program was in the second year of adoption for the first funding cohort and a new funding cohort had just experienced its first year. New programs often show their largest effects in the second year, when the program is still fresh but no longer unfamiliar. The effect declined in 2005, a little more in 2006, but then started increasing.

Does this curve match what we would expect of a classic "plateau" effect? No, for two reasons:

- Years 2007 and 2008 show an increase in the Reading First effect, not what one would expect from a plateau.
- A true plateau in achievement would correspond to a straight-line decrease in the RFII effect, eventually going to zero. Instead, we see that up to 2006 the rate of decrease gets smaller, that the RFII effect actually finds a "floor" around 0.080, which is still a respectable and significant effect size. This is not what one would expect of a true plateau effect.

To explore the question further, we break out the schools by YIP and Year.

Analysis #6: The plateau effect broken out by YIP

Figure and Table 2.17 shows how the Reading First effects for YIPs 3, 4, and 5 have evolved over time. It decomposes the curve in Figure 2.16 into its YIP components.

Figure 2.17: RFII Effect by YIP and Year, to Assess YIP-specific "Plateau Effects"

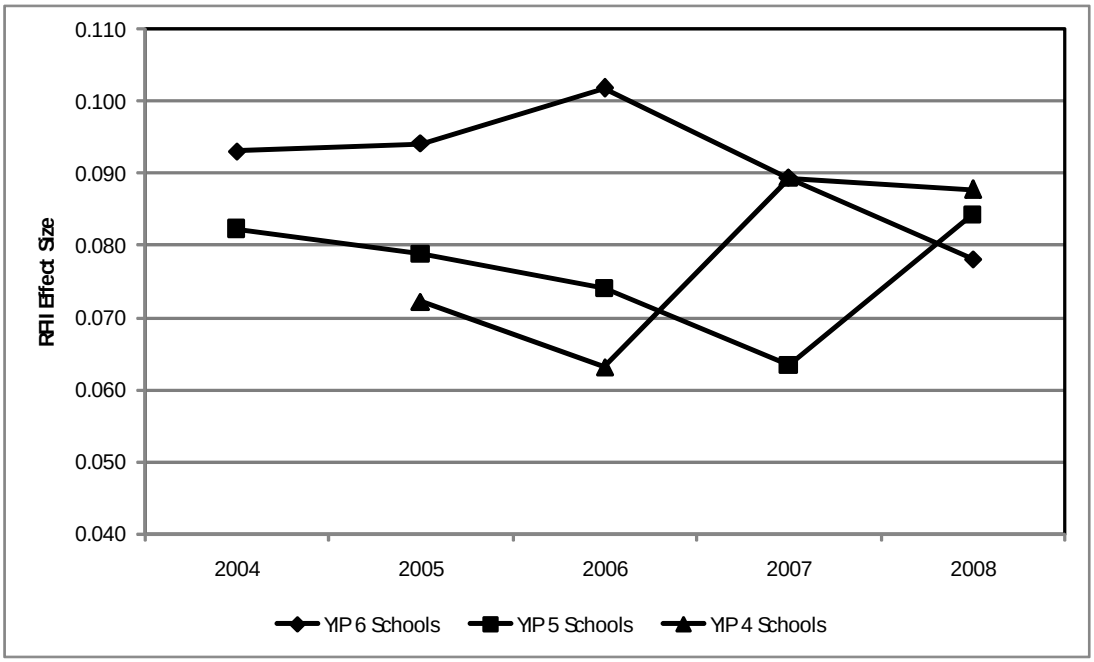


Table 2.17: Average RFII Effect by YIP and Year, to Assess YIP-Specific "Plateau Effects"

	2004	2005	2006	2007	2008
YIP 6 Schools	0.093	0.094	0.102	0.089	0.078
YIP 5 Schools	0.082	0.079	0.074	0.063	0.084
YIP 4 Schools		0.072	0.063	0.089	0.088

Here we see the cause of the "floor" in Table 2.16 and of the uptick in 2007 and 2008. Reading First in YIP 4 schools became more effective starting in 2007. YIP 5 schools, after a steady decline, became more effective in 2008. These two turn-arounds suffice to compensate for the declining effect that seems to be establishing itself in the YIP 6 schools.

YIP 6 schools do seem to have entered a period of declining Reading First effectiveness and a possible "plateau effect", strongly apparent in the grade 2 and grade 3 trend-line charts for YIP 6. While the three points in Figure 2.17 show the steady decline characteristic of a plateau effect, we have no way of knowing whether the trend will reverse itself as happened with the YIP 5 schools. Unfortunately, it will not be possible to track the YIP 6 schools further as their funding has been discontinued as of the 2008-09 school year.

We do not have an explanation for why YIP 6 might be reaching a plateau while the effectiveness of Reading First for YIPs 4 and 5 is increasing. We hypothesize that students in YIP 5 schools, which are largely suburban, were under-performing through 2007, but that as Reading First hit a saturation point across all the elementary grades they were able collectively to perform at a much higher level, and that there is room for quite a bit more achievement growth in these schools, probably more than in the predominantly urban YIP 6 schools. We suggest that the predominantly rural schools in YIP 4 had trouble getting started and that it took three years for the program to gain traction with the teacher and student population. But as of 2007, Reading First is having a strong, steady growth effect on YIP 4 schools, which also may have higher achievement potential than the urban YIP 6 schools.

So, is there a Reading First plateau effect? In general, no. For YIP 6 schools alone, quite possibly.

Conclusions

The conclusions in the Year 6 Report reinforce and extend those of the Year 5 Report. We began the chapter by stating that Reading First would be said to show evidence of being effective to the degree that:

1. Achievement gains in Reading First schools are positive for grades 2, 3, 4, and 5.
2. Reading First schools show higher achievement gains than non-Reading First schools for grades 2, 3, 4, and 5.

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

The Year 6 Report finds that Reading First is effective across all five criteria with a small number of exceptions in particular instances. As regards the fifth criterion, this report notes that the Reading First effect has been found to be extremely statistically significant when averaged across all possible achievement metrics over the years. Statistically the Reading First effect is real, no matter how it is calculated. We find the effect to be meaningful, as well. Reading First is more than 60% as powerful in impacting achievement as such well-established demographic variables as percent of SED, EL, black, and migrant students per school.

The Year 6 report extends findings of effectiveness to grades 4 and 5, even though these grades are not directly addressed by Reading First. This confirms the findings of the previous evaluation reports 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.

In addition, we find some evidence that the amount of *additional* achievement gain produced by a given amount of Reading First implementation is lessening for YIP 6 schools. This might be the beginning of the "plateau effect" predicted in previous evaluation reports. However, the effect does not appear in the YIP 5 and YIP 4 schools.

We conclude by restating from the Year 4 and Year 5 Reports 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 Reading First schools since 2002 is the result of the program and not some other factor. Because the rest of the state K-3 schools have shown similar, though less dramatic, upward trends over the same time period, and because many non-Reading First schools have been found to be using Reading First-style program elements, it is likely that the statewide trend in non-Reading First schools is being driven by the same program elements that are driving the Reading First gains. This conclusion validates efforts to make such program elements available to all California elementary schools, not just those in Reading First.

References

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

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

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

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

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

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

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

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

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

Chapter 3: Implementation

This chapter presents data gathered from surveys of Reading First participants used to address the question: How well has the Reading First program been implemented in each participating school and district? Principal, reading coach, and teacher surveys provide a global perspective on implementation in Reading First schools as well as information about specific dimensions of program implementation such as professional development, material and instructional resources, understanding of Reading First Assurances and curricular materials, and perceptions of the Reading First program.

To evaluate the implementation of Reading First in California, Educational Data Systems (EDS) developed three surveys – one each for Reading First teachers, coaches, and principals – and administered them annually from 2004 to 2008. 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 2008, a total of 16,442 usable surveys were received from teachers, 468 from special education teachers, 887 from reading coaches, and 826 from principals, totaling 18,623 and yielding an estimated response rate of 95%.[22] Results of the surveys can be found in Appendices A – D of this report.

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

Key points in this chapter are:

- Measuring implementation is an essential element in assessing program effectiveness (i.e., the potential of a program to produce achievement gains given a sufficient level of implementation).
- The RFII can be interpreted as a (theoretical) percentage of times that teachers rate their schools "more than adequate" on relevant survey questions.
- Most schools in the Reading First program are implementing the program "adequately."
- The average level of implementation has risen only modestly through the duration of the Reading First program. The average (RFII) across all schools was 39 in 2006, 2007and 2008, compared to 36 in 2004 and 2005.

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

- There is no reason why most Reading First schools could not significantly increase their implementation of the program. The majority of schools could substantially improve their achievement scores by doing so.
- School-level implementation by the principal, and teacher evaluations of Reading First, are the two strongest predictors of achievement gains. This positive predictive effect offsets the negative predictive effects associated with having high percentages of SED, EL, black, and migrant students.

Measuring Reading First Program Implementation

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

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

In this chapter, we use survey data to quantify the degree of implementation occurring within each Reading First school. For each school, multiple respondents completed the survey, providing the perspectives of the site principal, the reading coach, and participating teachers. A school that may report a low level of use of curricular materials, neglects professional development, or inadequately allocates or employs instructional time, for example, would not be considered to be implementing the program. When "implementation" is defined in this more tangible way, assuming it can be measured with reasonable accuracy, it becomes feasible to decide whether the program has the *potential* of working if it is well implemented.

Rationale for Using a Survey

To directly measure the presence, absence, or degree of implementation of Reading First in *all* participating schools and districts is a daunting task. There is no statewide database that would definitively reflect Reading First implementation, and it is impossible within the scope of this evaluation to conduct observations at all sites. In 2008 there were 863 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 solicited in its Request for Proposals an implementation measure for *all* Reading First schools. To obtain information about implementation from all Reading First schools and districts, teachers, principals, and reading coaches in all Reading First schools were asked to complete a comprehensive survey constructed to gather information about the presence, absence, and degree of utilization of the critical elements that define the implementation of the Reading First program. For the first time in 2008, special education teachers were asked to complete a survey that paralleled the teacher survey. The special education teacher responses were not included in the construction of the school implementation measures.

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 five years of the survey use.[23] To summarize, the above issues are addressed as follows:

1. Schools are measured relative to each other rather than against an absolute standard.

2. Teachers complete the survey anonymously, enhancing their ability to report truthfully about the program. Because in most schools there is only one principal and one reading coach, their responses are not entirely anonymous, though school code numbers and not school names are used in the analysis process. A school's implementation measure pools together the teacher, principal, and coach responses.

3. Questions are worded so that their "correct" answers are not immediately obvious, increasing the chance that respondents select truthful answers.

4. There are numerous opportunities for cross-verification of findings across respondents within a school. Respondents not only report their own use of program elements but also rate other respondent types (coaches rate teachers, teachers rate coaches, etc.).

5. The implementation survey provides data that are used for making program adjustments and no "high-stakes" funding decisions rest on results. The "significant progress" regulations24 approved in fall 2007 are based entirely on achievement data.

6. Equating methods are used to equate responses across respondent groups and across program years.

7. The potential ambiguity at the question level is addressed by using statistical methods to group items' coherent dimensions that seem to cluster together statistically and are validated by experts

[23] 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.
[24] Information on "significant progress" is available at: http://www.cde.ca.gov/pd/ca/rl/rdfst06achievedef.asp.

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 reasonably valid and reliable as a means for measuring implementation at the school level.

Changes to the Survey

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

Individual questions throughout the survey underwent editorial modifications, often to clarify routing from section to section on the web survey. In 2005, based on a change in the Reading First program to include Spanish curricular materials for waiver classrooms (instruction in Spanish), the teacher survey was expanded to include additional questions involving the receipt and use of the Spanish versions of curricula. In 2006, further revisions were made to clarify which curricular materials were referenced in specific questions. In 2007, very minor wording changes clarified some items thought to be potentially confusing or no longer relevant in a program that has been in place for several years. In 2008, the addition of a special education survey necessitated some minor changes to the teacher survey to facilitate routing on the web survey, but there were no changes made to items included in the calculation of the RFII. In 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 (IRT) program called Facets, subsets of questions across the three surveys are used to generate measures on 17-19 dimensions.[25] IRT equating designs rely on common items that serve as links across forms and survey administrations. In 2008, the item difficulty calibrations which are the basis of survey equating were refreshed to correct for the effects of item "drift" over time. Most of the items showed little change in difficulty over time, but some types of items, in particular those asking about usage of program materials, have become easier over time. This most likely reflects the effect of increased familiarity and practice. Item analysis was performed individually for each of the 19 dimensions.

Three of the 19 dimensions are used to calculate each school's RFII. They are: School Implementation Overall (SIO), Overall Reading First Understanding (OUND), and Teacher/Coach Professional Development (TCPD).

The measures on these dimensions are weighted and combined to calculate the school's RFII. The weights are:

School Implementation Overall (SIO) = 70%

Overall Reading First Understanding (OUND) = 20%

Teacher/Coach Professional Development (TCPD) = 10%

The resulting RFII statistic is scaled to be between 0 and 100 and to have a distribution similar to that of the Reading First Achievement Index (RFAI). It is called the "Preliminary RFII". Based on advice from the EAG, as of 2007 the Preliminary RFII of a school in a given year is averaged with its Preliminary RFII from the preceding year (if one exists) to come up with a "Final RFII." Thus, the Final RFII assigned to each school in 2008 is an average of its 2008 Preliminary RFII and its 2007 Preliminary RFII. It is hoped that this 2-year rolling average approach makes each school's Final RFII more robust to changes in the sample of teachers in each school who take the survey while allowing it to be reflective of the school's recent implementation history. For purposes of this report, unless otherwise stated all references to the 2008 RFII signify the Final 2008 RFII, not the Preliminary 2008 RFII.

[25] 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.

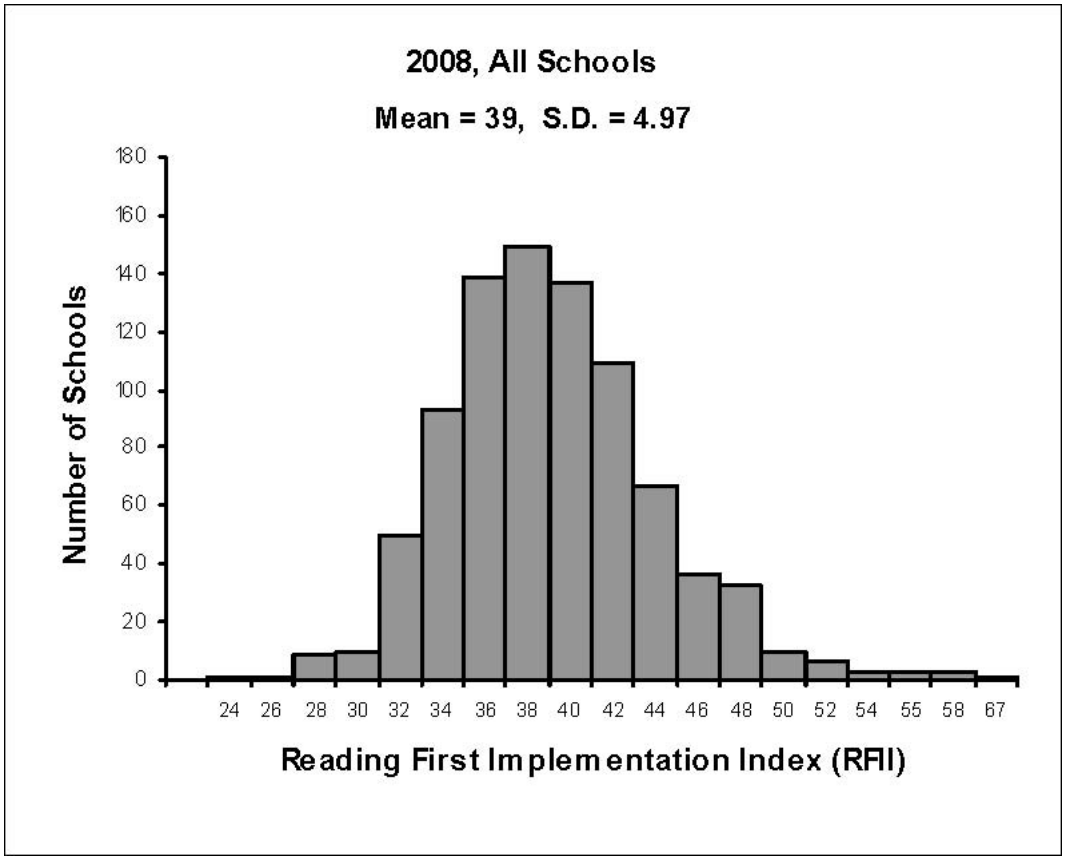
Implementation Results

Distribution and Interpretation of the RFII

Figure 3.1 shows how the RFII was distributed across all Reading First schools in 2008. The mean 2008 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 – 2008 Reading First Implementation Index (RFII), Distribution of Schools



Degree of implementation develops over time. Schools with more years in the program (Years In Program, or YIP) tend to have different implementation profiles than schools newer to the program.

Dimensions of Implementation

Table 3.1 shows the dimensions derived in the RFII calculation process as well as the RFII itself, their means for each year from 2004 to 2008, and the standard deviations for 2008. In the bottom row, we see the mean school RFIIs for 2004 through 2008 for all schools in the Reading First population. The 2008 RFII had a mean of 39 and a standard deviation of 5. The 18 dimensions are listed along with the sections of the implementation surveys they most depend on, and the number of items in each dimension. Three of these dimensions, set in bold type, were used to calculate the RFII. The means in the columns by year may be interpreted as the average percent of times (items) that teachers rated their school "more than adequate" on that dimension, averaged across schools. This is the same standard used for the RFII.

For this report we introduce several new columns. The two columns under the heading "Effect on Achievement" use the results of the meta-analysis discussed in Chapter 2 to quantify the effect of each dimension on achievement gains. Effect sizes are reported as a "standardized-beta coefficient," which is the number of standard deviations that the dependent or outcome variable (i.e., achievement) in a regression equation increases for each one standard deviation increase of the predictor variable. The table is sorted in descending order according to this effect size. All of these effects, except for Coach Professional Development, are significantly greater than zero with 95% confidence.

To provide context, an additional column called "Effect Relative to Demographics" is provided. This tells how large the implementation dimension effect is relative to the average effect sizes of the demographic variables that are included in the regression equations. The demographic variables are: percent of Socio-Economically Disadvantaged (SED) students in the school, percent of English Learners (EL), percent of blacks, percent of migrant students, and number of students in the school (included even though it is an institutional variable). Their average effect size, generally negative, is converted to a positive number called the "Mean Absolute Demographic Effect Size." Thus, each dimension's effect size is divided by this demographic effect size to get its "Effect Relative to Demographics." A value of 1.00 means that the dimension predicts (and causes, one assumes) achievement to the same degree that the demographic variables do, i.e., it is a powerful, meaningful effect.

The last column reports the correlation between each dimension and the 2008 RFII, thus the degree to which they are aligned. Correlations range from -1.00 to +1.00, where 0.00 means there is no relationship at all.

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

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

Table 3.1: All Schools, N (2008) = 863, Mean for Each Dimension, 2004-2008[1, 2, 3] Sorted by Effect on Achievement

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

[1] Dimensions 3, 11, and 17 are in bold because they are weighted contributors to Dimension 19, the RFII. The 2008 statistics are across 863 schools from the point of view of teachers for most dimensions. Dimensions 6, 10, and 18 are from the point of view of coaches. Dimensions 5, 7, and 13 are from the point of view of principals. The 2004 statistics are across 628 schools; the 2005 statistics across 808 schools; the 2006 statistics are across 856 schools, the 2007 statistics are across 885 schools; the 2008 statistics are across 863 schools.

[2] The two columns called Effect on Achievement are effect sizes derived using a meta-analysis of 221 regressions. The left column is the "standardized-beta coefficient." The asterisk "*" means the effect is statistically significant at the 95% confidence level. The right column is the RFII dimension effect relative to the Mean Absolute Demographic Effect for that dimension. A value of 1.0 means the dimension has the same predictive power as the demographic variables, on average.

[3] The statistics in the right column report each dimension's correlation with the RFII. The closer to 1.00, the more it captures what is meant by "implementation" as embodied by the RFII.

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

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

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

The most powerful non-implementation predictor of school achievement is "Teacher Reading First Evaluation" – how teachers evaluate the program at their school. When teachers are positive, schools grow rapidly. When they are negative toward the program, schools suffer. It is tempting to discount any causal relationship here, to say that when teachers feel negative about the program it is because of problems with the program, not with their own pre-dispositions toward the program. But that does not accord with the findings of this and previous reports. We already know that when the program is implemented, it is effective. Therefore, the high predictive value of teacher evaluations of Reading First suggests that when teachers feel positive about the program they actually do a better job of using it, and this has a causal relationship to achievement. When teachers feel negative about the program they do not use it properly and gains are small. Reading First has demonstrated potential to turn schools around in short order, even in very challenging demographic circumstances, but it cannot happen without teacher buy-in. Strong principal participation and positive teacher attitudes create Reading First effect sizes

sufficiently large to counter-balance the serious demographic challenges posed by high SED, EL, black, and migrant student populations.

Conclusions

Are Schools Implementing "adequately"? To interpret the implementation data, we rely on the procedures developed in prior reports that validate the RFII as a satisfactory measure of implementation. The RFII serves as a comparative benchmark for examining implementation by every school in the Reading First program. The RFII of an individual school can be viewed relative to some standard reference point that characterizes the population of schools as a whole. In the first year of implementation, the average RFII was 36. This became the (somewhat arbitrary) cut-point between "High Implementation" schools and "Low Implementation" schools. This distinction was used in conjunction with school achievement measures in other chapters to track the different achievement trend-lines for high implementing and low implementing Reading First schools (see Chapter 4 of the Year 4 Report, Chapter 2 of this report and of the Year 5 Report). To preserve comparability over time, the 36 as a cut-point continues to be used to define the upper boundary of the lower implementing schools. However, based on advice in 2007 from the EAG, the "High Implementation" schools 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."[26] By that criterion, the histogram in Figure 3.1 and Table 3.1 above reveal that schools are on average doing an "adequate" job of implementing the Reading First program, since the mean 2008 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 modestly. In 2004 and 2005, the mean RFII was 36 while in 2006, 2007 and 2008 it was 39. However, the RFII indicates that implementation has become stagnant: though adequate, schools have not been improving their level of Reading First implementation. Chapter 2 and an analysis of Table 3.1 show that increasing Reading First implementation is both highly achievable and beneficial.

[26] 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: Program Elements

The purpose of this chapter is to examine the importance of elements of the Reading First program according to the perceptions of Reading First participants. Chapter 3 reported the use of the Reading First Implementation Index (RFII) to measure school-level implementation. In that chapter, we examined how implementation developed statewide over time and that longevity in the Reading First program had a significant and positive impact on implementation. Now that the Reading First program has been implemented for six years, it is important to further examine how the key program elements have become integrated into reading instruction in California. In this chapter, we use individual items from the Reading First Implementation Survey to examine how teachers, coaches and principals viewed specific aspects of the Reading First program. We draw from four years of survey data to compile "lessons learned" from Reading First implementation in California.

This chapter yields the following key findings:

- Teachers, coaches and principals in Reading First schools consider several program elements as essential enough to sustain, including a protected Reading/Language Arts time block, collaborative planning and lesson study time, professional development, assessment and data analysis, reading coaching, and small group or universal access time.
- Initial, advanced and ongoing professional development were viewed as effective in preparing teachers to teach their adopted reading/language arts programs.
- In Reading First schools, the adopted reading/language arts curriculum use was high. Nearly 80% of teachers reported that 80% to 100% of their reading/language arts instruction relies on their adopted curriculum materials.
- The majority of teachers reported allocating more than the minimum number of daily minutes required for reading/language arts instruction in Reading First schools consistently over time. The use of a pacing schedule or guide for completing the grade-level curriculum in the academic year increased over time.
- There was moderate adherence to the required twice-monthly collaborative planning meetings at Reading First schools. A majority of survey respondents reported two or more meetings per month. However, despite the view that these meetings were important, a relatively high number of teachers over time, 43-48%, reported fewer than two planning meetings per month.
- Generally, school principals were viewed as supportive of Reading First and requiring full implementation.

- The coaching force has become more experienced and highly qualified over time in Reading First schools. Coaches are viewed as valuable resources to support implementation.

Research on Implementation

From previous reports and in this report (Chapter 2), we have shown that level of implementation has a significant and positive relationship with achievement. Higher implementing schools consistently show higher rates of growth for students in terms of reading achievement. The relationship between achievement and implementation noted in this Reading First evaluation is remarkable considering that research has demonstrated the inherent difficulty of linking achievement and implementation. Measuring implementation is difficult and complex, 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). In this evaluation study, we have consistently found that being in the Reading First program for three or more years has led to higher implementation.

Power, et al. (2005) suggest that measuring fidelity 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, an approach that has guided our study of implementation. This evaluation report takes both of these elements into consideration. In this chapter, we examine responses of teachers, coaches and principals on specific items of the survey that represent the main program elements and other items that examine the process of professional development and support provided by the Reading First program.

Data Sources

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

Perceptions of Program Elements

Sustainability of the Reading First Program

In this sixth year of evaluating California's Reading First program, the survey included an item to examine the sustainability of the program. Teachers, coaches and principals were asked, "If elements of your Reading First program had to be cut for funding or other reasons, which elements of the program would you most strongly support keeping in place? Select all that apply." Table 4.1 displays the results. Note that percentages do not total 100% because respondents were able to select multiple elements. In this table, we see some similarities across respondent groups in the elements they would prioritize for sustaining; teachers, coaches and principals are similar in their desire to continue to have a protected Reading/Language Arts time block, collaboration or lesson studies, professional development, assessment and data analysis, a reading coach, and small group or universal access time. The percentages are relatively lower overall for teachers than for coaches and principals for all elements.

The element with the highest percentage of selection by teachers (65%) was the protected and uninterrupted Reading/Language Arts time block, an item that also appears to be a priority for coaches (85%) and principals (78%). The Reading First assurances ask schools to allocate an uninterrupted 60 minutes in kindergarten and 150 minutes in grades 1-3. Along with this, planning instruction appears to be a high priority for teachers, coaches and principals, as is evident in the selection of Collaboration/Lesson Studies (Teachers, 42%; Coaches, 73%; Principals, 70%) and Structured Teacher Planning Time (Teachers, 53%; Coaches, 69%; Principals, 68%).

Coaching is another element of high priority, especially for principals and coaches. For principals (81%) and coaches (83%), the coaching element received the highest percentages, while for teachers (49%), it was the fifth highest. This reinforces the finding in the Year 5 report that coaches play a vital role in supporting reading instruction in California. Assessment and data analysis appears to be another element with high sustainability, though for teachers it was not as high in rank order (teachers, 45%; coaches 82%; principals 78%), while small group instruction and differentiating instruction through universal access also received high priority (teachers, 58%; coaches 81%; principals, 69%). Professional development, a cornerstone of the Reading First program was highly supported (teachers, 48%; coaches, 80%; principals, 76%). Curriculum materials for both waivered and non-waivered classrooms received relatively low rankings from all respondent groups. Other items appear to be fairly desirable to maintain, as can be seen in Table 4.1.

Table 4.1: Percentages of Teachers', Coaches', and Principals' Responses Regarding Sustainability of Elements of Reading First

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

Professional Development

A primary component of the Reading First initiative is professional development. The program is designed to provide school personnel with depth of knowledge of scientifically based reading practices and the ability to apply them in the classroom. 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 how much they participated in professional development and to rate the quality and helpfulness of their professional development experiences. Appendices A, B and C include the responses of participants to all survey items. Only 65% of teachers reported attending a Reading First related training during the 2007-2008 year. Because there were no new cohorts of LEAs this year, only new teachers, or teachers new to a grade level, would have attended SB 472 professional development, and only 3-5% at each grade level reported doing so. Likewise, only 3-5% at each grade level attended the Advanced second-year training, designed for those in their second year, while 33%, reported attending Advanced Mastery training, for those beyond the second year, and 32% reported that they did not attend Reading First related training. Two percent of teachers (2%) reported attending Coach training.

This year's survey results for teachers were similar to previous years' findings regarding the quality of the professional development institutes. Table 4.1 below provides results for an item that appeared on the survey from Year 3 through Year 6, asking how well the training prepared teachers to teach their adopted curriculum. Though fewer teachers attended a professional development institute in Year 6 than previous years, most teachers reported that the institute prepared them adequately or very well.

Table 4.2: Teacher Survey Results for Quality of Professional Development

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

Note: Results are in percent of total surveys collected on the item.

Each year of the program, teachers were to engage in 80 hours of follow-up professional development provided at the district or school level. Table 4.3 demonstrates the percentages of teachers reporting their level of involvement in the follow-up sessions. The percentages of teachers reporting participation in the full 80 hours declined slightly over time, with the percentage of teachers declining to report increasing. Given the perceived importance of collaborative planning and lesson studies, it is possible that some of the time that would have been devoted to follow-up professional development was reallocated to collaborative panning activities.

Table 4.3: Teacher Survey Results for Follow-up Professional Development

Item B5: How many hours of the 80-hour follow-up to the Reading Professional Development Institute will you have completed by the end of the school year?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. Not applicable	8	8	10	10
b. Less than 20 hours	4	3	3	2
c. 20-39 hours	3	2	2	2
d. 40-59 hours	5	4	4	3
e. 60-79 hours	3	3	2	2
f. 80 or more hours	65	63	58	49
No response to this item	12	17	21	32

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 in the appendix. One item in particular provides insight into trends in curriculum use over time. Item F3 from the teacher survey, asks, "What percentage of your total reading/language arts instruction relies on materials from your district's adopted program?" Table 4.4 displays the results of this question asked in Years 3 – 6. Consistently over time, approximately 80% of teachers have reported that 80% to 100% of their instruction relies on their adopted curriculum materials.

Table 4.4: Teacher Survey Results for Curriculum Use

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

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

Teachers, coaches and principals were asked to rate the overall effectiveness of their district's adopted reading/language arts program in item I1 in each year of the survey. Results from this item over time are included in Table 4.5 below. Results for teachers, coaches and principals have been consistent over the four years, with most reporting the curriculum to be either good or excellent in its effectiveness.

Table 4.5: Percentages of Teachers', Coaches', and Principals' Responses Regarding Curriculum Effectiveness

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

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 4.5 shows the amount of time reported by teachers spent in teaching their adopted curriculum from 2004 forward. The time allocation results are fairly stable over time.

Table 4.6: Percentages of Teachers Responses Regarding Time Allocation

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

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

Collaborative Teacher Meetings

The Year 5 report included evidence that the collaborative planning meetings promoted in the Reading First program have been viewed as important and useful. Collaborative planning meetings are supposed to occur twice monthly and should focus on analyzing student data, understanding the curriculum materials, improving instructional strategies, and assisting struggling readers. Table 4.6 presents findings from a question asked of teachers (Question D2), coaches (Question E2) and principals (Question E2) regarding how often the school provided time for teachers to plan collaboratively. Data are displayed for Years 3 – 6. Findings for each respondent group were consistent across years, though fewer teachers reported two or more times per month than coaches and principals. A majority of survey respondents reported two or more meetings per month. However, despite the view that these meetings were important, a relatively high number of teachers over time, 43-48%, reported fewer than two planning meetings per month. This may be an indication that it is difficult for schools to allocate teacher meeting time for planning due to competing demands on limited out-of-classroom time for teachers.

Table 4.7: Percentages of Teachers, Coaches, and Principals Regarding Collaborative Planning Time

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

Pacing of Instruction

The Reading First program has required districts to develop pacing plans or guides for ensuring consistency across classrooms in terms of content covered and to ensure that students move through the grade-level standards and aligned curriculum. Pacing plans provide guidelines for what lessons should be taught in time periods spaced throughout an academic year. If teachers adhere to the pacing guidelines, they should cover the entire year's curriculum. The survey asked participants whether they had a pacing schedule and how closely they adhered to it. Table 4.7 presents results from teachers (Question D1), coaches (E1) and principals (E1) on this question. Nearly all participants reported that they have a pacing schedule. A smaller proportion of teachers than coaches and principals reported that their pacing schedule provides detailed guidance about what lessons to teach on a daily or weekly basis. These percentages are consistent with previous reports.

Table 4.8: Percentages of Teachers, Coaches, and Principals Regarding Pacing Plans

Item D1, Teachers; Item E1, Coaches/ Principals: Does your school have a pacing schedule?	Teachers %	Coaches %	Principals %
a. My school does not have a pacing schedule	2	0	1
b. My school has a pacing schedule based only on the assessment schedule	31	17	13
c. My school has a pacing schedule that identifies lessons on a daily or weekly schedule and when to give assessments	66	82	84

Additional information about pacing schedules was provided by teachers (Question F4). Table 4.9 shows that teachers who reported precise adherence to the pacing schedule significantly increased from Year 3 to Year 4, and then increased slightly for Years 5 and 6. This suggests that, as teachers developed expertise and depth of knowledge through long-term participation in the Reading First program, they were more able to fully implement the pacing schedule.

Table 4.9: Teacher Survey Results for Adherence to Pacing Schedule

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

Leadership

Improving the capacity of school leadership to support an effective reading program has been a central feature of the Reading First program. Teachers responded to two questions regarding the role of the school administrator in program implementation, displayed in Table 4.10. Question D11 indicates that over 80% of teachers have consistently reported adequate or more than adequate support from their principals. Question D12, asked only from Years 4-6, indicates that 88-90% of teachers reported that they were required by the school principal to fully implement their adopted reading/language arts program.

Table 4.10: Teacher Survey Results Regarding School Leadership

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

Coaching

The Reading First program has played a significant role in establishing reading (or literacy) coaches in districts throughout the state. Coaches are out-of-the-classroom teachers with expertise in research-based instructional strategies and state-adopted reading curricula. The Year 5 report included a separate chapter examining the role of reading coaches in the Reading First program. Here, we highlight a few key survey items related to the use of reading coaches. The reader is referred to the Year 5 report for an in-depth discussion of the Reading First coaching model (www.eddata.com).

The Year 5 report indicated that, through the Reading First program, the state has built a highly qualified coaching force. Despite turnover in the coaching force, many of the coaches who served in the Reading First program stayed with the program and developed sufficient expertise to serve as instructors for the professional development institutes, acquired Reading Specialist certification, and moved on to administrative roles. Table 4.10 shows the level of coaches' experience with their districts' adopted reading/language arts program over time. In the Year 6 survey, 28% of coaches reported having five years of experience and 57% reported 6 or more years of experience with the adopted program, or 85% with five or more years of experience. From Year 3 to Year 4, we see a significant drop in the percent of coaches with two or fewer years of experience with the adopted program.

Table 4.11 Level of Coach Experience

Coach Survey A2: How many years of experience do you have with your district's adopted reading/ language arts program?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. Less than 1 year	3	1	1	0
b. 1 year	4	1	1	0
c. 2 years	28	7	2	2
d. 3 years	26	29	8	2
e. 4 years	11	24	28	9
f. 5 years or more	27	36	59	28
g. 6 years or more				57

By definition, reading coaches should serve as resources for schools and classroom teachers in implementing the reading curriculum. Additionally, coaches are likely to be in the role of facilitating the collaborative teacher planning meetings that occur twice monthly. Table 4.12 shows the perceptions of teachers, coaches and principals regarding the level of support provided by coaches to teachers regarding the implementation of the curriculum from the perspectives of teachers, coaches and principals. Results were consistent across years for teachers, coaches and principals. Teachers consistently had a lower frequency of reported instance of useful, specific, and detailed answers to questions than did coaches or principals.

Table 4.12: Percentages of Teachers', Coaches', and Principals' Responses Regarding the Coach as a Resource

Teacher Survey (E2): How helpful is your coach in answering questions about how to teach the program?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
Not Applicable. My school does not have a reading coach (only available for Year 3)	0			
a. The coach often doesn't know more than I do about how to teach the program	8	7	7	8
b. The coach gives general answers to questions	24	24	24	24
c. The coach gives specific, detailed answers that teachers can use	64	66	66	65
Coach Survey (F5): How helpful do you feel you are in answering teacher questions about how to teach the program?				
a. I often don't know more than the teachers about how to teach the program	0	1	1	0
b. I am able to give general answers to questions	15	10	12	9
c. I give specific, detailed answers that teachers can use	83	87	86	90
Principal Survey (F5): How helpful is your coach in answering questions about how to teach the program?				
Not Applicable. My school does not have a reading coach (only available for Year 3)	1			
a. The coach often doesn't know more than I do about how to teach the program	1	0	0	0
b. The coach gives general answers to questions	10	7	8	7
c. The coach gives specific, detailed answers that teachers can use	86	88	87	88

The coach typically has the role of facilitating grade-level teacher meetings designed to focus on data analysis, instructional strategies, struggling students, and general program implementation. A question on the surveys asked teachers, coaches and principals about the coach's role in these meetings. Table 4.13 shows results from Years 3 through 6. Findings were consistent across years. Teachers reported at a higher rate than coaches and principals that the coach was not involved in these meetings, but overall coaches seemed to be facilitating the meetings and maintaining focus on instructional needs.

Table 4.13: Percentages of Teachers', Coaches', and Principals' Responses Regarding the Coach as a Facilitator

Teacher Surveys E4: Does the coach facilitate regular grade-level teacher meetings related to your district's adopted reading/language arts program?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
Not applicable, my school does not have a reading coach. (only available for year 3)	1			
a. The coach is not involved with the grade-level meetings	23	21	23	25
b. The coach helps facilitate the meetings regularly	45	45	46	44
c. In addition to facilitating meetings, the coach keeps them focused instructional needs of teachers	28	31	29	28
Coach Surveys F7: Do you facilitate regular grade-level teacher meetings related to your district's adopted reading/language arts program?				
a. I am not involved with the grade-level meetings	12	11	12	15
b. I facilitate the meetings regularly	38	35	37	34
c. In addition to facilitating meetings, I keep them focused on instructional needs of teachers	46	49	49	48
Principal Surveys F7: Does the coach facilitate regular grade-level teacher meetings related to your district's adopted reading/language arts program?				
Not applicable, our school does not have a reading coach. (only available for year 3)	1			
a. The coach is not involved with the grade-level meetings	8	7	7	7
b. The coach helps facilitate the meetings regularly	34	34	33	33
c. In addition to facilitating meetings, the coach keeps the focus on instructional needs of teachers	55	53	56	56

Assessment

Utilizing ongoing, curriculum-embedded assessment to determine students' specific learning needs, monitor progress and adjust instruction is another key element of the Reading First program. The survey included several items regarding participants' use of the 6-8 Weeks Skills Assessments. Based on findings in the Year 5 report expressing the importance of using this type of assessment, as well as specific concerns regarding the process or tools, this Year 6 report explores the importance and use of the assessments in detail in Chapter 5. Additionally, findings from an open-ended survey question are discussed.

Conclusions

The findings reported in this chapter demonstrate that the Reading First program has led to the development of a well-integrated structure and process of providing reading/language arts instruction in California. The program elements outlined in the Reading First assurances are integral parts to a whole that is more than the sum of its parts. Taken individually, each program element would not likely have the effect of impacting reading instruction nor would they individually be sustainable. The use of state-adopted curricula, professional development, coaching, leadership support, protected time blocks, and other program elements together form an integrated reading program that has impacted reading achievement in the state. It is important for state leadership and policy makers to consider the importance of sustaining these program elements as interconnected and essential ingredients of an effective reading/language arts program.

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.

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.

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.

Chapter 5: The Role of Assessment in Reading First

The purpose of this chapter is to examine the perceptions of Reading First participants regarding one of the essential elements of the Reading First program, assessment. The Reading First program requires that LEAs and schools use data from curriculum-embedded assessments to identify students' learning needs, monitor program effectiveness and make instructional decisions. The C-TAC makes the 6-8 Weeks Skills Assessments available to every Reading First LEA. The assessments are conducted every 6 to 8 weeks and are designed to measure student progress in specific skills taught during that 6-8 weeks time span.

In this chapter, we use information from the Reading First surveys to examine the responses of teachers, coaches and principals to survey items related to the assessment process in Reading First. 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 assessments used. This question asked, "*What is your opinion of the 6-8 Weeks Skills Assessments (from SCOE)*? *How are they useful? How could they be improved?*"

This chapter yields the following key findings:

- The 6-8 Weeks Skills Assessments were consistently used to monitor student progress and guide instruction during the past four years. The use of these assessment tools appeared to increase from 2004-2005 to 2007-2008.
- School principals have become increasingly involved in a leadership role over the past four years regarding the use of data to guide reading instruction. This is likely due to their involvement in professional development for administrators, working closely with coaches and other experiences acquired through the Reading First program. However, coaches and principals reported a higher level of principal involvement than did teachers.
- Coaches provide valuable support in data analysis and interpretation. They help to keep the focus of collaborative planning meetings on the use of data to better understand students who are struggling with reading development and appropriate reading intervention strategies to address their needs.
- Teachers, special education teachers, coaches and principals generally found the 6-8 Weeks Skills Assessments to be useful for monitoring student progress, guiding instruction, helping to identify students who need additional assistance, and helping to plan reading intervention for small-group follow-up instruction.
- Suggestions for improving the 6-8 Weeks Skills Assessments included improving the alignment of the assessments with the specific skills taught and the state standards; improving the test format,

procedures, or timing; and specific suggestions regarding how fluency, comprehension and vocabulary are assessed.

Research on Reading Assessment

Assessment is an integral part of teachers' pedagogy. Skilled reading teachers intuitively gather information informally about their students' mastery of concepts and skills through observation as their student read aloud and participate in classroom tasks and through planned periodic data collection using appropriate assessment tools. Yet, intuition and informal observation are not sufficient for making instructional decisions and maximizing students' learning. Studies have shown that when teachers also systematically use specific curriculum-based data regarding students' acquisition of skills, they make better-informed decisions about what to teach and how to effectively deliver instruction (Fuchs, Fuchs, Hamlett & Stecker, 1991; Haager & Klingner, 2005). The Reading First program guidelines require that participating schools use ongoing assessment with valid and reliable assessment tools to "determine whether students are making adequate progress or need more support to achieve grade-level reading outcomes (U. S. Department of Education, 2002)."

In California, prior to implementation of the Reading First program, there was no statewide directive for systematic and ongoing assessment. LEAs may or may not have incorporated district-adopted assessments into their instructional plans. The annual state testing, currently the California Standards Test (CST), was the only statewide requirement for obtaining data regarding students' reading achievement. Though the CSTs provide information about individual students' reading proficiency, they do not provide the ongoing and curriculum-based data needed for adjusting and refining instruction. The 6-8 Weeks Skills Assessments developed and distributed by the C-TAC at the Sacramento County Office of Education (SCOE) have provided the tools for the systematic, ongoing, classroom-based assessment required of the Reading First program. These assessments were designed to parallel the curriculum and assess the skills taught during the previous 6-8 weeks of study. Additionally, oral reading fluency is assessed through timed readings during each assessment period in first through third grades. The fluency scores are compared to national norms to determine if students are building fluent reading skills. These assessments have formed a comprehensive package for examining student progress statewide, district-wide and at the local school and classroom levels.

Survey Responses Regarding the 6-8 Weeks Skills Assessments

Several items on the teacher, coach and principal surveys asked about the role and use of the 6-8 Weeks Skills Assessments in Reading First schools. Results from relevant items are organized into five categories: Purpose and Use, Leadership Role, Management, Data Analysis and Guiding Instruction.

Purpose and Use of Assessments

The intended purpose of the assessments is to provide school and district personnel with data regarding student progress in reading and to assist them in making adjustments to instruction that would boost student achievement. School personnel administer the assessments every six to eight weeks and have the option to enter the data into a data management system that provides data reports to be used in data analysis and planning meetings. The assessments for each grade level, K – 3, focus on students' competency with skills taught during the six to eight weeks of instruction and, in grades 1 – 3, include a timed oral reading fluency check that is compared with grade-level, research-validated norms. Students who lag behind in specific skills or in overall fluency should be targeted for instructional support that would be provided during small group instruction.

The survey asked participants their views of the purpose of the assessments. Table 5.1 shows the responses of teachers, coaches and principals from Years 3 – 6. Results were consistent across years and across respondent groups. Most teachers, coaches and principals reported that these assessments were used to monitor student progress or guide instructional decisions, the two primary purposes of the assessments. Because respondents were asked to select only one, it is understandable that responses were fairly equally split between these two purposes.

Table 5.1: Percentages of Teachers, Coaches, and Principals Regarding Assessment Purpose

Teachers Item D6: What is the primary purpose of the 6-8 Week Skill Assessments in your school, at your grade? Select only one.	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. Skill assessments are not administered	4	2	2	2
b. To monitor student progress	42	38	38	37
c. To guide instructional decisions	48	54	55	56
d. To challenge students to achieve	2	3	2	3
e. To compute grades for report cards	2	2	2	2

Coaches Item E4: What is the primary purpose of the 6-8 Week Skill Assessments in your school, at your grade? Select only one.	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. Skill assessments are not administered	1	0	0	0
b. To monitor student progress	23	22	21	21
c. To guide instructional decisions	73	73	76	76
d. To challenge students to achieve	1	0	1	1
e. To compute grades for report cards	1	1	1	1
Principals Item E4: What is the primary purpose of the 6-8 Week Skills Assessments in your school, at your grade? Select only one.				
a. Skill assessments are not administered	0	0	0	0
b. To monitor student progress	19	15	18	15
c. To guide instructional decisions	79	81	79	81
d. To challenge students to achieve	2	1	1	2
e. To compute grades for report cards	0	1	0	0

Teachers, coaches and principals were asked which assessments were used in reading every six to eight weeks. Schools were required to report data from the 6 – 8 Week Skill Assessments for the purpose of monitoring the Reading First program, and, at the end of the year, as part of the evaluation process. Accomplished teachers often use multiple data sources to better understand their students and to make instructional decisions, so it is not surprising that teachers in Reading First schools reported using teacher-developed, publisher-developed, district-developed or other assessments in addition to the 6-8 Weeks Skill Assessments. The reported use of the 6-8 Weeks Skills Assessments increased over time according to teachers, coaches and principals. It is surprising, however, that teachers consistently reported a lower level of use of The 6-8 Weeks Skills Assessments than did coaches and principals from Year 3 through Year 6. Table 5.2 provides data from teachers, coaches and principals from Years 3-6. Percentages reported by each respondent group were consistent across years.

Table 5.2: Percentages of Teachers, Coaches, and Principals Regarding Assessment Use

Teachers Item F6: If you assess your students in reading every six to eight weeks, which assessments do you use? Select all that apply.	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. I do not assess students in reading progress every six to eight weeks	5*	3	3	3
b. I use teacher-developed assessments that my colleagues or I have written	15*	18	20	22
c. I use assessments that come from the publisher with the adopted program	49*	42	42	40
d. I use the 6-8 Week Skill Assessments	60*	74	77	77
e. I use district-developed assessment	*	28	30	29
f. I use assessments other than those listed above	18*	13	14	13
Coaches Item G6: If teachers assess their students in reading every six to eight weeks, which assessments do they use for this purpose? Select all that apply.				
a. Teachers do not assess students in reading progress every six to eight weeks	1	0	0	0
b. Teachers use teacher-developed assessments that they or their colleagues have written	9	12	17	19
c. Teachers use assessments that come from the publisher with the adopted program	43	37	41	39
d. Teachers use the 6-8 Week Skill Assessments	86	89	95	95
e. Teachers use district-developed assessment	*	22	23	21
f. Teachers use assessments other than those listed above	13	10	9	11
Principals Item G6: If teachers assess their students in reading every six to eight weeks, which assessments do they use for this purpose? Select all that apply.				
a. Teachers do not assess students in reading progress every six to eight weeks	1	1	0	0
b. Teachers use teacher-developed assessments that they or their colleagues have written	9	10	12	17
c. Teachers use assessments that come from the publisher with the adopted program	44	45	48	46
d. Teachers use the 6-8 Week Skill Assessments	82	87	91	90
e. Teachers use district-developed assessment	*	33	37	39
f. Teachers use assessments other than those listed above	17	13	15	13

* The selection choices for this item on the 2005 surveys were worded slightly differently than subsequent years, but the meaning was essentially the same. The selection regarding use of district-developed assessments was not given as a choice in that year.

An additional item (Item I5 for teachers and I7 for coaches and principals, discussed in Chapter 4) asked participants, "If elements of your Reading First program had to be cut for funding or other reasons, which

elements of the program would you most strongly support keeping in place? Select all that apply." One of the several choices offered was, "Assessment and Data Analysis." Only 45% of teachers reported that they would continue the use of assessments compared to 82% of the coaches and 78% of the principals. It appears that the data, and data analysis process, at least with the use of these assessment tools, may be viewed as more important to coaches and principals. Later in this chapter, this topic is explored in greater depth through a qualitative analysis of an open-ended question.

Leadership Role in the Assessment Process

The roles of the reading coach and principal are important in ensuring that data collection and analysis are meaningful and relevant to Reading First implementation. One survey item asked participants about the principal's role. Table 5.3 below displays teachers', coaches' and principals' views of the principal's role in the assessment process. From the perspective of each respondent group, it appears that principals' involvement in using the data to make instructional adjustments increased consistently from Year 3 to Year 6. This is evidence that the administrative force in Reading First schools developed capacity through the professional development, support and experience acquired through the Reading First program. The percentage of teachers who viewed the principals as integrally involved is smaller than that of the coaches and principals.

Table 5.3: Percentages of Teachers, Coaches, and Principals Regarding the Principal's Role in the Assessment Process

Teachers, Item D5: How involved is your school principal with the 6-8 Week Skill Assessments?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. The principal is generally not involved with the skill assessments	34	26	24	22
b. The principal makes sure the skill assessments take place, but does not track results	14	8	7	6
c. The principal makes sure that the skill assessments take place and keeps track of the results	30	32	32	32
d. The principal makes sure that the skill assessments take place, tracks results, and requires that instruction be adjusted as necessary	18	33	36	38

Coaches Item E3: How involved is your school principal with the 6-8 Week Skill Assessments?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. The principal is generally not involved with the skill assessments	22	13	14	10
b. The principal makes sure the skill assessments take place, but does not track results	17	12	9	7
c. The principal makes sure that the skill assessments take place and keeps track of the results	33	28	30	29
d. The principal makes sure that the skill assessments take place, tracks results, and requires that instruction be adjusted as necessary	27	44	45	52
Principals Item E3: How involved are you with the 6-8 Week Skill Assessments in your school?				
a. I am generally not involved with the skill assessments	9	4	5	3
b. I make sure the skill assessments take place, but do not track results	12	6	5	4
c. I make sure that the skill assessments take place and keep track of the results	43	31	30	29
d. I make sure that the skill assessments take place, track results, and require that instruction be adjusted as necessary	35	57	59	63

How do the coach and the principal work together as site leaders regarding the assessment process? An item on the Year 6 Coach survey asked, "As a reading coach, the conversations you have with your principal focus on what topics? Check all that apply." Of the wide range of choices to select, one directly addressed the use of assessment, "Analyzing the 6-8 Week Skill Assessment data." This response category was third highest with 76% of respondents selecting it, following "Addressing instructional needs of teachers (86%) and "Planning site professional development programs and services (83%).

Management of Assessment

A common complaint from teachers is the time and effort required to conduct frequent assessments. How, then, have the Reading First schools managed this endeavor? Though teachers are responsible for administering and scoring their students' assessments, the coach may provide support in various ways. Table 5.4 below presents findings from a survey question regarding the role of the coach in assisting with assessments. Results were consistent across years and across respondent groups. Generally, the coaches were very involved in reviewing and interpreting assessment results.

Table 5.4: Perceptions of Teachers, Coaches, and Principals Regarding the Coach's Role in the Assessment Process

Teachers, Item E6: Does the coach help you with the 6-8 Week Skill Assessments?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. Not applicable. My school does not administer the 6-8 Week Skill Assessments.	3	3	2	2
b. The coach is not involved with these assessments	14	11	12	13
c. The coach makes sure the assessments take place, but does not review results	16	12	11	12
d. The coach helps interpret the assessments and reviews results	63	71	72	70
Coaches, Item F9: Do you help the teachers with the 6-8 Week Skill Assessments?				
a. Not applicable. Our school does not administer the 6-8 Week Skill Assessments.	2	0	0	0
b. I am not involved with these assessments	2	1	2	1
c. I make sure the assessments take place, but do not review results	5	2	2	1
d. I help interpret the assessments and review results	91	93	95	97
Principals, F9: Do coaches help the teachers with the 6-8 Week Skill Assessments?				
a. Not applicable. My school does not administer the 6-8 Week Skill Assessments.	1	0	0	0
b. The coach is not involved with these assessments	2	2	1	1
c. The coach makes sure the assessments take place, but does not review results	4	4	3	2
d. The coach helps interpret the assessments and reviews results	91	90	92	93

Data Analysis and its Link with Instruction

The use of data to inform instruction is a key feature of the Reading First program. The data that are collected every six to eight weeks are compiled into reports that are used by principals, reading coaches, individual teachers, and grade level groups to better understand how the students are progressing in the grade level curriculum, which individual students may need additional support, and how instruction might be adjusted to achieve better results. The collaborative grade level meeting forum is the primary avenue for openly discussing assessment results and how they link to instruction. Question D9 on the teacher survey (E6 for coaches and principals) asked, "What topics are discussed at grade-level meetings? Select all that apply." Table 5.5 below shows the findings over time for this question. In this table, we see that the discussion of assessment results was consistently reported as a topic for meetings. We also see an

increase over time in the discussion of students experiencing difficulty with reading, as well as the use of intervention strategies, both issues that would likely be informed by the data.

Table 5.5: Perceptions of Teachers, Coaches, and Principals Regarding the Focus on Data Analysis at Collaborative Grade Level Meetings

Teachers, Item D9: What topics are discussed at grade-level meetings? Select all that apply.	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. Not applicable	3	1	1	1
b. Instructional reading/language arts strategies	88	88	89	88
c. School-level administrative issues and announcements	46	49	51	50
d. Students who are having trouble	66	70	72	72
e. Extracurricular activities	24	28	30	30
f. Reading/language arts assessment results	85	88	88	87
g. Intervention strategies	77	81	82	81
h. The school's and district's mission	20	21	22	23
i. Issues in the field of education	28	29	29	28
j. Teacher professional development issues	45	46	45	44
k. Upcoming special events	42	48	50	49
l. Issues related to specific teaching practices that are part of your adopted reading/language arts program	76	79	78	76
Coaches, Item E6: What topics are discussed at grade-level meetings? Select all that apply.				
a. Not applicable	2	1	0	1
b. Instructional reading/language arts strategies	89	93	96	95
c. School-level administrative issues and announcements	27	39	37	36
d. Students who are having trouble	62	74	78	80
e. Extracurricular activities	13	22	24	25
f. Reading/language arts assessment results	87	94	94	94
g. Intervention strategies	78	85	90	90
h. The school's and district's mission	13	19	23	22
i. Issues in the field of education	16	25	23	20
j. Teacher professional development issues	41	47	52	49
k. Upcoming special events	26	36	36	37
l. Issues related to specific teaching practices that are part of your adopted reading/language arts program	82	88	91	90

Principals, Item E6: What topics are discussed at grade-level meetings? Select all that apply.	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. Not applicable	2	1	0	0
b. Instructional reading/language arts strategies	91	95	97	96
c. School-level administrative issues and announcements	18	23	25	20
d. Students who are having trouble	66	76	80	82
e. Extracurricular activities	9	15	17	14
f. Reading/language arts assessment results	90	94	96	96
g. Intervention strategies	86	92	94	94
h. The school's and district's mission	17	28	30	28
i. Issues in the field of education	13	23	22	19
j. Teacher professional development issues	47	57	59	56
k. Upcoming special events	18	23	25	22
l. Issues related to specific teaching practices that are part of your adopted reading/language arts program	82	91	92	90

The link between data analysis and instruction is further demonstrated by a survey question regarding the use of assessments. Table 5.6 below shows perceptions over time of how the assessment results were used, according to teachers, coaches and principals. Here, we see consistent findings over time and across respondent groups. The percentage of teachers, coaches and principals selecting the use of assessments to guide instruction were higher in Year 3, but this may be due to the fact that monitoring student progress was not an option, and this item only allowed one choice.

Table 5.6: Perceptions of Teachers, Coaches, and Principals Regarding the Use of Assessments

Teachers, Item F7: How do you primarily use results of the 6-8 Week Skill Assessments?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. (Year 3 only) I don't assess student progress every 6-8 weeks.	4*			
a. I don't use the results (Year 3, "I give the assessments but I don't use the results.")	5*	2	2	2
b. I use the results to monitor student progress every six to eight weeks	*	33	33	33
c. I use the results to guide my teaching	88*	61	61	61
Coaches, Item G7: How do teachers primarily use results of the 6-8 Week Skill Assessments?				
a. (Year 3 only) They do not assess reading progress every 6-8 weeks.	1*			
a. Teachers don't use the results	11*	1	1	1
b. Teachers use the results to monitor student progress every six to eight weeks	*	32	35	33
c. Teachers use the results to guide their teaching	87*	63	63	64

Principals, Item G7: How do teachers primarily use results of the 6-8 Week Skill Assessments?	2004-2005 %	2005-2006 %	2006-2007 %	2007-2008 %
a. (Year 3 only) Teachers do not assess reading progress every 6-8 weeks.	1*			
a. Teachers don't use the results	6*	1	1	0
b. Teachers use the results to monitor student progress every six to eight weeks	*	27	29	26
c. Teachers use the results to guide their teaching	90*	69	69	71

* The selection choices for this item on the Year 3 (2004-2005) surveys were worded slightly differently than subsequent years but the meaning was essentially the same. The selection regarding use of results to monitor student progress was not given as a choice in Year 3.

Participants' Perceptions of the 6-8 Weeks Skills Assessments

In this section, we use qualitative research methodology to examine findings from the open-ended question included on the survey, "*What is your opinion of the 6-8 Weeks Skills Assessments (from SCOE)*? *How are they useful? How could they be improved?*" Teachers, coaches and principals who work in Reading First schools are positioned to provide important insight into the assessment process promoted by the Reading First program because they are the personnel who work directly with assessment results at the classroom and individual student levels. The open-ended question regarding the assessments asks participants to explain how they are useful and make suggestions for improvement. This question provided an opportunity for participants to provide feedback and voice their opinions about the assessments.

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.

A Note about Qualitative Analysis

Previous reports have used qualitative research methodology to distill open-ended responses into categorized findings. 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 analyses, such as those reported for survey items in the previous section of this chapter. When used in conjunction with quantitative statistical analysis, qualitative research can provide a deeper explanation of statistical results. Of the 16,442 teacher surveys collected, 12,018 wrote codable narrative responses to this question, or 73.1%. Of the 468 special education teacher surveys collected, 315 wrote codable responses to this

question, or 67.3%. Of the 887 coach surveys collected, there were 727 codable comments submitted, or 82.0%. Of the 826 principal surveys collected, there were 598 codable comments submitted, or 72.4%. This is a high response rate for an 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. Because all participants did not choose to provide narrative comments, it is impossible to determine if those that did are representative of the entire 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 the multiple perspectives of teachers, special education teachers, reading coaches, and principals. The high response rate for each respondent group adds credibility to the findings.

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-fourth of the total codable responses indicated that the assessments are used to monitor student progress. This does not mean that three-fourths did not use the assessments to monitor progress. It merely means that, for one-fourth of the respondents who chose to comment, 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 three-fourths who did not elect to write about progress monitoring would agree that the ability to monitor progress with these assessment tools is a useful aspect of the assessment program. In fact, in Table 5.6 above, 32-35% of teachers indicated that they use the assessments for progress monitoring. However, in that item, respondents could select only one choice and the primary use of the assessments was viewed as to guide instruction. It is important to bear in mind that in an open-ended format, the responses are not directed by multiple-choice items or suggested responses.

Data Analysis Methodology

For this qualitative analysis, the text file extracted from the online survey was subjected to analysis using a qualitative software package, Qualrus (Brent, Slusarz & Thompson, 2002). Data reduction involved coding individual responses, or "chunks" of data, according to its meaning relative 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. 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.

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. For each category of responses, the sample comments help to interpret the findings relative to that category. The sample comments were selected on the basis of being collectively representative of those coded with that response category.

"*I believe they [the assessments] effectively monitor the progress of my students. They are tools that I use, not only to monitor my students' progress but to also monitor my own teaching.*" This quote taken

from the teacher responses to the open-ended question demonstrates the intended link between assessment and instruction in Reading First. This teacher sees the potential power of fully implementing the assessment system because she mentions not only how this information helps her to make decisions about students but also about the instruction she provides.

For each response category below, the definition used by the research team to assign codes is provided along with a sampling of representative comments from teachers, special education teachers, coaches and principals. When possible, comments are provided across respondent groups. In some cases, representative comments were taken from only one or two groups (e.g., teachers) because illustrative comments could not be found within the other groups. Though every effort was made to maintain the original meaning of the quote, some quotes received minor editing- no more than a word or two - to correct faulty grammar or spelling, or to clarify context when the quote was pulled from a longer response.

Table 5.7 Response Category Descriptions and Representative Comments

Monitor Progress *The assessments are useful for monitoring student progress. The school uses the assessment data to determine which students have made progress or whether the class as a whole is making adequate progress. The idea is that there is an ongoing process of looking at data to make sure students are keeping step with the curriculum, or making appropriate growth in skills.*	<u>Teachers</u> " I feel they (the assessments) are an adequate indicator of student progress and can also show areas where different strategies can be utilized for student advancement." "The assessment is solid review of the unit. It allows me to recheck student memory and development." "They allow for structured monitoring of student progress. They provide guidance for lesson planning and student help needed or not needed." <u>Special Education Teachers</u> "I use the assessments to evaluate student progress and to determine teaching strategies to meet student needs. I also use it to compare RSP Student levels of performance with their grade level peers." "It is extremely useful for monitoring student progress, identifying students needs, and guiding instruction." <u>Coaches</u> "The 6-8 Week Skills Assessments have been very helpful in showing teachers the importance of assessment and monitoring student learning." "All assessment information informs instruction and helps monitor progress. While they may not be strongly connected to the direct instruction provided in the program, they help us understand a full picture regarding student progress."

	Principals "Assessments are good to drive instruction and guide teachers to make flexible groupings to address individual needs."
Guide Instruction *The assessments provide information that is useful in guiding instruction. The assessment data helps the teacher to determine what to teach or how to teach. Some may describe it as the assessments drive the instruction or drive their planning. The assessment data provides a basis for changing or improving instruction*	Teachers "I find them very useful in driving my instruction. I find the data very helpful in determining the "next step." "I find them very useful when I plan. The data helps me guide my instruction. It also serves as a tool to show the parents the various areas where their children are doing well or need more help. "I find them very useful! I use the assessment data to monitor progress and then adjust and plan instruction to meet the needs of my students according to their performance results. It helps me plan for instruction during class time as well as provide meaningful intervention beyond the bells." Special Education Teachers "The assessments have helped me significantly in monitoring student progress and planning strategies to help [students] reach grade level standards." Coaches "The data generated from the Skills Assessments is very helpful for teachers to plan IWT instruction based upon specific student needs (horizontal analysis). "These assessments help us recognize the effectiveness of our instructional strategies. They allow us to see areas of weakness or confusion, so we can change our delivery of practices in order to assure understanding of skills being taught." Principals "Assessments are good to drive instruction and guide teachers to make flexible groupings to address individual needs." "Extremely useful in analyzing data to drive instruction and best practices. It also is helpful in organizing student groups for IWT and teacher flex groupings."
Positive General Comment *This code is used for a general positive comment. These comments indicate that the respondent generally likes the assessments.*	Teachers "I believe that they are very useful." "I feel the assessments are VERY useful! I happen to have a current Reading Specialist credential and find this information essential." Coaches "They are extremely useful." Principals "They are excellent summative tools. The data reports are informative to the teachers."

Identify Student Needs *The assessments are useful for identifying student needs on an individual or whole class level. The assessments tell the teacher which students need more instruction or support in specific skills or units of study. The assessments help the teacher target or pinpoint specific skills for students.*	Teachers "Informative. Assist in targeting students who need extra help to strengthen particular reading skills in small group work (IWT)." "The SCOE Assessments are a great tool for teacher to use, to see what needs to be re-taught during Workshop." "Data from the assessments are useful to guide your teaching. Also, to see the areas of need and move students from intensive to strategic and to benchmark." Special Education Teachers "I think they are necessary. It helps teacher guide their instructions and also which student needs extra help." Coaches "They provide additional data to inform instruction, and meet the needs of individual students." Principals "Excellent resource for teachers to assess students needs and guide teachers instruction." "Excellent at informing teachers of student's weaknesses in order to guide differentiated or changes in instructional practice."
Intervention In Small Groups *The assessment data provide important information for forming intervention groups and determining what to re-teach in intervention time.*	Teachers "Allows me to view progress and differentiate teaching for specific reading groups." "Analysis of the results has proven useful in identifying and helping struggling students. It helps target instruction both in whole group and in small groups." "Assessments drive my focus groups." Special Education Teachers "The assessments are quite useful because they allow me to plan and implement Response to Intervention in the area(s) where my students struggle." Coaches "Teachers use the results of these assessments to determine needs of students, to determine workshop small group instruction and to determine areas of their instruction they may need to refine." "Extremely useful in analyzing data to drive instruction and best practices. It also is helpful in organizing student groups for IWT and teacher flex groupings." "The data from these assessments guide instruction. Teachers study the results and form Universal Access and intervention groups based on the data." "Assessments are used to guide teachers in planning and providing

	differentiated instruction during workshop and other intervention periods."
Align Assessment with Skills Taught *Includes comments that indicate the assessments do not align with what is taught in the instructional program. May state that skills included in the assessments are not taught in the unit of study covered by the tests, there is a mismatch between what is tested and what is taught, or skills tested have not yet been taught.*	Teachers "I find that some parts of the SCOE assessment are not directly correlated with what the students are learning. This is unfair for students and teachers." "I have never felt that SCOE assessments are properly aligned with what and how we teach. They are more difficult and tricky for the students." Special Education Teachers "The SCOE comprehension test theme and vocabulary are related to the theme that is about to be taught not the one that has just been taught." "I find that the SCOE tests are not 100% aligned to the instructional manual. In addition, some of the questions on the test can be misinterpreted by the student and in result could cause students to answer questions incorrectly." Coaches "The SCOE should better align with what is being taught in each theme. This has been a common complaint across the grade levels." "I believe this is a comprehensive assessment but at times there will be items tested in SCOE that have not been introduced yet in the program." Principal "There needs to be more alignment of the skills taught during the theme and the skills tested on the SCOE." "My teachers often complain that they [assessments] are not written to truly match the way Houghton Mifflin teaches a skill."
Suggestion for Test Format or Procedures *Includes comments about the test format, such as font size, how the directions are worded, too few questions per skill, how the students write answers, etc.*	Teachers "They can improve [the assessment] by being more specific with the instructions." "The directions for the Checking Skills sections are too complicated and do not really assess knowledge but following directions." "Checking Skills is structured so differently that many children are utterly confused by the directions, then different directions, then a third style of direction, all in the same subtest, i.e. during one part they are asked to underline verbs, another rewrite sentences in the past tense, and perhaps, underline the "action words." Too varied, especially for ELD students." Special Education Teachers "I do not always agree with the way the Checking Skills is laid out. I feel students should either get partial credit for answers or the

	questions need to be broken down for more accurate data." "Multiple choice questions are not accurate for first grade (lucky guessing and limited vocabulary)." Coaches "There are parts of the tests that generate great frustration on the part of the teachers because they feel the reading selections / questions are unnecessarily confusing for students." "I appreciate the SCOE tests, but they are very flawed in places. Sometimes the format of the assessments interferes with truly finding out what students actually know. For example, in third grade there are questions that require two steps. It is often hard to tell if the student failed due to not comprehending the concept, or not understanding the directions." Principals "The format is not aligned to the standardized testing format." "I believe the directions for the "checking skills" portion of the assessment it too difficult for the students to read independently. Also, it is more an exercise of following the directions."
Align Assessment with Standards *Indicates that assessments do not adequately align with the state standards or what is necessary to score at proficiency on CSTs.*	Teachers "The material included on these tests does not always correlate with the state standard being taught as the pacing charts tell us, but everything tested on SCOE is important and should be included somewhere in the school year." "6-8 Week Assessments should modify writing prompts according to 3rd. grade key standards. I believe students will benefit more if we could focus on 2 or 3 genres." "I feel that they need to be more aligned with the California State standards." " Many important standards are not covered on assessment but on the CST (compound words, prefixes, suffixes, quotation marks, and syllabication)." Special Education Teachers "The Checking skills section of the assessments are not based on standards or aligned with the CST's. There are several key standards that are never tested in the CST format, for example syllabication, quotations, etc." Coaches "It would be more helpful if the assessments were standards based and correlated more closely with the STAR. It would also be helpful if question stems mirrored the STAR test questions."

	Principals "I have a concern that the 6-8 week skills assessments are not more closely aligned with the released test questions from the CST's. For schools and district that are in Program Improvement status, it is imperative that instruction, assessment and delivery lead as directly as possible to improving student achievement aligned with standards, not programs."
Suggestion for Pacing or Timing *This category indicates a problem with the pacing or timing of the assessments. The assessments occur too frequently or not at opportune times in the school year.*	Teachers "The assessments occur too often. I would prefer an assessment each trimester." "I do not have enough time to reteach since the pacing guide is not taking into consideration testing." "6-8 weeks is too often for assessing. The SCOE should only be given three times a year." "Administering, correcting and analyzing theme tests is time and effort consuming. Theme tests should be condensed or simplified to be able to use results data more effectively." Coaches "There is little time for re-teaching with the tight pacing schedule." "I see many teachers rushing through many of the components in order to meet the pacing and assessment deadline." "First grade assessments move too fast. Week 6 is First Grade Friendly, but Week 12 is testing long vowels, and students aren't ready for them. Slow down the pace." Principals "The pacing and assessments can be difficult to schedule for all tracks to gain a full benefit. The amount of assessment that is being done sometimes interrupts instruction." "I wish we could have flexibility to adjust the pacing guide. In my opinion it does no good to analyze data and continue on." "In some cases the 6-8 assessment is too late. Teachers have to assess the students almost immediately. Waiting two months can let students fall behind."
Suggestion regarding Test Difficulty *The assessments test skills that are too difficult for students. Students are not at a skill level that allows them to be successful on assessments. The expectations of the assessments are beyond the students' capabilities.*	Teachers "The test tries to trick the students with too many phonetically correct answers. The test is given before students have time to practice what they have learned or given before they have learned the material." "[In the assessments], fluency stories are more advanced than the decodables and spelling skills practiced during the core phonics lessons." "For the comprehension assessments, try to choose stories the children have some background information about."

	"For the most part, they are too difficult for the majority of the students." Special Education Teachers "The end of year fluency assessment for first grade is written at a beginning of third grade level as measured by Flesch-Kincaid readability levels. Please adjust to reflect first grade." "The Fluency expected isn't reasonable for grade level, and it seems to be written at a higher grade level." Coaches "I think that skills assessments for certain themes in certain grade levels should be rewritten for improved correlation to grade level content and difficulty." "The fluency assessments are not grade level. Many have been leveled at one or two grades above. This makes it difficult in measuring accurate fluency levels and progress." Principals "The fluency passages are not appropriate to the grade levels. The questions are asked in a tricky manner especially in the first and second grade versions. Because each test is unrelated to any other it is difficult to monitor growth." "I believe the current assessment could be improved to better guide instruction by appropriately assessing the students. Some items are too difficult... they are more difficult than the CST items."
Fluency, Specific Suggestion *Provides a specific suggestion of the fluency part of the assessments. They may comment about the inappropriateness or appropriateness of the fluency assessments, how fluency is assessed, whether fluency should be assessed, etc.*	Teacher "The fluency section of the test is far too advanced. Children are asked to read quickly words that they are unfamiliar with. They are able to use their sounds, however that slows down the fluency." "Fluency is measured as reading rate. Fluency should include comprehension." "Fluency portion of the test seems to be 1-2 grade levels above my students and often contains content that is foreign to them, making context clues difficult or irrelevant to them since they have no previous experience to scaffold from. They would have more success decoding unfamiliar words in the text if they had more base knowledge of the subject." "Fluency passages do not coordinate with the theme and it does not reinforce the vocabulary we have studied throughout the theme." "The fluency passages are also too difficult and do not accurately assess reading fluency. They have led to an over emphasis on fast reading rather than reading for meaning."

	Special Education Teachers "I'd like to see the fluency assessments be more on the level of the anthology reading selections." "The fluency compared to the practice given in the anthology is too difficult for the students." Coaches "Many of the fluency passages and reading comprehension passages are difficult for our students because they lack the background knowledge to comprehend the text. For example, first grade has a passage on Ice Fishing, We need stories that culturally diverse students in California can relate to if we are using them for assessment." "Fluency scores also seem to be somewhat dependent on the topic. While I understand that students need to automatically recognize spelling patterns and high frequency words, I have seen students stop to ponder the meaning of what they are reading, slowing down the process. As in comprehension, scores tend to rise when the topic is one students are familiar with." "I do feel the fluency passages are not at grade level. From 1st through 6th grade they are too hard. Their readability does not match the HMR program." Principals "Also, the fluency passages have been tested out much higher than grade level, often 2 or 3 grade levels above where a student is at. It is often difficult to justify this to our teachers." "For many students, the level of difficulty of the fluency passages and comprehension texts create frustration, a sense of failure."
Comprehension, Specific Suggestion *Indicates concern with reading comprehension test format and content. For example the reading comprehension passages may include content that is out of the realm of experience or knowledge of students.*	Teachers "The Reading Comprehension passages are confusing and need to be revised so that students are actually being tested on reading comprehension." "Sometimes the Comprehension sections require too much of a leap for kids who may not have had middle class experiences." "Choose excerpts from the textbook for the Reading Comprehension or choose material that is similar to that presented in the textbook." "Comprehension tests go from easy to hard, back to easy.... It is impossible to truly measure improvement in student reading comprehension due to level of difficulty and story type (narrative/expository) varying from unit to unit. Genres should be tested after the Unit and not before the students have had a chance to experience it."

	Special Education Teachers "The ultimate goal is comprehension, yet the stories tend to be too advanced for the majority of [special education] students." Coaches "Providing a comprehension selection for students reading below grade level would be helpful to analyze students' comprehension ability while at the same time using a diagnostic to assess and instruct in decoding to help increase the reading ability of the child." "Several of the stories do not relate to the students' own life experiences. The third grade comprehension story about the newspaper boy for example, is very difficult for students to comprehend as our students have never seen a paper boy."
Vocabulary, Specific Suggestion *Provides a specific suggestion for improving the way vocabulary is tested. May indicate that vocabulary words on the test are not the right words, vocabulary is too difficult, vocabulary is not in enough depth, etc.*	Teachers "I do not like the way the vocabulary section is assessed. The students are required to do four different things on the two-page assessment and I feel that the SCOE does not assess vocabulary in a way that really measures students' understanding." "I do not like it when certain vocabulary words are used in the SCOE tests when they are not used in the story." "I don't believe that the vocabulary adequately assesses vocabulary strategies, it is more of a test on individual vocabulary words. I would revise the test to assess the student's ability to apply vocabulary strategies." Special Education Teachers "The multiple meaning sentences in the vocabulary portion of the assessment is especially difficult for second language learners." "Sometimes the vocabulary is used so sparsely that the words are difficult for the kids." Coaches "I think the vocabulary could be more aligned to the vocabulary that is in the units themselves, especially with second language learners." "The vocabulary portion is the least effective, since the students are not required to apply vocabulary strategies." "Multiple choice questions on vocabulary do not effectively measure student's knowledge of the words and their multiple meanings." Principals "Vocabulary could be presented in a format more consistent with the way the vocabulary is taught using the strategies of context clues, word structure, and apposition."

	"The vocabulary does not always match what has been taught, high risk students may not have sufficient background knowledge and life experiences to be successful. However if the vocabulary that is specifically taught and used in stories was used, assessments could be improved."

Relative Importance of Coded Responses

The coded responses, or categories, were sorted according to the frequency with which they occurred and are listed in rank order in Table 5.8. This table provides the overall findings and subsequent tables break down and further explain findings. This rank ordering is based on the combined set of classroom teachers, special education teachers, reading coaches and principals, or "All." Rankings and percentages of responses by category are then listed for each group: teachers, special education teachers, coaches and principals. This allows the reader to compare the relative perceived importance of the responses by participant group. Note that this table depicts the relative frequency with which the codes occurred and should not be interpreted as individuals' rankings or ratings. Table 5.8 also lists the percentage of the total "codable" responses for 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 single 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 assessment used at the school, it was considered an irrelevant comment and not coded. Response categories that occurred in less than 4% of the All Respondents combined column are not reported here because they occurred with such low frequency across respondent groups that they are not considered to have sufficient weight to call a relevant finding.

In this table, we see that the order of importance, or ranking, 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. For the two highest-ranking codes, "Monitor Progress" and "Guide Instruction," two common uses of the assessments, there was agreement across groups. This is consistent with the findings reported in Table 5.1. There is some variability in the rank orders and percentages across groups beyond these two categories. Tables 5.9 and 5.10 break out these codes into two main themes.

Table 5.8: Rank Order and Percentages of Responses for Categories

Question: "What is your opinion of the 6-8 Weeks Skills Assessments (from SCOE)? How are they useful? How could they be improved?"

Response Category (Code)	All N = 13,658		Teachers N = 12,018		Special Education Teachers N = 315		Coaches N = 727		Principals N = 598	
	Rank	%	Rank	%	Rank	%	Rank	%	Rank	%
Monitor Progress	1	23	1	24	1	18	4	21	3	22
Guide Instruction	2	23	2	22	2	14	1	31	1	30
Positive General Comment	3	14	3	15	3	14	9	6	4	10
Align Assessment with Skills Taught	4	12	4	12	5	12	5	18	5	10
Suggestion for Test Format or Procedures	5	11	5	10	4	13	2	31	6	8
Identify Student Needs	6	8	6	8	8	4	10	5	8	4
Align Assessment with Standards	7	7	8	6	9	3	3	21	2	23
Suggestion for Pacing or Timing	8	7	7	7	7	5	8	6	7	5
Suggestion regarding Test Difficulty	9	5	9	6	10	3	13	3	14	1
Fluency, Specific Suggestion	10	5	10	5	12	2	6	8	9	3
Positive- Intervention in Small Groups	11	4	11	4	11	2	7	6	10	3
Comprehension, Specific Suggestion	12	4	12	4	13	2	12	4	13	1
Assessments Not Useful	13	4	13	4	6	6	14	1	11	1
Vocabulary, Specific Suggestion	14	4	14	4	14	1	11	4	12	1

Note: Rank order is based on calculated percentages. Rounding percentages to whole numbers makes some categories appear equal in percentage while not in rank. For example, Monitor Progress and Guide Instruction are both listed as 23% in the All column, but have different rankings based on the full calculated percentage.

Five of the response categories fit under the heading of "Positive Perceptions." In Table 5.9, we see consistency across respondent groups regarding the positive aspects of the assessments for monitoring progress, guiding instruction, identifying student needs and informing small group intervention. The category of "Positive General Comment" was examined for possible sub-categories of responses, but this was not possible. A response was coded with this response if the statement was a general positive comment about the assessments such as, "I like the assessments," or "they are useful" but did not

elaborate on how they were used. The frequency of this occurrence indicates a general overall positive attitude toward the assessment process. It is interesting to note that special education teachers' perceptions were similar to the other groups. Special education teachers are not tied to the use of the assessments and often use other informal and formal assessments in the Individualized Education Plan (IEP) process (see Chapter 7 for further information about special education).

Table 5.9 Positive Perceptions of the 6-8 Weeks Skills Assessments

Response Category (Code)	Teacher %	Special Education Teacher %	Coach %	Principal %
Monitor Progress	24	18	21	23
Guide Intervention	22	14	31	30
Positive General Comment	15	14	6	10
Identify Student Needs	8	4	5	4
Intervention in Small Groups	4	2	6	3

The open-ended question invited respondents to make suggestions for how to improve The 6-8 Weeks Skills Assessments. There were eight response categories coded as suggestions. Table 5.10 provides the percentage of codable responses from each respondent group that were categorized with these codes. The percentages are similar across respondent groups with two noticeable differences for coaches and principals. A higher number of coaches than teachers made suggestions to better align the content of the assessments and the skills taught in the units covered by the tests and suggestions regarding the test format or procedures (e.g., how the directions are provided or how students' record responses). A higher percentage of coaches and principals suggested that the assessments should be better aligned with grade level standards that are the focus of the CST exams. The category of suggestions regarding test difficulty, a relatively low percentage of responses, should be taken with caution. The idea of making the assessments less difficult should be balanced with aligning the assessments with state standards. California has adopted a rigorous set of grade level standards and making the assessments less difficult should not be interpreted as lowering standards for students. Rather, it is important to ensure that grade level content is taught and mastered by students and any future revisions to the assessments should take grade level standards into consideration.

Table 5.10 Suggestions for Improving the 6-8 Weeks Skills Assessments

Response Category (Code)	Teacher %	Special Education Teacher %	Coach %	Principal %
Align Assessment with Skills Taught	12	12	18	10
Suggestion for Test Format or Procedures	10	13	31	8
Align Assessment with Standards	6	3	21	23
Suggestion for Pacing or Timing	7	5	6	5
Suggestion regarding Test Difficulty	6	3	3	1
Fluency, Specific Suggestion	5	2	8	3
Comprehension, Specific Suggestion	4	2	4	1
Vocabulary, Specific Suggestion	4	1	4	1

Only one remaining response category did not fit with positive or suggestions themes, the category of "Assessments are Not Useful." It is included here because it occurred in 4% or greater of the All Respondents column. Though 6% of special education teachers' comments were coded with this category, only 4% of teachers, 1% of coaches and 1% of principals made comments that fit this category. It is likely to be a more important issue to special education teachers because they perceive that assessments linked with the general education curriculum are not appropriate for their students or do not fit with the specialized curriculum that they may use.

Conclusions

In conclusion, this chapter finds that the 6-8 Weeks Skills Assessments used in Reading First schools play an important role in monitoring student progress, planning and carrying out instruction, and providing appropriate support for struggling readers. Teachers, principals and coaches play vital roles in collecting, analyzing, interpreting and using data to ensure that the instruction is delivered in a way that best meets students' needs. Participants made several suggestions for improving the assessments. In developing future versions of the assessments, these suggestions should be considered.

References

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

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematical operations. American Educational Research Journal, 28, 617–641.

Haager, D., & Klingner, J. K. (2005). Differentiating Instruction in Inclusive Classrooms: The Special Educator's Guide. Boston: Allyn & Bacon.

U.S. Department of Education (2002). Guidance for the Reading First Program. Washington, DC: Author.

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 the English language development (ELD) of English learners (ELs). The Year 5 Reading First evaluation study investigated the impact of Reading First on ELs, including an analysis of reading achievement and the perceptions of school personnel. Based on concerns expressed by respondents in that analysis about the need to better address the English language development (ELD) of ELs, we include further investigation of this topic in the Year 6 report. In this chapter, we report EL achievement outcomes and then present findings from responses of teachers, reading coaches, and school principals to an open-ended question on the Reading First Implementation Survey specific to the impact of Reading First on ELD. It is important to note that in California (and in the Reading First program), there is no single instructional model for teaching reading and ELD to 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.

This chapter yields the following key findings:

- Achievement gains for English learners in Reading First schools are positive for grades 2, 3, 4, and 5.
- 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 almost uniformly 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 grades 4 and 5. However, the non-Reading First EL subgroups in these grades show higher growth than the corresponding EL subgroups 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, participants offered suggestions 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 Outcomes 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 are higher in the primary grades. In this Year 6 report, the percentage of ELs was 52.7% for Cohort 1, 54.8% for Cohort 2, 57.7% for Cohort 3, and 33.2% for Cohort 4. National studies and reports have consistently reported pervasive academic difficulties for ELs (e.g., National Assessment of Educational Progress) and students whose primary language is other than English are very likely to score below competency markers on academic achievement. There is a continued gap between EL and non-EL reading achievement (National Center for Education Statistics, 2007) and over 50% score in the bottom third in reading or mathematics. In California only 32% of second grade and 17% of third grade students classified as ELs scored at proficient or advanced in Reading/Language Arts testing in 2008.

The importance of early effective reading instruction for ELs is underscored by the long-range academic difficulties experienced by this group. Nationally, only 22% of Hispanic students scored at proficiency in reading in grade 12 compared to 42% for white students (National Center for Education Statistics, 2007) and the graduation rate for Hispanics is at 52% compared to the national average of 70%. In California, 85.5% of Latino students passed the most recent high school exit exam compared to the overall pass rate of 90.2%.

The Reading First program has provided guidance regarding reading instruction for ELs based on contemporary research through professional development, coaching and curricular handbooks that provide guidance for providing instructional support during reading/language arts instruction. Districts have provided ongoing professional development and coaching through the Reading First program. Coach

institutes have included strategies for supporting teachers who serve ELs. These efforts were designed to represent research-based strategies for ELs.

Though the research base regarding ELs is more limited than the reading research base in general, some elements of effective reading instruction are validated by research. Research on effective reading instruction for ELs has demonstrated the importance of explicit instruction in the foundational reading skills in the early grades just as these methods have been well documented with native English speaking children (Shanahan & Beck, 2006). However, teachers may need to adjust instruction, particularly in the areas of vocabulary and comprehension, to make it more accessible for EL students. 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). The Reading First program is the first comprehensive effort to date in California to provide instruction that relies on practices that are validated by scientific evidence.

A series of observational studies in first grade California EL classrooms found certain instructional practices to correlate 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 teaching practices included modeling, providing explicit instruction, prompting to guide student learning, making adjustments for low performers, providing explicit phonemic awareness and decoding instruction, monitoring of student performance, providing extensive vocabulary development, and using 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 (See Chapter 5 of this report). 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.

Developing English language proficiency, particularly academic language, is a critical need for California's ELs. Limited oral language proficiency in English impacts EL students' learning across curriculum areas. A research synthesis on effective reading instruction for ELs states,

> "Given the linguistic basis of developing knowledge in academic content areas, [ELs] face specific challenges to acquiring content-area knowledge, given that their academic language, and therefore achievement, lags behind that of their native English-speaking peers. It is important to distinguish academic from conversational language skills as many [ELs] who struggle academically have well-developed conversational English skills. To be successful academically, students need to develop the specialized language of academic discourse that is distinct from conversational language." (Francis, Rivera, Lesaux, Keiffer & Rivera, 2006, p. 7)

Data Sources

For this chapter of the evaluation report, we examine student achievement results to determine the extent to which the Reading First program has specifically impacted ELs in California. Additionally, to gather information about the impact of Reading First on English language development for ELs, in this Year 6 Reading First survey, teachers, coaches and principals had the opportunity to write in responses to an open-ended question, "*How has the Reading First program supported the English language development (ELD) of English learners and how could this be improved?* 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

Chapter 2 provides a full discussion of the various achievement metrics presented in this chapter. Here, 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 (percent Proficient and Above) and the average CST English language arts scale score of EL students in the grade (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.

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 2008 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, again, it is important to note that the non-Reading First schools and the Reading First schools are demographically dissimilar, 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 (See Chapter 3 for explanation of the Reading First Implementation Index, or RFII). 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 one 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.[27] 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 (2008) 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 has had a significant impact on the reading achievement of English learners. Before presenting the achievement results, we repeat three points from Chapter 2, useful in interpreting the tables and charts:

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

[27] 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.

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, 2008) 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.

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 percent Proficient and Above achievement metric and the mean CST scale score metric. This table demonstrates the difference between a school's starting score (in the year previous to entry into Reading First) and its ending score (2008), 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 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 in Reading First schools.

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	High Implementation Schools	Low Implementation Schools	All Non-Reading First Elementary Schools
Grade 2, CSTs	792	749	139	246	2135
% Proficient and Above	3.1	3.0[b]	3.1[b]	2.5[b]	2.8
Gains in Scale Score	4.6	4.6[ab]	4.9[ab]	4.1[b]	4.2
Grade 3, CSTs	794	740	138	247	2030
% Proficient and Above	1.4	0.7[ab]	1.3[ab]	0.4[b]	0.4
Gains in Scale Score	3.0	2.6[ab]	3.8[ab]	1.8[b]	1.6
Grade 4, CSTs	584[1]	542	94	176	1791
% Proficient and Above	3.5	2.1[ab]	2.1[ab]	1.8[ab]	3.0
Gains in Scale Score	4.1	3.0[ab]	3.0[ab]	2.5[ab]	3.8
Grade 5, CSTs	238[2]	210	23	80	1744
% Proficient and Above	3.1	0.9[ab]	1.4[b]	0.7[ab]	1.9
Gains in Scale Score	3.6	2.0[ab]	3.0[b]	1.4[ab]	2.5

[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 and 6 schools, thus the smaller N.
[2]The grade 5 sample includes only YIP 6 schools, thus the smaller N.

In comparing the All Reading First schools column with All Reading First schools with ELs column, we see that for grade 2 the growth is similar, but for grades 3 through 5, the EL schools had smaller gains, but still experienced statistically significant gains. Given the achievement gap historically experienced by EL students, this is notable. Furthermore, when examining the High Implementation schools column, we see the gains for high implementation schools are markedly higher than low implementation schools with regard to EL students. The Reading First program has stressed the importance of fidelity of implementation of curriculum and instructional principles. The data here illustrate how critical implementation is for schools with EL students. In general, the relationship between implementation and achievement for ELs appears to confirm that found in Chapter 2 for all students. Reading First has significantly impacted both ELs and non-ELs.

One issue that complicates interpretation of achievement findings broken down for the EL subgroup is the possibility that a substantial number of ELs are being reclassified as fluent in English. This could substantially lower the performance of the EL subgroup in higher grades relative to earlier grades since the higher grades EL subgroup no longer includes the higher performing ELs from earlier. 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

percent 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.

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

Table 6.2 and the accompanying trend-line charts show the CST results for Grade 2, YIP 6, Reading First schools. The table includes the starting and ending mean scores for grade 2 in schools that have been in the program for six years. The first column of achievement gains duplicates the "All Reading First Schools" data that is reported in Table 2.3 (p.31). The gains in the four columns headed "English Learner Students" were computed using *only* data for the EL subgroup. As in the Year 5 Report, English Learners in high implementation schools show the strongest achievement gains.

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.

Table 6.2: English Learner CST Metric, YIP = 6, Grade = 2

Years in Program: 6 Grade: 2	Reading First Schools				
		English Learner Students			
	All Reading First Schools All Students	All Reading First Schools	High Implementation Schools	Low Implementation Schools	All Non-Reading First Elementary Schools
Number of Schools	253	237	26	94	2135
% Proficient and Above					
2002	15.5	11.3	9.3	12.1	17.8
2008	35.8	28.5	34.6	26.3	34.7
Change Since Starting Year	**20.4**	**17.2[b]**	**25.3[ab]**	**14.2[ab]**	**16.8**
Mean Scale Score Per Student					
2002	300.1	293.0	290.0	294.1	304.5
2008	330.0	321.0	330.1	317.6	329.6
Change Since Starting Year	**30.0**	**28.0[ab]**	**40.1[ab]**	**23.5[b]**	**25.0**

[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.

Figures 6.2a and 6.2b show the trend-lines for the EL subgroup for grade 2 in YIP 6 schools, on the percent Proficient and Above 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 = 6, Grade = 2

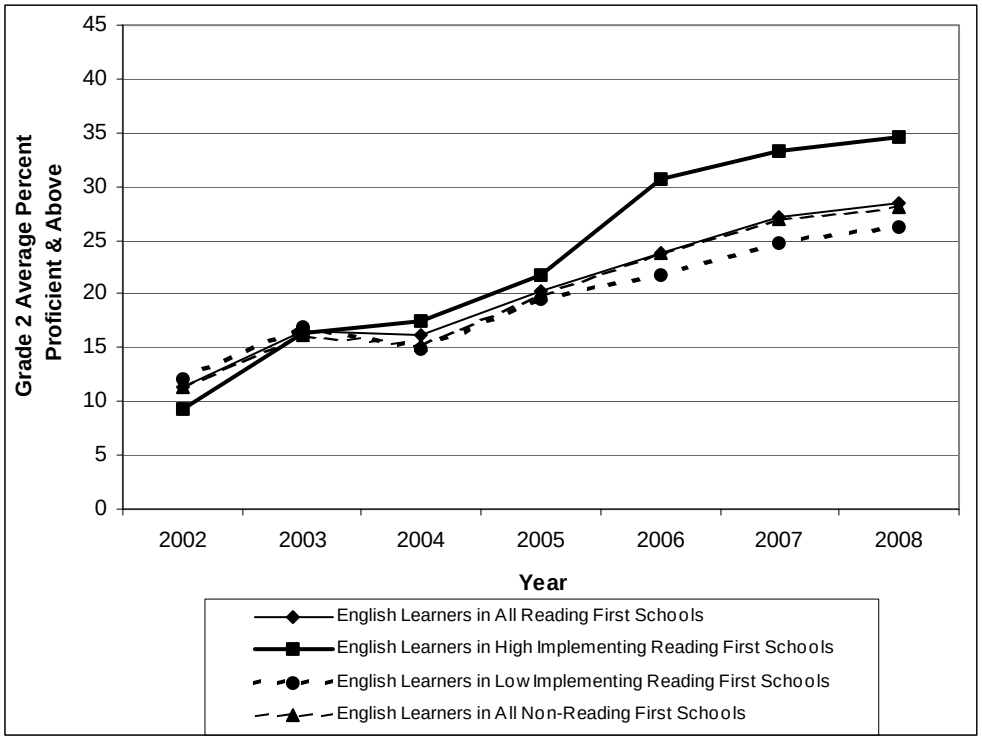
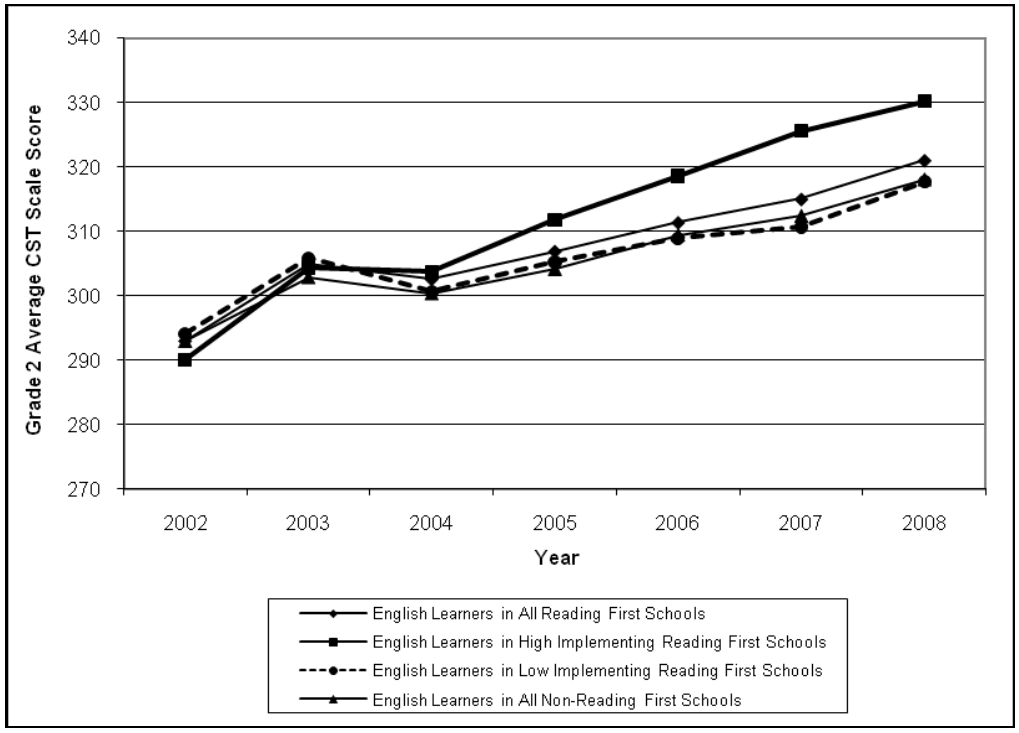


Figure 6.2b: English Learner CST Mean Scale Score, YIP = 6, Grade = 2



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 ELs (Table 6.3, 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 6 years. The first column of achievement gains duplicates the "All Reading First Schools" data that is reported in Table 2.4 (p. 35). The gains in the four columns headed "English Learner Students" were computed using *only* data for the EL subgroup.

Table 6.3: English Learner CST Metric, YIP = 6, Grade = 3

Years in Program: 6 Grade: 3	Reading First Schools				
		English Learner Students			
	All Reading First Schools All Students	All Reading First Schools	High Implementation Schools	Low Implementation Schools	All Non-Reading First Elementary Schools
Number of Schools	253	233	26	93	2030
% Proficient and Above					
2002	14.9	8.5	6.3	9.2	15.3
2008	21.7	9.3	10.3	8.0	17.9
Change Since Starting Year	**6.8**	**0.8[a]**	**4.1**	**-1.3[a]**	**2.6**
Mean Scale Score Per Student					
2002	294.9	283.7	280.4	285.1	296.7
2008	310.1	292.9	297.2	290.7	306.1
Change Since Starting Year	**15.2**	**9.2[b]**	**16.8[b]**	**5.6[ab]**	**9.4**

[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 6.8 gain in "% proficient" for all students versus only a 0.8 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 6 schools. YIP 6 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 6 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 = 6, Grade = 3

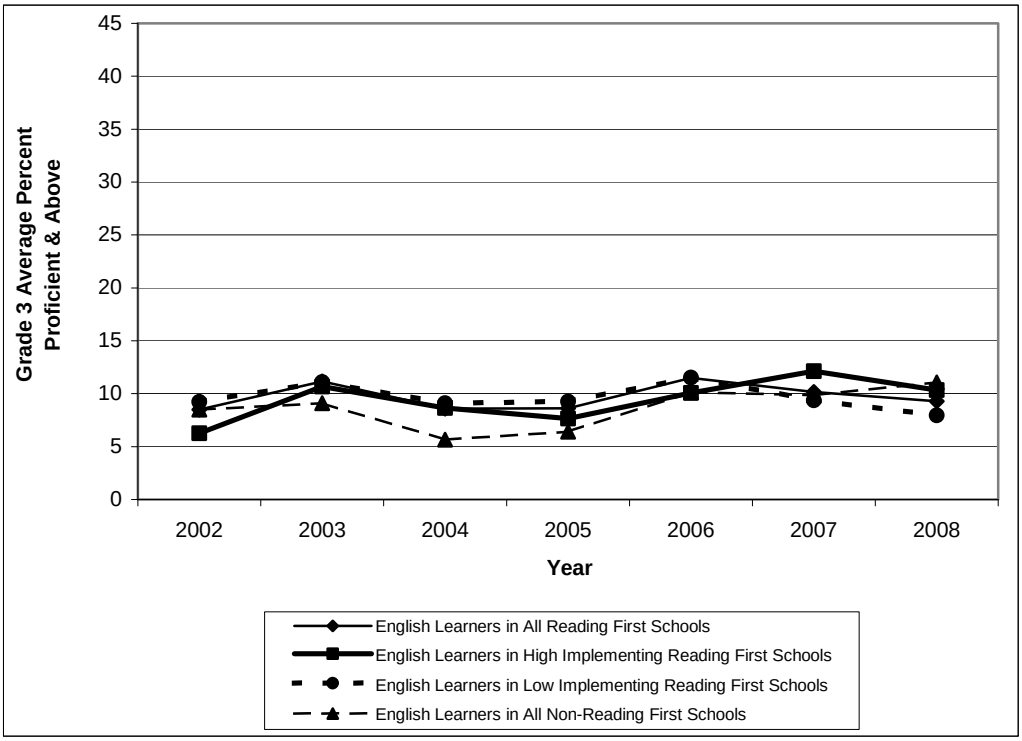
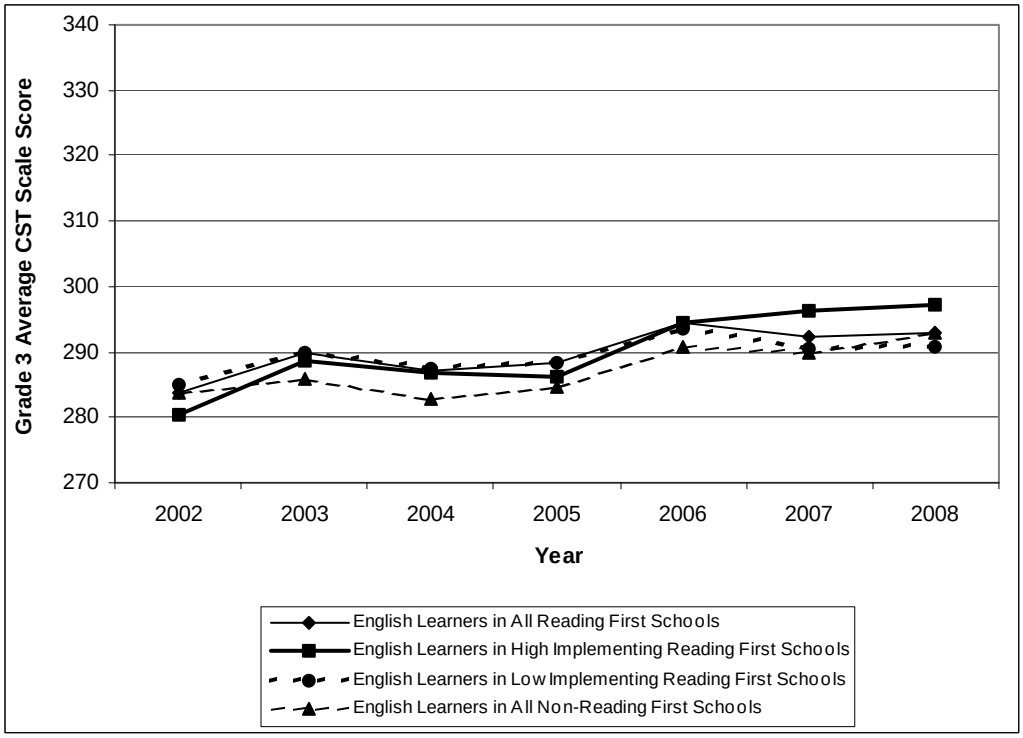


Figure 6.3b: English Learner CST Mean Scale Score, YIP = 6, Grade = 3



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 ELs (Table 6.4, 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 six 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: English Learner CST Metric, YIP = 6, Grade = 4

Years in Program: 6 Grade: 4	Reading First Schools				
		English Learner Students			
	All Reading First Schools All Students	All Reading First Schools	High Implementation Schools	Low Implementation Schools	All Non-Reading First Elementary Schools
Number of Schools	249	226	24	91	1791
% Proficient and Above					
2002	15.4	6.2	5.4	7.3	11.6
2008	37.2	17.1	18.9	15.6	29.6
Change Since Starting Year	**21.7**	**10.8[ab]**	**13.5[b]**	**8.3[ab]**	**17.9**
Mean Scale Score Per Student					
2002	307.1	292.9	288.1	295.3	303.3
2008	334.1	312.9	316.2	310.6	326.4
Change Since Starting Year	**26.9**	**20.0[ab]**	**28.1[b]**	**15.3[a]**	**23.1**

[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.

In contrast to all reading first students, the ELs in Reading First schools – even those in high implementation schools – show lower gains in the percent 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 percent 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 = 6, Grade = 4

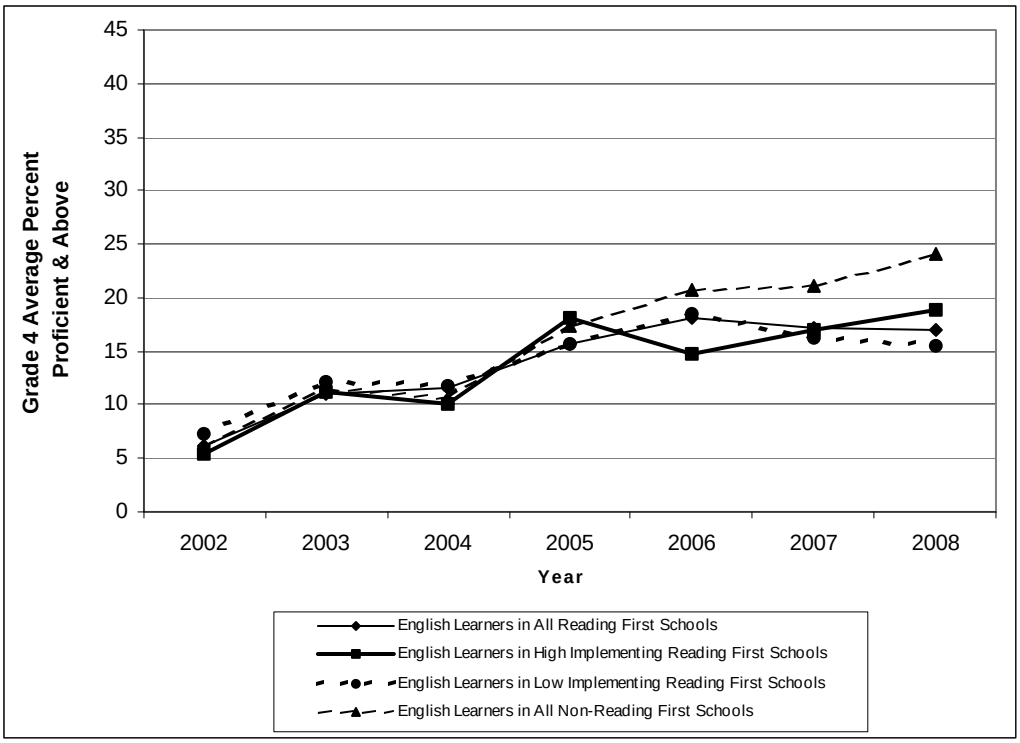
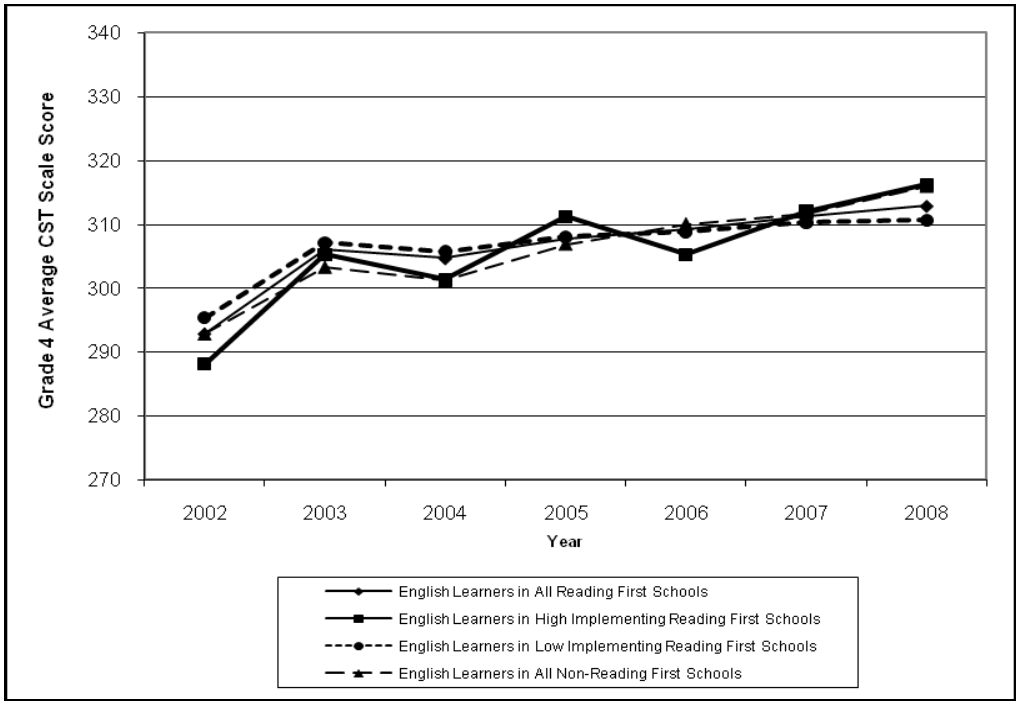


Figure 6.4b: English Learner Mean Scale Score, YIP = 6, Grade = 4



CST Results for Grade 5 ELs (Table 6.5, Figures 6.5a and 6.5b)

Table 6.5 reports the CST achievement results for grade 5 English learners in Reading First schools that have been in the program for six years. The first column of achievement gains duplicates the "All Reading First Schools" data that is reported in Table 2.6 (p. 43). The gains in the four columns headed "English Learner Students" were computed using *only* data for the EL subgroup.

Table 6.5: English Learner CST Metric, YIP = 6, Grade = 5

Years in Program: 6 **Grade: 5**	Reading First Schools				
		English Learner Students			
	All Reading First Schools All Students	All Reading First Schools	High Implementation Schools	Low Implementation Schools	All Non-Reading First Elementary Schools
Number of Schools	238	210	23	80	1744
% Proficient and Above					
2002	11.1	2.8	2.2	3.1	6.5
2008	29.9	8.5	10.7	7.4	18.0
Change Since Starting Year	**18.8**	**5.7[ab]**	**8.5[b]**	**4.3[ab]**	**11.6**
Mean Scale Score Per Student					
2002	303.4	288.6	287.7	289.4	298.6
2008	324.9	300.3	305.6	297.8	313.5
Change Since Starting Year	**21.5**	**11.7[ab]**	**17.9[b]**	**8.4[ab]**	**14.9**

[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.

As in grade 4 English learner results reported above, the grade 5 ELs in Reading First schools – even those in high implementation schools – show lower gains in the percent Proficient and Above metric than their EL counterparts at non-Reading First schools. Again, as in grade 4, the grade 5 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 5 are substantially lagging their counterparts in non-Reading First schools, extending the result which we saw above for grade 4.

Our hypothesis for this discrepancy is discussed above for grade 4: that this result is at least in part an artifact of LEA reclassification policies. However, we repeat that without data regarding the reclassification policies of Reading First LEAs, the theories are temporal and argue for follow-up data and analysis in further studies.

Figures 6.5a and 6.5b display these anomalies graphically for ELs. Note once again that an instructive comparison with their counterparts in Chapter 2 for all students, Figures 2.6a and 2.6c, show a reversal of the patterns.

Figure 6.5a: English Learner % Proficient and Above, YIP = 6, Grade = 5

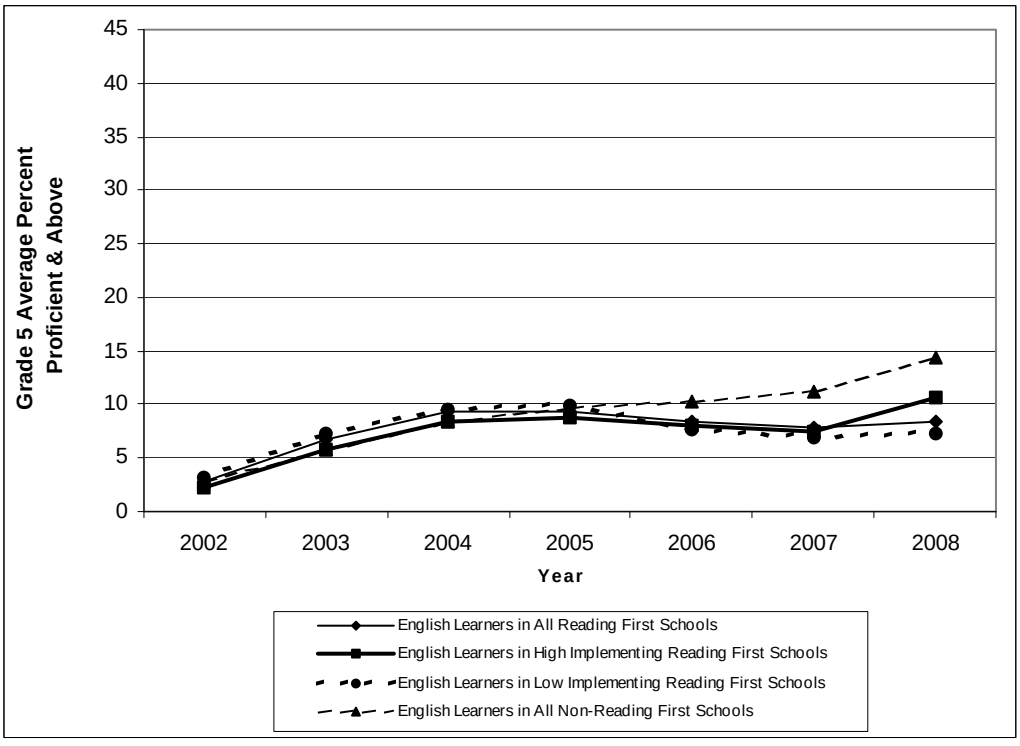
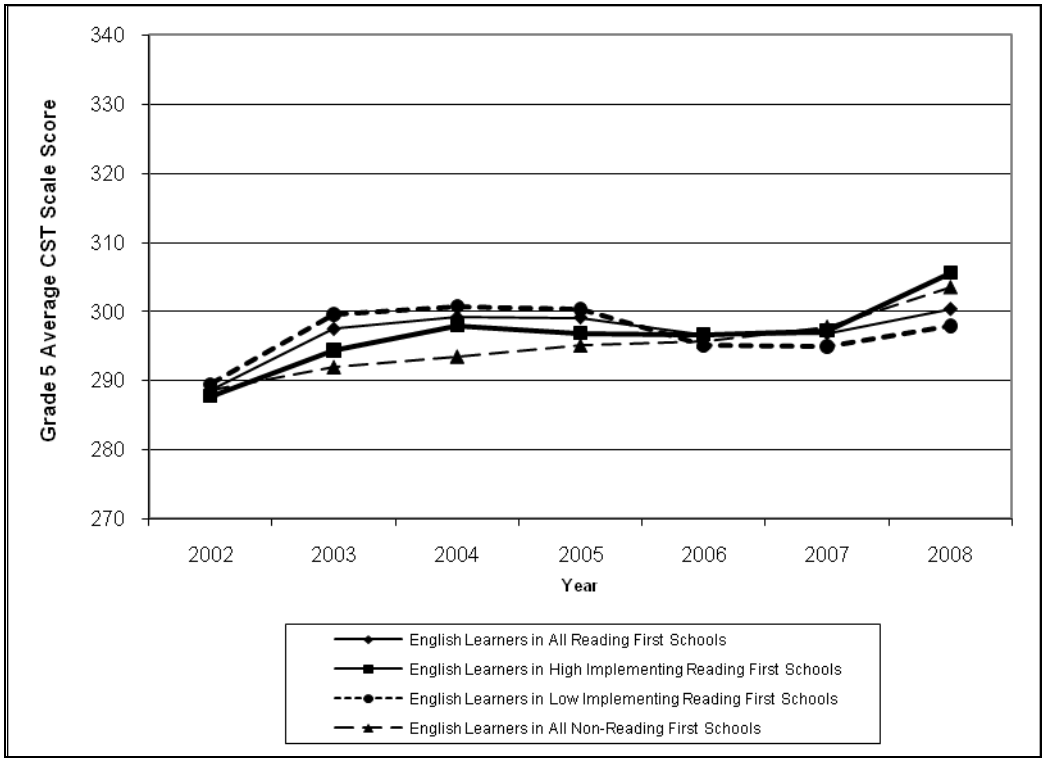


Figure 6.5b: English Learner Mean Scale Score, YIP = 6, Grade = 5



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

This section reports the qualitative analysis of an open-ended question included on the survey regarding ELD in Reading First schools. Teachers, coaches and principals responded to the question, *"How has the Reading First program supported the English language development (ELD) of English learners and how could this be improved?"* Similar to the format used in Chapter 5, in this chapter we use qualitative research methodology to examine findings from the open-ended question regarding ELD in order to better understand the perspectives of school personnel who have direct experience with this issue.

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 5 for information about the nature and benefits of qualitative methodology. Chapter 5 provides a description of the qualitative methodology and a discussion of the limits to generalizability of the findings. We refer the reader to Chapter 5 for an explanation of the coding and categorization procedures as well.

Of the 18,664 surveys collected in total, 60% of the respondents provided codable, written comments to the open-ended question regarding ELD. Of the 16,482 teacher surveys collected, 9,646 wrote narrative responses to this question, or 58.5%. Of the 468 special education teacher surveys, there were 277

comments or 59.2%. Of the 888 coach surveys collected, there were 652 narrative comments submitted, or 73.4%. Of the 886 principal surveys collected, there were 559 comments submitted, or 63.1%. These response rates are considered high given the fact that not all teachers teach EL students and the open-ended questions were voluntary.

Code Characterization

For each code, or response category, in Table 6.6 below, a brief definition is provided along with representative comments from teachers, special education teachers, coaches and principals. These are listed in the order of frequency occurring within all respondent groups combined. For each category of responses, the sample comments help to interpret the findings relative to that category. The sample comments were selected on the basis of being collectively representative of those coded with that response category.

For each response category below, the definition used by the research team to assign codes is provided along with a sampling of representative comments from teachers, special education teachers, coaches and principals. When possible, comments are provided across respondent groups. In some cases, representative comments were taken from only one or two groups (e.g., teachers) because illustrative comments could not be found within the other groups. Though every effort was made to maintain the original meaning of the quote, some quotes received minor editing- no more than a word or two - to correct faulty grammar or spelling, or to clarify context when the quote was pulled from a longer response.

Table 6.6 Response Category Descriptions and Representative Comments

Response Category Description	Representative Comments
Positive Perceptions of Materials *The RF program has had a positive impact because it has provided curriculum or materials specifically for ELD instruction. The comment is about how these materials have helped teachers to provide ELD instruction.*	Teachers "I have used both the EL and the Extra Support Handbooks and have found these resources to be effective with my students, who are primarily English Language Learners. The vocabulary development, read alouds, repeated readings, and the CDs for audio repeated listening have all been effective. The Sound Spelling Cards are also visually and instructionally strong supports for EL learners." "I have observed that many EL students have come better prepared in reading and writing. The program addresses many of their needs to develop language components and also provides them with opportunities to develop comprehension skills and fluency, via, for example, I Love Reading Books, UA (Universal Access) Resource Book, Phonics Library etc." "English Language Development is supported through the ELD Handbook and also differentiated for teachers in the Teacher's Edition. Reading First provides opportunities for

	EL students to strengthen their language skills through the way [ELD] is embedded in the HM program and also during U.A. time." "I feel that Reading First has supported English Language Development here at my school. Each year I see an increase in their academic abilities within the program. I feel that they are provided with enough support to be successful in the program because we are supplied with so much additional materials for them." <u>Special Education Teachers</u> Reading First has supported English Language Development by providing more support through materials and trainings for staff development. All this additional support helps the students and teachers." <u>Coaches</u> "RF has provided the English Language Development Guide and English Language Support Guide (ELSG). These materials are making the curriculum assessable to English Language Learners. Materials can be used for instruction of all students learning Standard English." <u>Principals</u> "Reading First has provided support for English Language Development, because the Into English and OCR programs are closely aligned. The other components, such as ELSG, Intervention Guide, and ELD Support Guide helps teachers meet the unique needs of English Learners." "EL students are supported through the Reading First Program through specific materials, principal, reading coach and assistant principal trainings/institutes, and direct instruction used on a daily basis."
Positive General Statement *Indicates a positive perception or positive regard for how Reading First has impacted ELs and supported ELD. States that the program is effective but does not elaborate.*	<u>Teachers</u> "The core Language Arts program does an admirable job of providing support for English Language Learners." "I think the program is helpful. Students who are second language learners need a lot of practice in using the language and by making sure they understand the meaning of words and the way to use the vocabulary. It will help them with their comprehension." <u>Coaches</u> "ELD has been heavily supported both by RF and by the district itself." "English Language Development has been supported again through the strong Reading First belief that all students can achieve. The Reading First program recognizes the special complexities that learning English can offer. Students have many opportunities to achieve."

	Principals "I believe that Reading First has helped our EL students succeed with its ELD component. Teachers are very happy with the program. All we need to do is to show fidelity to the program, and with that we should improve." "It has exposed EL students to a lot of language that they would not receive in pull-out ESL programs or English Language Acquisition Reading Programs."
Positive Perceptions of Instructional Strategies *The RF program has equipped teachers with more or better instructional strategies for teaching ELD.*	Teachers "We group our EL learners daily as a second grade team. I use the ideas set forth in the English Language Learners Handbook almost exclusively for that time period. I like the way that the lessons tie in with the stories in the theme. I particularly like the multi-level response ideas so that I can differentiate my instruction." "Reading First has given us many meeting/collaboration times where teachers can discuss what is helpful and what is not, as well as taught us new techniques and methodologies to teach all students." "I like the focus on learning the basics of English reading in a structured way, via discrete skills. It helps me see how children with a particular language background tend to have the same difficulties in learning English, helping me to target instruction to those children in different ways (i.e. Chinese speaking children tend to learn words as sight words, and often have trouble using rhyming word families as well as using letters to form changing words, as their language is picture-based)." Coaches "Reading First has helped ELs by providing teachers with specific Direct Instruction strategies and training on differentiated instruction. This has led to grouping students during their ELD time into instructional level groups based on their English Proficiency, including an Academic Language strand for the EOs during this block." "It has been a wonderful support for our EL students because it gives us the opportunity to look at the data and have a very clear understanding of who needs extra support and exactly what area or areas that support is needed in. It truly does guide our teachers' instruction. The program gives the teachers strategies that they can use to pre-teach or reteach the students. Our EL students in first grade are outperforming our EO students in all areas. Our EL students in second grade are outperforming our EO students in the support the areas of fluency and vocabulary. I am sure that really analyzing our data and letting it guide our instruction has made that difference."

	Principals "Teachers are using the ELD Practicum strategies and embedding them into OCR instruction. Applying these effective strategies, i.e.: communication guide, think-pair-share, thinking maps, etc. into the core curriculum instruction has improved the skills of English learners. Teachers work together weekly during Reading First grade level collaboration meetings on how to effectively embed these strategies into the curriculum."
Positive Perceptions of Training *States that the training or professional development that teachers, coaches or administrators received was helpful or beneficial for ELs. Participants learned how to teach ELD in the training.*	Teachers "Reading First has supported ELD learners by providing training for our teachers. Some of us attended the EL training in the fall." "As a Lectura teacher, I have been very satisfied with the additional material provided (such as the EL Handbook) and the one week collaboration training at the beginning of the year which allowed me to get ideas from other veteran teachers. PIE instruction has also helped my ELs greatly." "Making ELD a part of our entire school day is important in our school. The Reading First summer institute was extremely helpful to our teachers, last summer. We also have an ELD coach who works hard to help students during ELD time." Special Education Teachers "I have participated in the Frontloading training which has assisted me in better implementing the core and meeting the needs of my ELs. Reading First provides training, materials, and coaches which help me with to have the skills and materials necessary to meet the needs of my English learners." "Reading First has supported English Language Development by providing more support through materials and trainings. All this additional support helps the students and teachers." Coaches "During grade level Content Focus Coaching Professional Development, teachers learn new strategies and collaborate to ensure the needs of our ELD students do not go unmet. Teachers plan unit openers together to help build students' background knowledge. Teachers need more practice looking at ELs' writing and monitoring their learning in order to meet their needs during IWT." "Reading First has been a great part of professional development that has built teacher practices through planning and collaboration. Grade Level meetings, Lesson Study, and focused classroom observations have led to consistent and effective delivery of instruction in all grades (K-5th)."

	Principals "Last summer the majority of our teachers and I (the principal) attended Reading First's EL Institute. Teachers' increased awareness of the ELSG within their adopted text has resulted in increased usage and better service to English Language Learners. Teachers continue to need more professional development geared to providing differentiated instruction to address the diverse needs of their students. This and a review of all those (good teaching) SDAIE techniques would be helpful." "Reading First has recently provided professional development, specifically addressing vocabulary development targeting our ELs. This has assisted our teachers in helping their ELs access and make greater strides in meeting grade level content standards. The Reading First Coach also provides valuable assistance to our teachers in helping to analyze student work and plan appropriate instruction to meet their needs."
Positive Perceptions of Coach Support *Coach provides support for ELD. Positive comment about the role of coaches in helping teachers to teach EL students or teach ELD specifically*	Teachers "We have an ELD coach who supports teachers with any questions. She has made herself available to model lessons and assist us in any way possible." "As a team we have specific goals and expectations for our students as well as ourselves throughout the year. Many dialogues were happening between teachers and the reading coach. The reading coach helps train our volunteers from different organizations to reinforce specific skills that our EL and struggling readers would need in order to help them meet our academic expectations and goals." "Because of the population of our school, our Reading First coach has supported the needs of teachers to address English Language Development of the students in our classes. Through the use of the EL handbook, UA time and materials for vocabulary development she has helped us to meet the needs of our students." "Having a bilingual Reading First coach has been extremely helpful to support our English Language Development at this school. There are many strategies that he has been able to provide that are specific to the language development of our students. If a student is able to succeed in their primary language, they will succeed in their second language. He has been available to support the teachers inside and outside of the classroom. Special Education Teachers "The coach is always researching and providing new strategies for us to reach all students." "The Reading First Coach helps with techniques and materials for ELD as part of the overall literacy program."

	Coaches "Our Reading First Coaches have had extensive training in working with English Language Learners. Therefore, we are able to support the teachers in this area as well. I feel the most valuable aspect of Reading First is the coaches. We are highly trained and are able to build capacity in literacy while supporting our ELs which is 90% or our population." Principals "Reading First has supported our literacy coach. She has had the biggest impact on ELD instruction in the classroom. HM has some effective components but it does not provide enough depth with language acquisition strategies. The materials are adaptable and scaffolded for EL students. Use of assessment data and collaborative planning has been valuable to target EL instruction, as well as UA and Core Plus." "The coach is the most important piece in the implementation of the program at our school. She is very well versed in what she does to assist the teachers. EL students at the primary grades have very little or no difference from any other student at our school. Teaching the program with rigor from the beginning will ensure mastery for all students."
Positive Perceptions of Vocabulary *Reading First program supports learning in vocabulary of EL students.*	Teachers "Reading First has supported our students with vocabulary development and encourages the students to read and write. It provides explicit instruction in English." "Reading First has provided the opportunity for EL learners to have the vocabulary development and support to be successful readers and writers. The more adults working on their behalf to target instruction and practice after assessments is in their best interest." "Reading First has helped to introduce students to theme related vocabulary. It's helped them to learn the alphabet letters and sounds, and through the reading aloud the students have learned how to comprehend the English language." "Reading First has given us opportunities to create small groups and differentiate using the English Learner handbook. It allows the students to learn vocabulary before reading the story to help them understand it better." (There were no comments in this category from special education teachers, coaches or principals.)

Suggestion Regarding Materials – Curriculum *Suggestion regarding different or more materials or curriculum for teaching ELD. This would include literature selections, stories, passages, worksheets, or teacher guides that are appropriate for ELs. Some state that they need more pictures or visuals to go with the curriculum so that ELs can understand the content. Others make suggestions about the cultural relevance of text.*	Teachers "Additional support could include more use of realia, visuals, narrative input charts, and the cognitive content dictionaries." "We would like to see more reading materials with tapes or CDs so that our EL students can listen in our listening center." "A reading curriculum that is culturally relevant. The curriculum needs to match the students' experiences more. The writing needs more scaffolds and materials that provide for the scaffolds. There are no good supplemental materials that augment the students' learning of the genres of writing as well as the writing process itself." Special Education Teachers "More pictures! We print up our own vocabulary pictures. It would be nice if the program had laminated cardboard pictures that represent the vocabulary words. On the back I could break down the words into morphological parts so that the cards themselves are a lesson on vocabulary development. Thick laminated cardboard would make them last longer. Vocabulary picture cards are very time-consuming otherwise." "We need curriculum that is more matched thematically to OCR. Sometimes it works together and it drives a point for our ELs and sometimes it doesn't." Coaches "The curriculum, as well as the assessments, needs to be tailored to meet the different ELD levels of the students." Principals "Additional support for English language learners is needed in the form of read-along tapes/books, vocabulary supports, decodable translations available for parent informational use (especially when parent(s) cannot read or understand English and child is unable to translate), also Decodable listening materials/tapes, etc. to help students with pronunciation." "English Language Learners need more visuals and realia when delivering lessons. The use of technology to assist with vocabulary development for ELs would also be beneficial."

Suggestions, Other *Respondent gives a suggestion for how to improve ELD support in Reading First that is not covered by other suggestion categories. This category included various suggestions.*	Teachers "Special training for parents is needed because currently they are in the dark as to exactly how to assist. Then the teacher must continuously try to contact and work with each parent on home motivation and reading skills." "More support for EL students is definitely needed. It would also be helpful to all students (and especially ELs) to have a writing program that is consistent for the whole school." "More parent involvement and more encouragement for parents to learn English." "More Oral Language opportunities are definitely needed. Children have to spend so much time just listening to Open Court instruction when, in fact, they should be doing more interacting in order to have time to practice oral language skills." Special Education Teachers "I believe EL students need to correctly know their language and I believe they need reinforcement in their home language in order to correlate their learning to their environment." "We have a supportive staff, principal, and reading coach who are always willing to support the staff. It would be helpful if we had a magnet school for our newcomers (students who speak no or very little English)." Coaches "Issues of language transfer need to be addressed with more vigor. The monitoring of language development needs to be connected to lessons taught rather the subjective observation from teachers. A language pacing guide is needed for teachers of ELs in order to avoid the end of year rush to complete levels of competency." "Please help develop, support and train us to assist in the instruction of transition from Spanish to English reading and language skills for students. Teachers desperately need the tools to make transition possible. Also help us to address Dual Immersion classrooms so that we are integrating SDAIE strategies like GLAD but not dropping vital components like routines, and Read Alouds." Principals "We just need to make sure the students are able to talk, talk, talk. They need to experience the language, both background and academic, in order to succeed." "We need to have more opportunities for kids to write and engage at a high level with the content concepts vs. trying to cover every page and component of the program."

Suggestion Regarding Instruction *Suggestion about how to improve the instructional procedures or routines. For example, may state that they need to differentiate instruction better, or reteach, or teach differently than they are currently doing*	Teachers "I believe that the program needs to offer more activities and strategies for English Language Learners. The literature should be selected in ways that will connect to English Language Learners. If students are familiar with the stories, settings, characters, and words they will comprehend the story in a deeper level." "Based on K-6 CELDT scores in Listening and Speaking, I know we need to collaborate and develop strategies for increasing student learning in these areas in each grade level." "I would like to see instruction broken down into the 5 levels of EL to make it more teacher friendly, especially for new teachers." Special Education Teachers "Providing us with strategies that will benefit English Language Learners with Learning Disabilities would be helpful." "I think it has helped but I think we need more explicit instruction and opportunities for our ELD learners to practice what they have learned." Coaches "Having an EL coach to specifically raise students' skills in English Development Language for improving ELD levels is paramount in providing the necessary opportunities for each EL student to practice speaking, listening and writing English on an academic basis regularly. The ELSG from OCR is a great beginning. Now the teachers, across the board, need to implement its linguistic patterns regularly for student success---getting over the ELD Level 3 hump." "There needs to be an intensive, small group, pull-out instruction in order for the students to keep up with the district adopted program." Principals "Helping to bring an awareness of the need of small group time, universal access time, and use of strategies and techniques that will help students to move faster." "Our English learners currently have access to the core program. Students identified for intensive intervention do not have access to materials at their assessed level of need. We need to be better at identifying specific instructional strategies to help long-term ELs meet all criteria for reclassification to fluent-English proficient. We currently do not have curriculum-embedded assessments for placement and ongoing monitoring of ELs' progress other than the yearly CELDT. We need 8-10 week assessments to plan and provide instruction based on student language proficiency and academic performance levels."

Need More Time for ELD *Response indicates that there is not enough time in the instructional period or school day to provide adequate support for ELD in the Reading/ Language arts program*	Teachers "English language learners need more TIME, just as we would if we were whisked away to take University courses in Danish, Latvian or Portuguese." "I feel that Reading First has made an attempt at helping our English Language Learners by providing the ELSG. As a teacher who primarily works with English Language Learners I find that I already implement the strategies the guide suggests. What is lacking is enough time to complete the daily lessons prescribed in our teacher's guide and provide the much needed background information our EL students need. This is especially true during the early months of the first grade pacing plan." "As a bilingual teacher, I would like to be allotted extra time within the school day to use these materials (EL Support Activities workbooks) in the bilingual classes. Sometimes it is hard to get to these materials because of all the other things we have to cover in the pacing guide." Special Education Teachers "Same as special education, they need longer to work on their assignments. The pacing plan is ridiculous. You have to do too much in a short amount of time." "We have students grouped for ELD according to CELDT levels. In my opinion, our students need more time to simply practice using grammatically correct English orally." Coaches "ELD fights for time in the daily schedule with the core curriculum and other subjects. The new adoptions, aligning ELD and ELA sound like a promising option." "Reading First has provided our teachers six years of weeklong trainings in the program. They have improved immensely over that time. It's wonderful to go into the classrooms and see the difference. I think ELs need more time to learn to read and write than other students." Principals "We need more time for ELs in school. We are always trying to find a way to give them all the time in core, all the time in ELD AND intensive time when appropriate--but there's not enough time in a day for everything."

Suggestion Regarding Vocabulary *Suggestion that the vocabulary component of the curriculum or the ELD procedures should be improved. May state specifically how to improve the vocabulary or just generally that it should be improved*	Teachers "I feel Open Court Reading should have more vocabulary development. In some of the blending lessons, there are words that students do not know or understand." "Most of the lessons in the EL handbook are too easy for the students that are at the Early Advanced level. They need more practice with the sound/spelling cards and intensive vocabulary instruction." "I liked the Open Court ELD support. When I worked with the language structure and vocabulary, I found that to be very helpful for EL students. More vocabulary instruction geared to EL students would be helpful." "We need more help using vocabulary in the selection and relating it to real life experiences. This is very hard for the EL students. They have not experienced these words or places, or characters. For example, the story of the Grimm brothers goes into Napoleon, Europe, conquering of cultures-- hard for our kids to understand. They need videos or audios to explain this vocabulary." Special Education Teachers "Our Reading First program has supported our English learners adequately. I feel that our program for English Language Development needs to be stronger in terms of addressing English language skills, vocabulary and linguistic patterns." Coaches "It would be great if the language arts program would have its own ELD component that teachers can use to provide background knowledge and vocabulary." "Additional lessons in vocabulary and comprehension would be beneficial." Principals "EL students need more opportunities to develop vocabulary and use it in stories immediately - stories and lessons sometimes require too much background explanation." "The units provide an array of stories that the English Language Learner can relate to or connect to their own lives. The EL students need a tremendous amount of exposure to quality literature and to a very structured language and vocabulary development component."

Suggestion Regarding Change Pacing *Suggestion about the pacing of the program. Suggests giving more time to ELs or making an allowance in the pacing schedule to go back and catch ELs up*	Teachers "Adjust pacing schedule to meet needs of English Language Learners. Slower pacing to develop vocabulary, small group instruction and additional support." "The pacing plan does not allow for additional instructional time for ELD students. The pacing plan also does not allow for teachers to reteach any skills students might need help in, as determined by unit testing." "We need to be allowed to adjust our pacing for the EL students, especially because we have such a high population of EL students." "It might be necessary to develop a different pacing for ELs which will allow teachers to reinforce skills and vocabulary." Special Education Teachers "The pacing guide has us moving too fast and I don't have time to spend using realia and scaffolding the lessons." "EL and special education students need a slower pacing plan and the program needs to be taught at the children's instructional and language level." Coaches "It has supported ELD very well. In some cases, students are able to achieve academically on grade level. I think the additional support needed is the pacing of lessons aligned with the adopted reading program." Principals "I believe that the pacing plan needs to be relaxed for teachers who are teaching ELD Levels 1 and 2. Those students need more time to master skill development. All of my kindergarten and 1st grade English Immersion classes have 18-20 EL students and they need more time to take tasks to mastery." "The pacing plan hinders our ability to provide additional time and opportunity for English Learners in use of language and practice or go deeper in an area/topic."

Suggestion Regarding More Training *More training and professional development in the ELD topic would be helpful or necessary*	Teachers "I still feel the need for more professional development and strategies for working with our EL students to get them up to speed in English faster." "Even though we are doing great with the ELD, we still need more structure and training to become better teachers and support English Language Learners." "I would like some training on how to integrate the components of the ELD program with the OCR program. And also be able to use it in the allotted amount of time." Special Education Teachers "We should have more trainings for special education teachers using the district program with ELs in the general education class. Principals "Teachers need continuous training opportunities for professional development in working with English learners, sharing experiences with each other in working with English learners and talking about instructional strategies that have been effective in bridging language needs through reading instruction."

Relative Importance of Factors Associated with Reading First and ELD for ELs

The research team read each response and assigned codes, or category labels as appropriate. In many instances, a single written response included more than one code because the respondent made several relevant statements in a single written response. The codes (or categories of responses) were sorted according to the frequency with which they occurred and are listed in rank order in Table 6.7. This table shows the relative importance or weight of each response category by respondent group. Rankings are listed for the whole data set combined and then for teachers, special education teachers, coaches and principals. This allows the reader to compare across participant groups. Note that this table depicts the relative frequency with which the codes occurred and should not be interpreted as ratings. Additionally, the rank order for teachers is likely to be very similar to the all participants column because there were many more teachers than special education teachers, coaches or principals.

Table 6.7: Rank Order and Percentages of Responses for Categories

Question: "How has the Reading First program supported the English language development (ELD) of English learners and how could this be improved?"

Response Category (Code)	All N = 11,134		Teachers N = 9,646		Special Education Teachers N = 277		Coaches N = 652		Principals N = 559	
	Rank	%	Rank	%	Rank	%	Rank	%	Rank	%
Positive Materials	1	18	2	18	2	21	2	22	2	21
Positive General Statement	2	18	1	18	1	24	7	11	3	17
Suggestion: Curriculum Materials	3	17	3	17	4	12	5	16	4	14
Current Practice Described	4	14	4	12	3	19	1	25	1	22
Suggestion: Other	5	10	5	10	5	10	6	12	8	8
Insufficient Support	6	9	6	10	8	5	9	7	6	13
Positive Instructional Strategies	7	9	7	8	6	6	3	18	5	14
Suggestion: Instruction	8	7	8	7	12	2	8	10	9	8
Positive Training	9	5	11	4	9	4	4	18	7	8
Need More Time for ELD	10	5	9	5	10	4	11	3	11	2
Suggestion: Vocabulary	11	4	10	4	11	3	10	4	10	4
Don't Know/Not Sure	12	4	12	4	7	6	12	1	12	0

Note: Rank order is based on calculated percentages. Rounding percentages to whole numbers makes some categories appear equal in percentage while not in rank. For example, Positive Materials and Positive General Statement are both listed as 18% in the All column, but have different rankings based on the full calculated percentage.

There were six response categories that fit under the heading of "Positive Perceptions." In Table 6.8, we see consistency across respondent groups regarding positive perceptions of the impact of the curriculum materials on ELD for EL students. The category "Positive General Statement" did not yield subcategories or information about exactly what was perceived as positive. Comments coded into this category were short, general positive statements such as "Yes, it (Reading First) has had a positive impact," or "It has helped," but did not elaborate on the specific nature of the positive impact. The Positive Instructional

Strategies category indicating that the instructional strategies included in Reading First had a positive impact on ELD was stronger for coaches and principals than for either teacher group.

Table 6.8 Positive Perceptions of the Impact of Reading First on English Language Development for English Learners

Response Category (Code)	Teacher %	Special Education Teacher %	Coach %	Principal %
Positive Materials	18	21	22	21
Positive General Statement	18	11	17	24
Positive Instructional Strategies	8	6	18	14
Positive Training	4	4	18	8
Coach Support	3	4	4	6
Positive Vocabulary	2	0	0	0

Another theme was those responses that fit under the heading of "Suggestions for Improvement." Respondents were specifically asked to comment on how the support of the Reading First program for ELD could be improved, so it is not surprising that eight response categories provided suggestions. Table 6.9 presents the percentage of comments from teachers, special education teachers, coaches and principals that fit into each category. Note that more elaborate descriptions of the categories and sample comments are provided in table 6.11. In Table 6.9, the weight of responses is similar across respondent groups for each category with only two exceptions. Coaches provided more suggestions regarding improving instruction than other groups while few special education teachers commented on this topic. Coaches also provided more suggestions regarding collaborative lesson planning.

Table 6.9 Suggestions Regarding Improving Reading First Support for English Language Development for English Learners

Response Category (Code)	Teacher %	Special Education Teacher %	Coach %	Principal %
Suggestion: Materials-Curriculum	17	12	16	14
Suggestion: Other	10	10	12	8
Suggestion: Instruction	7	2	10	8
Need More Time for ELD	5	4	3	2
Suggestion: Vocabulary	4	3	4	4
Suggestion: Change Pacing	3	1	1	1
Suggestion: More Training	2	1	0	7
Suggestion: Collaborative Planning	1	1	10	0

Five other response categories did not fit into the Positive Perceptions or Suggestions themes. Labeled as "Other Perceptions," these categories are listed in Table 6.10 below and further described in Table 6.11.

Many respondents provided a description of their current ELD practices or models used at their school sites as can be seen below. Teachers and principals were especially concerned that the support they did receive was insufficient to meet the needs of their students.

Table 6.10 Other Perceptions Regarding the Impact of Reading First on English Language Development for English Learner

Response Category (Code)	Teacher %	Special Education Teacher %	Coach %	Principal %
Current Practice Described	12	19	25	22
Insufficient Support	10	5	7	13
Don't Know/Not Sure	4	6	1	0
No Impact/Negative Impact	3	3	2	2
Other Relevant Comments	2	4	5	5

Conclusions

In conclusion, this chapter finds that achievement gains for English learners in Reading First schools are positive for grades 2, 3, 4 and 5. 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 grades 4 and 5. However, the non-Reading First EL subgroups in these two grades show higher growth than the corresponding EL subgroups in Reading First schools. We hypothesize that this trend 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, participants offered suggestions 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.

Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). Practical Guidelines for the Education of English Language Learners: Research-Based Recommendations for Instruction and Academic Interventions. (Under cooperative agreement grant S283B050034 for U.S. Department of Education). Portsmouth, NH: RMC Research Corporation, Center on Instruction. Available online at http://www.centeroninstruction.org/files/ELL1- Interventions.pdf

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.

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.

Shanahan, T., & Beck, I. (2006). Effective Literacy Teaching for English-Language Learners. Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth (pp. 415-488). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

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: The Impact of Reading First on Special Education and Reading Intervention Programs

The purpose of this chapter is to examine the impact of the Reading First program on special education programs for students with disabilities and reading intervention programs for students at risk for identification for special education. According to the national Reading First guidelines, districts are encouraged to include special education teachers in their implementation of Reading First. At a minimum, districts with Reading First funding must make the same professional development provided to K-3 teachers available to special education teachers in K-12. Additionally, the Reading First national guidelines support reading intervention for students showing early signs of struggling with reading development.

In past reports of this evaluation study, special education teacher survey responses were included with those of K-3 classroom teachers from each school. In order to better understand the role of special education teachers in the Reading First initiative, special education teachers completed the Reading First Special Education Teacher Survey in 2008. This survey included items from the Reading First Teacher Survey (reported in Chapters 3, 4 and 5 of this report) if those items were applicable to their role as special education teachers. In addition, items were added to gather more information about how the Reading First program has been integrated into their experiences and responsibilities as special education teachers.

This chapter also includes an analysis of responses to two open-ended questions from the teacher, special education teacher, coach and principal surveys regarding issues related to special education. One question examines participants' perceptions about the impact of Reading First on special education programs in their school. The second question examines participants' perceptions about their schools' efforts to implement response-to-intervention (RTI), that is, a program designed to identify struggling students in the early grades and provide appropriate intervention, as outlined in the 2004 reauthorization of the Individuals with Disabilities Education Act (IDEA 2004).

This chapter yields the following key findings:

- Some special education teachers have participated in professional development provided through the Reading First program and have found it effective in preparing them to teach their students using the core instructional materials.
- Over 50% of the special education teachers reported using their district's adopted reading/language arts curriculum for the majority of their instruction, while others use alternative curriculum materials or partially use the core materials.

- Some special education teachers reported that they had access to coaching, but the coaches may need to continue to develop expertise regarding the needs of students with disabilities and effective instructional strategies for this population.
- Despite legal mandates for inclusion and improving access to the general education environment for students with disabilities, this analysis suggests that communication barriers continue to exist regarding students with disabilities and their participation in grade-level reading/language arts curriculum. Some participants reported that they were not aware of how the Reading First program has impacted special education teachers and students while others reported a generally positive impact but did not elaborate.
- There were mixed evaluations of the benefits of Reading First for special education. Negative perceptions seemed to focus on the difficulty of using grade-level curriculum with students whose skills are far below grade level, while positive perceptions seemed to focus on the benefits of having professional development, collaboration, and curriculum materials for special education teachers and their students. Positive comments elaborated on specific aspects of Reading First including the curriculum and materials, professional development and assessments or about the processes of inclusion and collaborative planning.
- Many schools either have not yet begun to implement Response-to-Intervention (RTI) or are in the beginning stages of implementation. Yet, some participants reported implementation of intervention for struggling readers and were able to describe specific elements of their RTI approach.

Research on Special Education Reading Instruction

Instructional Needs of Students with Disabilities

The Reading First program promotes the use of scientifically based reading instruction and draws on a substantial knowledge base of what constitutes effective beginning reading instruction. Systematic, explicit instruction in the key foundational skills of phonemic awareness, letter sound correspondence and decoding, oral reading fluency, comprehension and vocabulary provides the backbone of the curriculum, professional development, assessment, coaching and other support structures that define the Reading First program. Many of the scientifically based instructional techniques and principles promoted in the Reading First program also apply to special education. Direct and explicit instruction in the key foundational skills of reading has been documented as critical for reading growth for students with disabilities. For students with reading-related learning disabilities, high quality instruction is important, but it must be adjusted to meet their specialized learning needs.

The majority of students with disabilities served in Reading First schools would be considered to have learning disabilities and would fall in the mild to moderate range of disabilities. Students with mild to moderate disabilities generally take the CST exam and work toward grade level standards and goals and are the focus of this chapter. Students with more severe or low-incidence disabilities are likely to have very specialized goals for reading, would require extensive modifications and support provided by a special education teacher and may take the alternative to the CST. For students with learning disabilities, daily intensive phonological awareness and decoding intervention is an essential component of reading instruction (Torgesen, Alexander, Wagner, Rashotte, Voeller & Conway, 2001). These students benefit from intensive intervention that focuses on both word-level and text-level processes (Boardman, Roberts, Vaughn, Wexler, Murray & Kosanovich, 2008; Denton, Wexler, Vaughn & Bryan, 2008; Scammacca, Roberts, Vaughn, Edmonds, Wexler, Reutebuch & Torgesen, 2007). Swanson (2001) synthesized the research base on reading instruction for students with learning disabilities and found a combination of direct instruction (including such features as explicit instructions, detailed feedback, breaking a skill down, and making the task concrete) and strategy instruction (e.g., elaborate explanations of process and purpose, modeling, teacher-student dialogue, and step by step prompts) to have the greatest effect on student outcomes.

Some students with disabilities are included in grade level instruction in the general education classroom for all or part of their reading instruction in a "resource" or "inclusion" model of special education services. Other students with disabilities receive some of their reading instruction from special education in a "pull-out" Resource Specialist Program (RSP) model of special education services while others receive all of their reading instruction in special education in a Special Day Class (SDC) model. Typically this involves using curriculum and materials at the students' instructional level, which is likely to be lower than their actual grade level.

Quality of Reading Instruction in Special Education

Though special education was created to be responsive to individual learning needs, some special education teachers do not always have the knowledge and tools to provide high quality reading instruction. Two classic observation studies documented the lack of differentiated reading instruction in special education (Moody, Vaughn, Hughes & Fischer, 2000; Vaughn, Moody & Schumm, 1998). For students with learning disabilities who by definition are behind in reading, low quality of reading instruction has a significant negative impact on student outcomes (Brownell, et al., in press; Brownell, et al., in preparation; Vaughn, Levy, Coleman & Bos, 2002). Teacher knowledge in reading, specifically knowledge of pedagogy and how children develop reading proficiency, is an important foundation for providing effective reading instruction for students with disabilities and special education teachers

typically lack access to high quality reading professional development and coursework (Brownell, et al., in press, in preparation). The professional development institutes of the Reading First program, ongoing follow-up support, and coaching are designed to extend teachers' knowledge in this regard.

In 2002, a national panel was formed to investigate the condition of special education. The report of the President's Commission on Excellence in Special Education (2002) indicated that the special education system often places too much emphasis on paperwork and procedures and too little on quality instruction. Additionally, they found a chronic lack of growth in achievement for students with disabilities. They also found that the general education and special education systems often operate separately without coordination of services for individual students. These findings highlight the importance of systematically including special education teachers in Reading First support and activities. Vaughn and colleagues found that students who are identified with reading-related learning disabilities fail to fully benefit from special education after experiencing two or more years of failure in the early grades prior to identification (Vaughn, Wanzek, Woodruff & Linan-Thompson, 2007), highlighting the importance of providing high-quality instruction both prior to and following identification.

Response to Intervention (RTI) as a Means of Prevention and Early Identification

The Individuals with Disabilities Education Improvement Act (IDEA) of 2004 (PL 108-446) included significant changes in the process of identifying students with learning disabilities and providing early intervention for students showing early signs of academic difficulties. This law specifies that school districts may now use a response-to-intervention (RTI) model for identifying students with potential learning problems and providing intervention. This law set the stage for schools to implement reading intervention procedures to not only identify potential candidates for special education, but also to circumvent reading failure for many students by providing intervention in the primary grades during the critical developmental period for acquiring reading skills.

The RTI provisions represent a significant change for schools. Under prior regulations for identifying and placing students into special education, teachers would have to recognize an achievement gap for an individual student and initiate a referral for consideration for special education. This would begin a process of testing and initiate a meeting of the IEP (Individualized Educational Program) team to determine eligibility, based on whether a significant discrepancy existed between the student's ability and actual achievement. Extensive research has documented the inherent difficulties with this discrepancy-based system. The overreliance on IQ tests and other psychological tests creates potential cultural and linguistic bias. Relying on teacher judgment to initiate the referral leads to inconsistency in the nature and severity of difficulties that would warrant a referral. Perhaps the greatest problem in the discrepancy-based system is the amount of time needed for a discrepancy to be evident. For most students in need of

special education services, the tests used do not reveal a significant enough gap between ability and achievement until the later elementary grades, leading to a "wait-to-fail" phenomenon (Vaughn & Klingner, 2007). In other words, though teachers would recognize learning difficulties in the early grades, students would not receive special education services until well after their peers had learned to read in grades K-3. The law states that districts should not require school personnel to use a discrepancy-based model for identifying learning disabilities.

The RTI approach to identifying students with learning disabilities specifies that three elements need to be in place. Schools should screen all students in the early grades to identify students experiencing difficulty with acquiring reading skills. At this point, no special education services or processes are required; the purpose is to identify potential candidates for the purpose of intervention. Once students are identified, schools should provide appropriate supplemental intervention and then measure and observe students' responses. If a student receiving intervention makes adequate gains, no disability is presumed to be present. However, a non-response would indicate a potential learning disability that would warrant further investigation by professionals and an IEP team.

Since the Reading First guidelines encourage implementation of a supplemental reading intervention program for struggling students, the RTI model aligns well with Reading First. The Year 6 survey included an open-ended question designed to gauge to what extent Reading First schools are implementing RTI and the nature of such programs.

Special Education Teacher Survey Responses

Several items on the special education teacher survey are reported in this section. Here, we report on the characteristics of special education teachers responding to the survey as well as their perceptions of the program elements of Reading First.

Special education teachers were asked about their experience with their district's adopted Reading/Language Arts program. Table 7.1 displays results from two survey questions.

Table 7.1 Special Education Teachers' Experience

A2. How many years have you been teaching your district's adopted reading/language arts program?	Percent
a. Less than 1 year	7
b. 1 year 3	7
c. 2 years	12
d. 3 years	11
e. 4 years	13
f. 5 years	15
g. 6 years or more	36

A3. How many years will you have taught in the primary grades (K-3) as of July 2008?	Percent
a. Less than 1 year	4
b. 1 year	10
c. 2 years	8
d. 3 - 5 years	24
e. 6 - 10 years	23
f. 11 - 20 years	21
g. 21 - 25 years	5
h. 26 or more years	6

In this table, we see that special education teachers have a range of experience with teaching the adopted program and teaching in the primary grades. The majority, nearly 55%, of the special education teachers had six or more of teaching experience in the primary grades, representing a fairly experienced special education teaching force. The majority also reported four or more years of experience with the reading/language arts program.

With regard to professional development, special education teachers reported on their level of preparation and their involvement in follow-up professional development. Table 7.2 presents findings from three relevant survey questions.

Table 7.2 Special Education Teachers' Professional Development

B4. How well did the Reading Professional Development Institute training prepare you to teach the district's adopted reading/language arts program?	Percent
a. Not applicable	4
b. It did not prepare me well	7
c. It prepared me adequately	41
d. It prepared me very well	11
B5. How many hours of the 80-hour follow-up to the Reading Professional Development Institute will you have completed by the end of the school year?	
a. Not applicable	12
b. Less than 20 hours	3
c. 20 - 39 hours	4
d. 40 - 59 hours	4
e. 60 - 79 hours	2
f. 80 or more hours	37
B6. If you completed at least 39 hours of follow-up, how well has it supported you in teaching your district's adopted reading/language arts program?	
a. Not applicable	16
b. It has not supported me well	6
c. It has supported me adequately	26
d. It has supported me very well	13

B7. How much professional development training in reading/language arts have you received this academic year that is not related to your district's adopted reading/language arts program?	Percent
a. None	23
b. 1 - 5 hours	23
c. 6 - 10 hours	15
d. 11 -15 hours	7
e. 16 - 20 hours	8
f. More than 20 hours	20

In Table 7.2, we see that over 50% of the special education teachers felt that the professional development prepared them adequately or very well to teach their adopted reading/language arts program. Regarding the follow-up training, 12% indicated that it did not apply to them, perhaps because they were not included in the follow-up hours or it was not available to them. At the other extreme, 37% indicated that they participated in the full 80 hours, with the rest participating in less than 80 hours. Of those who did participate in a substantial portion of the follow-up training, 43% indicated that they felt supported by the follow-up training. A final question asked about their participation in professional development that does not apply to the adopted reading/language arts program. Here we see a range of responses from no additional professional development to more than 20 hours.

Over 50% of the special education teachers reported using their district's adopted reading/language arts curriculum for the majority of their instruction. Table 7.3 shows the range of responses to this question. This finding is interesting, considering that many special education students are working far below grade level and are often requiring instruction at one or more grade levels below their actual grade level. In fact, many special education teachers use an alternative curriculum specially designed to accelerate learning for students far below grade level. The question did not ask if teachers were using materials at the students' actual grade level, so is possible that some, or even the majority, of the teachers were using the adopted materials, but at a lower grade level than their students' actual grades. Another possibility is that they may have been using the intervention component of the curriculum, designed for use with students working below grade level. A third possibility is that special education teachers may be working in an inclusive model, in which they go into the general education classroom during reading/language arts instruction and help to deliver instruction in grade level curriculum with appropriate modifications and supports. A future study could further address this issue by asking questions specifically designed to investigate the instructional models used. The open-ended questions discussed later in this chapter provide additional insight into the use of adopted curriculum materials.

Table 7.3 Special Education Teachers' Use of Curriculum Materials

F3: What percentage of your total reading/language arts instruction relies on materials from your district's adopted program?	Percent
a. 0%	4
b. 20 - 39%	6
c. 40 – 59%	11
d. 60 – 79%	20
e. 80-100%	56

Special education teachers were asked to estimate the amount of time allocated to teaching with the district's adopted reading/language arts program. Table 7.4 displays the results.

Table 7.4 Time Allocation for Adopted Reading/Language Arts Curriculum

F1: On average over the last four instructional weeks, how many minutes per day have you spent teaching the district's adopted reading/language arts program?	Percent
a. Less than 20 minutes	4
b. 20 – 39 minutes	3
c. 40-59 minutes	6
d. 60 – 79 minutes	10
e. 80 – 99 minutes	8
f. 100 – 119 minutes	7
g. 120 -139 minutes	17
h. 140 159 minutes	14
i. 160 – 179 minutes	6
j. 180 minutes or more	22

These results indicate a wide range of time allocation for the adopted program, probably representing the wide range of service delivery options for special education programs. As stated above, special education teachers work in a variety of instructional settings—self-contained, resource pull-out and inclusive push-in models—and provide individualized services for students with specific reading/language arts IEP goals. Some students may spend their entire time allocated for reading/language arts instruction with a special education teacher because that is their IEP-designated plan, while others may spend some or all of their time with a general education teacher for reading/language arts working in grade-level material. Additionally, special education teachers may be using an alternative curriculum for some or all of their students. It is reasonable to assume that these findings reflect the variety of service delivery options provided in the state.

Since special education teachers serve students across grade levels, and students with varying instructional needs and instructional levels, it is interesting to note how they responded to a question asking them about teaching split-grade combination reading/language arts levels. Table 7.5 shows that only 67% responded to this question, with most reporting teaching multiple levels of the curriculum.

Those who did not respond may have done so because they are using an alternative curriculum or because they are teaching more than two levels at once.

Table 7.5 Special Education Teachers with Split-Grade Combinations

A5: If you teach a split grade combination, are you teaching two program levels at once?	Percent
a. Yes, I teach both program levels	43
b. No, I teach the lower program level	21
c. No, I teach the higher program level	3

Reading First schools must allocate time for collaborative teacher planning to ensure that teachers work together to share ideas about implementation, work through problems or challenges in delivering their lessons, and receive collegial support. Special education research demonstrates the importance of co-planning time for general and special education teachers to plan for accommodating the specialized needs of students with disabilities. Table 7.6 shows the extent to which special education teachers reported that they were included in this Reading First collaborative planning time. Findings for special education teachers were similar to those of classroom teachers reported in Chapter 4.

Table 7.6 Responses of Special Education Teachers Regarding Collaborative Planning Time

I3: How often does your school provide time for teachers to plan collaboratively?	Percent
a. Hardly ever	15
b. Monthly	20
c. Twice monthly	21
d. Weekly	39
e. Daily	2

The pacing of reading instruction for students with disabilities is a controversial topic. Modifying the pace of instruction so that students can spend extra time on specific skills is a widely used modification for special education, but some argue that slowing down instruction for special education students leads to widening the achievement gap. How to accelerate growth for students who may need to learn more slowly is a significant dilemma for special education teachers. Two questions asked special education teachers about their school's pacing schedule for reading/language arts. Table 7.7 provides insight into how California's special education teachers may be addressing this dilemma.

Table 7.7 Special Education Teachers' Responses Regarding Pacing Schedules

D1: Does your school have a pacing schedule?	Percent
a. My school does not have a pacing schedule	3
b. My school has a pacing schedule based on the assessment schedule	29
c. My school has a pacing schedule that identifies lessons on a daily or weekly schedule and when to give assessments	64

D2: What is your role in adhering to your school's pacing schedule?	Percent
a. I must adhere to the grade level pacing schedule with my students	30
b. I am aware of the pacing schedule but I do not follow it with my students	5
c. I try to follow the pacing schedule as closely as possible, but my students cannot keep up with it	32
d. I am free to set my own pacing, based on my students' IEP goals	30

This table shows that special education teachers are aware of the pacing schedule established for their school. Additionally, it shows that 30% of the special education teachers adhere to the pacing schedule with their students and another 32% attempt to follow it but find that their students cannot keep pace with it. The pacing of instruction for students with disabilities is an issue that should be further addressed in policy and professional development.

The survey asked special education teachers about the support they receive from the school principal. Table 7.8 displays the results of this question.

Table 7.8 Special Education Teachers' Views of Principal Support

D14: In general, what level of support are you getting from your principal related to your teaching of the adopted reading/language arts program?	Percent
a. Little or no support	18
b. Adequate support	51
c. More than adequate support	27

Generally, special education teachers feel they are receiving support in Reading First schools. It is impossible to determine if this is different from non-Reading First schools, but historically, school principals often limit their involvement in special education to the administrative function of participating in IEP meetings, and may leave special education teachers to operate independently with regard to instruction.

Special education teachers answered four survey questions regarding Reading First coaching. Table 7.9 summarizes results from these questions.

Table 7.9 Special Education Teachers' Perceptions of Reading First Coaching

E1: What is your access to a reading coach?	Percent
a. I do not have access to a reading coach (skip to F1)	0
b. The coach is often unavailable	9
c. The coach is usually available	56
d. The coach seeks me out to assure that I have the support I need	29

E2: Is your coach helpful in answering questions about how to teach special education students?	Percent
a. The coach often doesn't know more than I do about how to teach my special education classroom	22
b. The coach gives general answers to questions	29
c. The coach gives specific, detailed answers that I can use	40
E3: If the coach has conducted one or more demonstration lessons for you, how helpful were they?	
a. The coach has not conducted a demonstration lesson for me	36
b. The coach's demonstrations do not help much	4
c. The coach provides adequate demonstration	29
d. The coach provides demonstrations that significantly improve my teaching	25
E8: In general, what level of support are you getting from your coach?	
a. Little or no support	14
b. Adequate support	44
c. More than adequate support	35

Supporting special education teachers may not be considered a central role of Reading First coaches, yet many special education teachers are using the adopted materials. The coaches' experience typically comes from teaching a general education classroom in the core curriculum and they may lack specialized knowledge about students with disabilities and the special education process, including the various service delivery models in use. Despite this, some special education teachers viewed coaches as valuable resources or sources of support. Question E1 indicates that coaches were perceived as available by 56% of special education teachers. Question E2 indicates that only 22% of special education teachers viewed the coach as lacking knowledge of special education instruction, while 49% found the information they received from coaches as useful. Question E3 indicates that coaches had not provided demonstrations for 36% of the special education teachers, but 29% reported the demonstrations as adequate and 25% as significantly improving their teaching. Question E8 indicates that nearly 80% of special education teachers found the coaches' support to be adequate or more than adequate. It is clear that in many Reading First schools, coaches and special education teachers have been working out procedures and activities to allow special education teachers to have access to the kind of coach support provided for general education classes. Often, a special education teacher is the lone special educator at a school and at most, there may be only two or three with different teaching assignments, and it is rare for them to have site-based support focused on improving their instruction. From these findings, it is clear that reading coaching has the potential to be a valuable source of support for improving special education reading instruction. As the coaching model continues in California's schools, it is important to consider how to provide coaches with the necessary expertise to support special education teachers.

Assessment is a key component of Reading First. Two questions asked special education teachers about their use of the 6-8 Weeks Skills assessments. These assessments are discussed in detail in Chapter 5, but here we look at two questions asked of special education teachers. Table 7.10 summarizes the results of these questions.

Table 7.10 Special Education Teachers' Perceptions of the 6-8 Week Skill Assessments

F8: If you assess your students in reading every six to eight weeks, which assessments do you use? Select all that apply	Percent
a. I do not assess students in reading progress every six to eight weeks (skip to Section G)	15
b. I use teacher-developed assessments that my colleagues or I have written	16
c. I use assessments that come from the publisher with the adopted program	26
d. I use the 6-8 Week Skill Assessments	51
e. I use district-developed assessments	20
f. I use assessments other than those listed above	24
F7: How do you primarily use results of the 6-8 Week Skill Assessments?	
a. I don't use the results	8
b. I use the results to monitor student progress every six to eight weeks	26
c. I use the results to guide my teaching	47

Question F8 shows that special education teachers use a variety of assessments including the 6-8 Weeks Skill Assessments (51%). Using a variety of data sources is a hallmark of special education and these findings are not unexpected. Question F7 did not give the option to select more than one response. Therefore, results showed that special education teachers use the assessments for two purposes, to monitor student progress and guide instruction, as discussed in Chapter 5.

Special education teachers were asked to rate the overall effectiveness of their school's special education program. Table 7.11 shows results of that question, with 66% reporting good or excellent.

Table 7.11 Special Education Teachers' Perceptions of Overall Effectiveness

I1: Overall, how would you rate the effectiveness of your school's special education program?	Percent
a. Poor	5
b. Fair	25
c. Good	49
d. Excellent	17

Participants' Perceptions of the Impact of Reading First on Special Education

In this section, we report findings from an open-ended question included on the teacher, special education teacher, coach and principal surveys regarding the impact of the Reading First program on special education. Participants voluntarily responded to a question focusing on the impact of Reading First, *"What impact, if any, has your school's involvement in Reading First had on special education teachers and students with disabilities?"* Similar to the format used in Chapter 5, in this chapter we use qualitative research methodology to examine findings from the open-ended question regarding special education in order to better understand the perspectives of school personnel who have direct experience with this issue. The reader is referred to Chapter 5 for information about the nature and benefits of qualitative methodology. Chapter 5 also provides a description of the qualitative methodology and a discussion of the limits to generalizability of the findings. We also refer the reader to Chapter 5 for an explanation of the coding and categorization procedures as well.

Of the 18,664 surveys collected in total, 44% of the respondents provided codable, written comments to the open-ended question regarding special education. Of the 16,482 teacher surveys collected, 6,862 wrote narrative responses to this question, or 41.6%. Of the 468 special education teacher surveys, there were 321 comments or 68.6%. Of the 888 coach surveys collected, there were 598 narrative comments submitted, or 67.3%. Of the 886 principal surveys collected, there were 515 comments submitted, or 58.1%.

Code Characterization

For each code, or response category, in Table 7.12 below, a brief definition is provided along with representative comments from the respondents. The definitions are those used by the research team to assign codes to comments. These are listed in the order of frequency occurring within all respondent groups combined. For each category of responses, the sample comments help to interpret the findings relative to that category. The sample comments were selected on the basis of being collectively representative of those coded within that response category. When possible, comments are provided across respondent groups. In some cases, representative comments were taken from only one or two groups (e.g., teachers) because illustrative comments could not be found within the other groups. Though every effort was made to maintain the original meaning of the quote, some quotes received minor editing to correct faulty grammar or spelling, or to clarify context when the quote was pulled from a longer response.

Table 7.12 Response Category Descriptions and Representative Comments

Response Category/ Description	Representative Comments
Don't Know/Not Sure *Respondent is not aware of the impact of RF on special education teachers and students, or not sure how RF relates to special educatio*n	Teachers "We have special education classes at our school, but I am not involved in those classes." "We are unaware of the true implication or impact of RF in the area of special Ed." "Not sure how this impacts our special education teachers or their students." Special Education Teachers "I don't know if there has been an impact." Coaches "I don't know if there has been a direct impact." Principals "not sure at this time"
Positive General Perception *General positive statement about the impact of RF on special education teachers or students*	Teachers "I feel that my students have really benefited from Reading first. I feel that they are truly prepared for the next grade level." "Based on the results discussed at the meetings, students with disabilities are improving significantly using the Reading First approach." Special Education Teachers "I have greatly benefited in my teaching from the Reading First Program. I have found the reading coach especially helpful." "I have been given many new strategies and ideas. I enjoy having a reading coach to be there for me." Coaches "I feel it has helped us better help all students reach a level of reading that is optimum for their ability." "It has heightened the awareness of what students need to learn to become successful. It has also deepened the level or professional knowledge on the part of the teacher." Principals "A very positive effect in K-3. We should see growth on the CST this year." "Better outcomes and more effective implementation."

No Impact *Reading first has had no impact on the special ed. classroom or students with disabilities.*	Teachers "I don't think the program has any impact on special education teachers and students with disabilities." "I don't think that Reading First adequately addresses the needs of special needs students." Special Education Teachers "Not helpful for SDC [Special Day Class] students" Coaches "I am not aware of any specific impact, positive or negative." Principals "No real impact on these students." "The program has had little impact on special education teachers and students with disabilities."
Negative General Perception *Response is a generally negative comment regarding Reading First's impact on special education and students w/ disabilities.*	Teachers "Our resource students are not getting the extra personalized instruction which they used to get and this is creating a bigger gap in their educational potential." "I feel that this increase in the number of additional students in the classroom makes learning and teaching more difficult. The special education students are also missing out on the individualized instruction they could be receiving in their own classroom." "I do not believe that Reading First should be a program for special education teachers and students with disabilities. These students have special needs that are not accommodated by Open Court." Special Education Teachers "The Reading First program has not had much benefit for my class. They should be involved in reading programs that have been developed and have been proven to be effective for special education like the Corrective Reading Program developed by SRA. Open Court and Voyager are not effective." "Reading First has put impossible demands on me and the special population I serve." "As a special education teacher, I am still looking forward to the day when our adopted reading program caters to the needs of our students with disabilities. I spend most of my time planning, modifying lessons and implementing my own teaching materials to support the program."

	Coaches "There does not seem to be a clear definition on what parts of the Reading/Language Arts program teachers are responsible for implementing." "The Special Education teacher finds it extremely difficult to implement the Reading and Language Arts Program, as there are students of various grade levels in her classroom." Principals "Special Education teachers are required to implement the "Voyager" program as well as OCR. These teachers are being overwhelmed with the two programs."
Not Applicable *Respondent states the question is not applicable to him/her.*	Teachers "Any students w/IEPs that require resource get the extra support through the resource teacher. The resource teacher's support may include programs other than Reading First." "They [special education teachers] are not involved in Reading First." Special Education Teachers "I use alternative curriculum with my students." "Special education goes by the IEP goals, so we use a different curriculum in most cases." Coaches "Special Education teachers do not participate." Principals "My special education teachers do not use the HM program." "Our Special Education Department uses another curriculum."
Positive Perceptions of Curriculum or Materials *Response indicates that the special education teachers receive materials just like other teachers. May also state that this is a benefit for the students, that having the curriculum materials has had a positive impact on the students.*	Teachers "Open Court has provided excellent Language Arts material supplements, which work well with the Open Court lessons used." "We have had many great resources that have made a dramatic impact on all students, especially those who have been identified with disabilities." "I believe the extra support handbook has been beneficial. " Special Education Teachers "Special Education is now provided with the complete set of curriculum." "As a special education teacher, I have been provided with the core reading/teaching materials to provide equal access to my students."

	Coaches "Our special education teachers and students have the materials that they need." "Differentiation of instruction is better because there are so many support materials available." "I think the biggest impact has been the amount of materials available from grade level to grade level to support RSP [Resource Specialist Program] or SDC [Special Day Class] instruction." Principals "Reading First participation helped to align the curriculum for the students in Special Education as fidelity to the adopted program is the expectation."
Positive Perceptions of Professional Development *Response is a positive statement regarding Reading First's providing training or professional development for special education teachers in RF schools*.	Teachers "I think that special education teachers have benefited from the professional development and it has helped them with implementation of the reading program." "[We are] more equipped to carry out strategies to further support students in class." "Our Resource Specialist teacher uses the strategies taught at the Reading First training." Special Education Teachers "I liked being included in the training when this reading program was adopted. It was the first or second year it was adopted and I still am helped by the training." "The positive is that we all received good training and strategies and now are looking at classroom needs with more flexibility." "Reading First trainings have helped me become a better teacher. The trainings have helped me become familiar with the use of interventions and support." Coaches "They have been able to participate in Lesson Studies which allows them to observe other teachers teaching lessons. This allows them the opportunity to tweak their lessons to meet the needs of their students." "For me as a coach, the training has assisted me in communicating with the special education teachers." Principals "The teachers in special education have benefited in lesson studies and other professional development."

	"Our special education teachers are now included in more of the professional development. Reading First has now made it evident that students with disabilities must be included in the strategies that support mainstream students."
Positive Perceptions of Assessments *Response is a positive statement regarding the use of assessments and data analysis required in Reading First's in the special education classroom.*	Teachers "Analysis of data during staff meetings has shown special education scores improving." "The assessments provide more data for SST's [Student Study Team meetings, focused on possible special education referrals] based on multiple assessments, and can help identify specific needs." "I feel that it has helped them [special education teachers] to help the children reach benchmarks." Special Education Teachers "It has helped to identify reading difficulties in special education students, and develop strategies to help the student of special needs to be very successful." Coaches "I believe the progress of students with disabilities has been more regularly monitored with Reading First." "It provides special education teachers with tools for monitoring student progress." Principals "We are able to reflect on student progress and identify areas that need to be targeted as well as their alignment to their IEPs." "It has helped identify students earlier; it has helped keep some students from inappropriate referrals; it has helped track individual student progress."
Positive Perceptions of Collaboration and Planning *Response is a positive statement regarding the inclusion of the special education teacher in the Reading First grade level or collaborative team planning meetings.*	Teachers "As lesson study is part of the Reading First program, it enables teachers to collaborate with their peers, resulting in support for the student with disabilities." "Classroom teachers have better communication with special education teachers and we work together as a team to help students with special needs." "Our special education teacher works very closely with the general education teachers and this is helpful." Special Education Teachers "It has allowed for collaboration between general education and special education and opportunities for student mainstreaming."

	Coaches "The impact has allowed special education and general education teachers to collaborate in planning, share strategy instruction, differentiate instruction and communicate school and district wide goals in terms of teacher and student expectations that work on increasing student achievement." "The special education teachers have been able to collaboratively plan with the general education teachers and gotten ideas to modify instruction to meet their students' needs." Principals "All general education and special education teachers collaborate and use the same core curriculum, with a significant jump in API scores for our special needs population." "It has facilitated planning and collaboration with the regular classroom teacher."
Positive Perceptions of Inclusion *Response is a positive statement regarding Reading First's impact on inclusion into the general education classroom of special education students in RF schools.*	Teachers "A good impact, full mainstreaming and inclusion. Students model good reading habits for other students." "Being that I mainstream some of the special education students in my classroom, I find that there is a consistent frame of reference amongst all classrooms with sound spelling cards being an integral part of the language arts program." "I believe it has helped with implementing mainstreaming more effectively." Special Education Teachers "It has had a positive impact on mainstreaming." "My Kindergarten students are mainstreamed for 3 hours in a regular K classroom. This has impacted their vocabulary, writing and fluency development." Coaches "I believe that it is beneficial with one Special Day Class at our school that has students going into mainstream classrooms." "I think our involvement in Reading First has had a significant effect on our special education teachers and students. Most of our students are mainstreamed during the language arts block and they are doing incredibly well." Principals "It has enabled a connection to mainstream education, allowing for our Special Education students to receive the same quality education."

Relative Importance of Coded Responses

The coded responses, or categories, were sorted according to the frequency with which they occurred and are listed in rank order in Table 7.13. This rank ordering is based on the combined set of classroom teachers, special education teachers, reading coaches and principals, or "All." Rankings and percentages of responses by category are then listed for each group: teachers, special education teachers, coaches and principals. This allows the reader to compare the relative perceived importance of the responses by participant group. Note that this table depicts the relative frequency with which the codes occurred and should not be interpreted as individuals' rankings or ratings. Table 7.13 also lists the percentage of the total "codable" responses for 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 single 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. Response categories that occurred in less than 4% of the All respondents column are not reported here because they occurred with such low frequency across respondent groups that they are not considered to have sufficient weight to call a relevant finding.

In this table, a few trends are important to note. The fact that the "Don't Know/Not Sure" category occurred with highest frequency for classroom teachers indicates a relatively low level of awareness about the role of special education in Reading First. This did not match with the other respondent groups but is consistent with research that reports the separation between the worlds of general and special education teachers and the importance of collaboration to address the needs of students with disabilities. Despite legal mandates for inclusion and improving access to the general education environment for students with disabilities, this analysis suggests that communication barriers continue to exist regarding students with disabilities and their participation in grade-level reading/language arts curriculum.

Another interesting finding is that professional development for special education teachers was noted with higher frequency in the coaches', and principals' responses than it was for both groups of teachers' responses. "Negative General Perceptions" represents 17% of special education teacher responses (first in rank order) while "Positive General Perceptions" represented 13% (third in rank order), indicating a mixed response from special education teachers. Negative perceptions seemed to focus on the difficulty of using grade-level curriculum with students whose skills are far below grade level, while positive perceptions seemed to focus on the positive benefits of Reading First for special education teachers and their students. Other categories represented positive perceptions about specific aspects of Reading First

including the curriculum and materials, professional development and assessments or about the processes of inclusion and collaborative planning.

Table 7.13 Rank Order and Percentages of Responses for Categories

Response Category (Code)	All N = 8,296		Teachers N = 6,862		Special Education Teachers N =321		Coaches N = 598		Principals N = 515	
	Rank	%	Rank	%	Rank	%	Rank	%	Rank	%
Don't Know/Not Sure	1	39	1	47	8	7	8	7	13	3
Positive General Perception	2	10	2	9	3	13	4	10	2	13
No Impact	3	7	4	7	6	7	5	10	5	10
Negative General Perception	4	7	5	6	1	17	7	7	10	6
Not Applicable	5	6	3	7	13	3	13	5	11	5
Positive Perceptions of Curriculum and/or Materials	6	6	7	3	2	16	2	19	3	11
Positive Perceptions of Professional Development	7	4	14	2	7	7	1	20	1	20
Positive Perceptions of Assessments	8	4	6	3	14	2	9	7	12	5
Positive Perceptions of Collaborative Planning	9	4	12	2	5	8	3	12	5	10
Positive Perceptions of Inclusion	10	4	8	3	13	3	11	5	9	6

Note: Rank order is based on calculated percentages. Rounding percentages to whole numbers makes some categories appear equal in percentage while not in rank. For example, No Impact and Negative General Perception are both listed as 7% in the All column, but have different rankings based on the full calculated percentage.

Reading First Schools' Implementation of Response-to-Intervention

An additional question focused on how schools are responding to a current initiative to develop prevention and intervention RTI models. The question on teacher, special education teacher, coach and principal surveys was, *"Response-to-intervention (RTI) is an initiative to identify struggling readers early and provide intervention to prevent reading failure and unnecessary identification for special education. If your school is implementing RTI, describe the components of your school's approach and how the Reading First program has or has not supported your school's efforts."*

For this analysis, we used the same qualitative analysis methodology reported in the previous section. Of the 18,664 surveys collected in total, 44.4% of the respondents provided codable, written comments to the open-ended question regarding special education. Of the 16,482 teacher surveys collected, 5,420 wrote narrative responses to this question, or 32.9%. Of the 468 special education teacher surveys, there were 206 comments or 44.0%. Of the 888 coach surveys collected, there were 417 narrative comments submitted, or 47.0%. Of the 886 principal surveys collected, there were 382 comments submitted, or 43.1%.

Code Characterization

For each code, or response category, in Table 7.14 below, a brief definition is provided along with representative comments from the respondents. The definitions are those used by the research team to assign codes to comments. These are listed in the order of frequency occurring within all respondent groups combined. For each category of responses, the sample comments help to interpret the findings relative to that category. The sample comments were selected on the basis of being collectively representative of those coded within that response category. When possible, comments are provided across respondent groups. In some cases, representative comments were taken from only one or two groups (e.g., teachers) because illustrative comments could not be found within the other groups. Though every effort was made to maintain the original meaning of the quote, some quotes received minor editing to correct faulty grammar or spelling, or to clarify context when the quote was pulled from a longer response.

Table 7.14 Response Category Descriptions and Representative Comments

Response Category/ Description	Representative Comments
Not Implementing *States that the school is not yet implementing or is not fully implementing RTI or supplemental intervention.*	Teachers "We don't have RTI in our district." "At this time we no longer provide support due to budget cuts." "I am not aware of the RTI program at our school. I do believe that if this program were in place that it would be helpful". Special Education Teachers "I am not involved in this aspect of the Reading Intervention Program." "Our district is not yet implementing RTI." Coaches "My school is not implementing RTI." "Our school is not currently using RTI. It wouldn't be a bad idea though since there seems to be a greater focus on intervention versus prevention. It would also help tremendously in

	misidentifying students." Principals "RTI has not been implemented at our school as of yet." "We are not implementing RTI, but hope to in the future."
General Statement Regarding Reading First Support for RTI *A positive general statement that the Reading First program has supported the school's implementation or learning about RTI in some way. May state that RF aligns with RTI and provides support for the components.*	Teachers "The Reading First Program has delivered very good intervention strategies and practices as a whole at our school." "RTI at our school is pretty well implemented via our Reading Mastery program." "Funding has been provided to have an RTI program. The RTI teacher has been given a certain amount of time to work with students in smaller groups and she is very useful." "I think our school has been very involved in helping those students that are not meeting grade level through intervention." "The Reading First program supports our intervention programs by providing assessments used to determine placement in these groups, and by supplying materials and activities which are used in some of these groups." Special Education Teachers "I Feel that the Reading First program has supported my school's efforts." "My school has developed a program that helps to identify reading problems early on. I feel that Reading First has been very helpful." Coaches "Our school does use RTI for struggling readers and our Reading First program has been crucial in implementing RTI. Reading First staff have helped identify students and helped with both planning and instructing struggling readers." "Reading First has supported our RTI strategies. They are very closely connected. RTI is all about good first teaching - so is Reading First." Principals "I believe that Reading First has supported our school's efforts to succeed" "Reading First has provided teachers with the skills and materials to deliver strong instruction and intervention." "RF has helped because teachers are expected to fully implement the core with high quality instruction for ALL students as the first level of prevention."

Don't Know About RTI *Respondent does not know about RTI and what it means.*	Teachers "I don't know anything about this program, but I think some students need some form of intervention." "I do not know what RTI is. This is the first I have heard of this." "I am unaware if our school is implementing RTI." Special Education Teachers "I don't know. When students get to me they have already been identified." "I am not familiar with RTI or if it is at this school." Coaches "RTI has been mentioned at a district level, but not explained, so I don't know exactly what it is." "I am not familiar with the RTI initiative." Principals "This is the first time I have heard of this initiative." "The connection of RTI and Reading First's impact on struggling readers cannot be determined at this time."
Curriculum & Materials for Intervention *Comment states the curriculum materials that are used for intervention. Response may give the name of a specific intervention program or discuss supplements to the core program (Houghton Mifflin, Open Court) that are designed for supplemental intervention.*	Teachers "We use Academic Academy, Waterford, Kaleidoscope, Read 180, and instructional aides." "We have PALS. It has helped our struggling students with comprehension skills, fluency, and vocabulary. The students love it and have received the program quite well." "At our school we use Intervention Guide and intervention materials to help at-risk students." Special Education Teachers "My school has given me adequate materials and support in implementing our Reading First program." "We are implementing RTI with use of the Voyager Reading Program after school." "We have been using Road to the Code, Sounds and Letters, Phonemic Awareness in Young Children, and Linguistics Guide. This has been very supportive." Coaches "We use Early Reading Intervention, Kaleidoscope, Language, and support vocabulary with Elements of Reading." "Reading First helped us with the instructional materials that we use."

	"Within our Reading first program we have interventions (Voyager) and Excel time- grouping students in a grade level to give them added help or intervention as needed." Principals "We are offering early kindergarten intervention with Voyager. Reading First supports our Excel time with UA and intervention or extension." "We currently utilize the HMR phonics intervention kit and other state-approved interventions with small groups. I know that materials we use were selected because of our coaches' attendance at HMR Coaches trainings and recommendations from those trainings."
Assessment Data used for Grouping or Instruction *States that some sort of assessment provides the basis for forming intervention groups or making instructional decisions about the RTI intervention.*	Teachers "Assessments are used to identify struggling readers and pinpoint specific areas of need thus enabling the school to take best advantage of the Reading First program" "After the assessment results are reviewed, we identify the students who are having difficulty in specific areas and provide the remediation necessary." "Based on assessment results, students are identified as intensive and strategic. These students are then given the intervention help that is needed for them. This has been very effective for those students." Special Education Teachers "It helps in using the assessments to identify and quantify the particular difficulties individual students are having and target instruction." "Based on the result of the assessment, we identify students who are far below basic, basic, proficient and advanced. Then we are asked to do a grade level planning for teaching strategies focusing on how to improve their reading skills." Coaches "After the 6-8 week assessment is given, teachers are given time during a Data Analysis Day to analyze the data and adjust their teaching based on student's results. Teachers regroup their students and come up with strategies to help those students who need the extra support." "The teachers use the data from SCOE to help form the groups. They monitor the students' progress using on-going assessments." "We use data to drive our RTI. The data comes from SCOE and other benchmark district assessments."

	Principals "Assessment data helps identify students that need more support." "Our school deploys students according their performance on several assessments." "We utilize fluency and comprehension results to monitor student's progress. These two components are part of the 6-8 week assessments and help us guide our instruction to provide reading interventions to our struggling readers."
Before or After School Intervention *States that intervention occurs during before or after school programs. Students may receive tutoring or small group instruction in before- or after-school programs or Saturday school.*	Teachers "Our school offers an after school reading program to help those students that might need to improve their reading skills." "Our school has an after school as well as a Saturday intervention program. I believe many children are helped by this." "Children who are under-performing are offered after school tutoring, three days a week. This is a small group setting and children make quite a bit of progress in this setting." Special Education Teachers "Students in need of extra help in reading are given that help twice a week after school." "Many of the teachers including myself have taken those students that struggle and tutor with them after school and provide a small group setting to help the students learn these skills." Coaches "Our school identifies struggling students and allows them to participate in a 30-hour program each year. The teacher can teach intervention before or after school, or they may teach during students' off-track time." "Teachers help struggling students in small groups during workshop time or during after or before school." "We have some after-school intervention groups which help struggling readers." Principals "We do workshop and before and after school extended day to provide intervention strategies." "Students are monitored, receive extra intervention during and after school, and parents are assisted to encourage practice at home."

Instructional Strategies for RTI *Comments regarding instructional strategies for RTI.*	Teachers "I find that the re-teach, pre-teach and ELD instruction guidelines are very worthwhile." "I like the lesson studies to help improve teaching strategies on specific skills." "Reading First provides helpful instruction tips to increase students' academic success." Special Education Teachers "I do my own intervention instruction for my special education students. I use the Re-teach Guide, small group intervention, and one-one assist strategies." "Our program provides intervention and teaching strategies which encompass "best practices," researched-based teaching strategies to support the learning needs of diverse learners. Intervention strategies are provided to support students who are not proficient in language arts." Coaches "We have an additional hour of intensive intervention daily for these students that specifically targets their academic deficiencies." "Teachers provide small-group instruction to promote reading comprehension and vocabulary for an hour and implement the resource handbooks and district adopted reading program. Instruction is direct, explicit, and systematic." Principals "Students in grades 1-3 receive 45 minutes of reading intervention and the various handbooks are utilized as the primary source for the instructional delivery." "RTI has assisted teachers in planning instruction, direct instruction training and implementation, coaching and staff development for teachers to better their delivery of instruction."
Universal Access or Workshop *Describes how the Universal Access (UA) or Independent Workshop Time (IWT or workshop) is used to provide instruction for struggling readers.*	Teachers "We have a UA time in the morning with two to three teachers. Each one is assigned a specific task to teach. Other intervention occurs in the afternoon." "In first grade we also do daily Universal Access time which is very valuable." "During workshop I pull out a few students that are struggling and I either pre-teach or re-teach the subject." "Every day during Independent Work Time a teacher comes and works with my students who are struggling in reading." Special Education Teachers "My school has a block of time called Independent Work Time

	(IWT). During this time, struggling students are put into a small group for re-teaching. These students are tested in fluency. They also review reading comprehension strategies during choral reading. We give the students phoneme segmentation drills to help increase reading fluency as needed." Coaches "Everyone does IWT for 1 hour daily using RF materials." "I think our UA time has helped to identify children who are having difficulty in Language Arts. These children are either pre-taught or re-taught difficult skills. This has enabled the teacher to help students gain skills rather than asking for them to be tested for resource help." Principals "Universal Access has provided opportunities for small group instruction, re-teaching and pre-teaching of difficult concepts. The program also provides us with the data necessary to help guide and monitor our instruction." "The Reading First Program has been very supportive. RTI is built into the Universal Access Time during Language Arts."
In Process of Developing RTI *Describes that the school is in the process of developing their RTI model or intervention program. May indicate what pieces are in place, but gives the sense that they are not ready to fully implement it yet.*	Teachers "Our RTI is in the beginning stages. We have the opportunity to review and assist with the needs of our students, but only targeting small groups of students at a time to raise scores." "It is new to us and we are trying to work out all the kinks. Our grade level is planning for better efficiency next year." "It is being fully implemented in 3 rd grade and has gone to 2nd grade. Next year will be implemented in K-3rd grade." "Our school is looking into RTI and headed towards its implementation. Our principal has given us web sites for us to read about RTI and has begun to talk about it with teachers. Our Reading First program and coach have, and are ready to support RTI." Special Education Teachers "My school has started with the RTI Process but it is not fully understood yet. The grade that made the most use of it was first." "We will be starting RTI next year but we do not yet have the specifics of the program." Coaches "Intervention is occurring at our site, but not under the RTI model. I believe the district is developing an RTI model currently which will happen in the 08/09 school year."

	"The District is in the process of implementing it district-wide. Coaches, RSP teachers, psychologists, some administration have been initially trained. There are plans to train all teachers in the District." Principals "We are in the process of implementing RTI. In September 08 we will be involved in more intensive training to fully and successfully implement RTI." "We are only beginning to implement RTI. The program assessments and supplemental materials have provided some guidance and support."
Positive General Comment *Comment is a general positive statement about RTI. The comment does not include details*	Teachers "Easy to use." "Helps struggling readers." "Intervention is key to improving a love of learning and lifelong learning." Special Education Teachers "RTI provides for intervention so that students can be successful in the core." Coaches "RTI works really well at my school." Principals "Our school is using the RTI model and now will become the model for the district." "RTI works really well at my school."
Negative Comment *Comment made is negative about RTI or how Reading First is involved with it.*	Teachers "At our school site, struggling students are not given support early enough." "I do not think students that need intervention are being addressed. The teachers are expected to deal with such issues in class and continue teaching the rest of the class in the curriculum. I feel other interventions are needed from outside of the classroom." "I don't think the Reading First program is supporting this RTI initiative." Special Education Teachers "New students do not receive services. Also there are big gaps of time when students are not serviced." Coaches "The major stumbling block we are facing in implementing RTI is finding time for students to receive intervention outside of

	Language Arts and Math instructional minutes." "There are lessons in the teacher's manual that are not focused on state standards. Teachers unknowingly teach these lessons because they trust the program. This wastes precious instructional time that could be used on state standards. This interferes with RTI because the students have not had a full opportunity to learn the standards due to the time spent on the distracting lessons that don't teach the standards." Principals "Intervention components are not strong or extensive enough." "OCR does not adequately support struggling readers." "The pacing guides have become an element of negative impact when teachers feel that they MUST keep with the pacing/assessment guides and do not take the time to re-teach if necessary."
Student Study Team *The school uses a team that is already established to oversee or guide the intervention provided on a case-by-case basis. Student Study Team (SST), Student Success Team, Student Assistance Team and other names are often used to describe a team that typically manages referrals to special education.*	Teachers "Our school has an SST panel that reviews students that cannot be retained." "In my school we have a Student Success Team that we can refer students to if they are struggling. This team gives us ideas and tries to identify other ways to help the students." "In our school we have SST meetings which are very helpful to identify students' needs and obstacles to learning. These intervention meetings are helpful if teachers, parents and school staff follow up on them." Special Education Teachers "Our school identifies students who are struggling and develops strategies to assist in addressing the needs of the students through SSTs." "Our school uses the SST process to intervene with these students in addition to the skills assessments." Coaches "We analyze data regularly and use the data to refer struggling students to the Student Intervention Team. This team then makes the necessary recommendations to help the student." "Reading First has been instrumental in providing the helpful data and information needed when SST teams meet and plan next steps for identified students in an effort to meet their needs in general education." "Our school started implementing RTI this year. It is conducted during monthly grade level meetings along with a SST representative to help identify and monitor the students in need."

	Principals "Utilizing the RTI model, we successfully use teaming and the Student Study Team process, as well as supplemental materials (state approved)." "Students are referred by our Student Study Team for focused small group instruction in the core English/Language Arts program."
Coach Supports RTI *Comment indicates that coaches support teachers and classrooms with implementing RTI.*	Teachers "Having a Reading First coach is really helpful. She is there to answer questions and come up with strategies and materials to assist our specific needs when asked." "Coach is very supportive and organizes all intervention." "I work with the literacy coaches to identify struggling students and give them the extra attention needed." Special Education Teachers "I think the reading coach is doing an excellent job of identifying and working with target students." "The coach and principal are integral in RTI. From a special education viewpoint, it is an excellent way to make sure that students receive appropriate intervention before being referred directly for special education testing." "The Reading First coach has been important in helping with analysis and interventions for grade levels." Coaches "Coaches and the RF coordinator have been very instrumental in designing lesson frames and selecting materials and strategies for use during these lessons." "Reading First supports this program by having the coach train the intervention teachers in the use of Sound Spelling cards and the components of a reading lesson (phonemic awareness, phonics, blending routines, high frequency words, and sentence dictation)." Principals "Our incredible Literacy Coach is participating in the "Advanced Reading and Language Arts Leadership Program" and has organized and trained our entire teaching and paraeducator staff." "Our coach has provided support to the K-2 teachers in how to monitor and review DIBELS data and what the next steps are."

Tier 2 Supplemental Intervention Provided *Describes how the school provides intervention for struggling students during time that is supplemental to that allocated for the core program.. Students are placed into small groups and a teacher provides extra instruction beyond the core program.*	Teachers "Our students are grouped at their levels and then receive 30 minutes of focused intervention at their level." "Certificated tutors target struggling readers and work with them in small groups." "Our school provides additional reading intervention for struggling readers by allotting an additional block of time geared to target struggling readers." Special Education Teachers "Our school has a pull-out program where the students that are struggling in the regular classroom get to work in a small group setting with an intervention teacher." "Our school has implemented an afternoon tutoring program for students at risk. They go to a reading tutor for 30 minutes a day four days a week. This is small group instruction at the students level." Coaches "I am helping coordinate a Tier 2 [supplemental intervention] program that is using SRA's The Early Reading Tutor." "We identify intensive students and place them in a Core +30 intervention which is a "dip-down" approach, using program materials from a previous grade level and specific strategies to support these students." "Tier 2: Coach and three teachers have been trained on TPRI to identify struggling readers, diagnose needs and provide intervention activities." Principals "Our school uses retired teachers to provide additional small group instruction (targeted) to students who are functioning at the Below Basic level." "Tier 2, using intervention programs in small groups has garnered some success." "We have Tier 2 reading teachers who provide 30 minutes of explicit instruction to strategic students using core materials."
Grade Level Meetings & Collaboration *Grade level planning meetings help teachers work together in implementing RTI.*	Teachers "At grade level meetings, we review our students scores from SCOE. We find an area that we are struggling in, and set a goal for our grade. Then we brainstorm strategies that we will all use to accomplish this goal."

	"At my school site, we have weekly grade-level collaboration meetings where our grade level celebrates our successes and gives feedback on how teachers are being effective in the classroom. Also, we target specific subgroups and monitor their progress." "Our school does Grade Level Intervention Meetings. At these meeting we are able to look at our students as a whole group and decide what students need additional help in reading and what types of interventions would work best." "Our (RTI) program team meets with a grade level each week to discuss struggling readers. The team is very effective." <u>Special Education Teachers</u> "Reading First fully supports special education teachers and includes us in planning, collaboration, and materials preparation." "We identify student needs and develop strategies during grade level articulation and academic conferences based on data." <u>Coaches</u> "At all of our grade level meetings, we review/add to our RTI pyramid chart. We discuss specific students and the interventions needed to help them improve." "Our teachers are spending a lot of time on the needs of the students at collaboration meeting and deciding what students need to be placed in intervention." "Teachers plan collaboratively on a weekly basis, analyze data, and share strategies, discuss strengths and areas of need to increase student achievement." <u>Principals</u> "The Reading Coach facilitates grade level meetings where teachers discuss students who are need of RTI. Teachers collaborate in the planning of interventions." "The RTI team works collaboratively with the classroom teacher and the Reading First coach."

<u>Relative Importance of Coded Responses</u>

The coded responses, or categories, were sorted according to the frequency with which they occurred and are listed in rank order in Table 7.15. The first column lists the response categories in rank order based on the "All" column. Rankings and percentages of responses by category are then listed for each group: teachers, special education teachers, coaches and principals. This allows the reader to compare the relative perceived importance of the responses by participant group. Note that this table depicts the relative frequency with which the codes occurred and should not be interpreted as individuals' rankings or ratings. Table 7.15 also lists the percentage of the total "codable" responses for 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 single 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. Response categories that occurred in less than 4% of the All Respondents combined column are not reported here because they occurred with such low frequency across respondent groups that they are not considered to have sufficient weight to call a relevant finding.

Notable findings in this table include the high frequency of "Not Implementing," "Don't Know about RTI" and "In Process of Developing RTI" indicating that many schools are just developing an awareness of or may not have started the process of RTI. The relatively high frequency in the category "General Statement Regarding Reading First Support for RTI" also may indicate that participants have an introductory level of awareness of the concept and principles of RTI.

Yet, some participants did indicate they are implementing at least some elements of RTI. This is evidenced in comments focusing on curriculum and materials, assessment, instructional strategies, and procedures for RTI. Grade level team meetings and collaboration seem to be important to successful RTI implementation and coaches may play a valuable role in assisting with RTI. Some participants discussed support personnel who provide intervention, such as paraprofessionals or intervention teachers hired specifically for RTI, while others indicated that they provide their own intervention instruction during small-group time.

Table 7.15: Rank Order and Percentages of Responses for Categories

Response Category (Code)	All N = 6,425		Teachers N = 5,420		Coaches N = 417		Principals N = 382		SpEd Teachers N =206	
	Rank	%	Rank	%	Rank	%	Rank	%	Rank	%
Not Implementing	1	15	1	15	5	14	7	9	2	16
General Statement Regarding Reading First Support for RTI	2	11	4	10	1	21	2	17	3	12
Don't Know About RTI	3	10	2	11	15	3	17	1	1	17
Curriculum & Materials for Intervention	4	10	3	10	6	13	6	11	2	16
Assessment Data used for Grouping or Instruction	5	8	8	7	4	16	3	15	9	5
Before or After School Intervention	6	8	6	8	8	8	10	8	10	4
Instructional Strategies for RTI	7	7	7	7	7	11	5	12	12	3
Universal Access or Workshop	8	7	10	6	2	18	4	14	6	6
In Process of Developing RTI	9	6	12	5	3	16	1	21	5	7
Positive General Comment	10	6	9	7	20	2	13	4	6	6
Negative Comment	11	5	11	5	19	2	17	1	7	6
Student Study Team	12	5	13	4	9	7	8	9	8	5
Coach Supports RTI	13	4	14	4	10	6	10	6	9	5
Tier 2 Supplemental Intervention Provided	14	4	15	3	13	4	9	8	11	4
Grade Level Meetings & Collaboration	15	3	16	3	12	5	12	4	12	3

Conclusions

In conclusion, this chapter finds that the Reading First program has yielded some benefits for special education teachers and their students. Special education teachers at some schools have had access to professional development, coaching and curriculum materials through Reading First participation. Over 50% of the special education teachers reported using their district's adopted reading/language arts curriculum for the majority of their instruction, while others use alternative curriculum materials or partially use the core materials.

Despite legal mandates for inclusion and improving access to the general education environment for students with disabilities, this analysis suggests that communication barriers continue to exist regarding students with disabilities and their participation in grade-level reading/language arts curriculum. Some participants reported that they were not aware of how the Reading First program has impacted special education teachers and students while others reported a generally positive impact but did not elaborate. There were mixed evaluations of the benefits of Reading First for special education. Negative perceptions seemed to focus on the difficulty of using grade-level curriculum with students whose skills are far below grade level, while positive perceptions seemed to focus on the positive benefits of Reading First for special education teachers and their students.

Regarding RTI, many schools either have not yet begun to implement RTI or are in the beginning stages of implementation. Yet, some participants reported implementation of intervention for struggling readers and were able to describe specific elements of their RTI approach.

Based on these findings, it appears that the state would benefit from continuing to support professional development for general and special education teachers to assist them in providing high quality reading instruction that is differentiated for students with disabilities as well as for struggling readers needing intervention. Coaches may play a valuable rule in this endeavor and would benefit from developing the appropriate expertise to support both general and special education teachers. It appears that the work of building RTI models in California's schools has just begun. Further professional development and support in the form of curriculum materials and personnel will be necessary for moving forward with RTI.

References

Boardman, A. G., Roberts, G., Vaughn, S., Wexler, J., Murray, C. S., & Kosanovich, M. (2008). Effective instruction for adolescent struggling readers: A practice brief. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Brownell, M. T., Bishop, A., Gersten, R., Klingner, J. K., Penfield, R., Dimino, J. A., Haager, D., Menon, S., & Sindelar, P. (in press). Examining the dimensions of teacher quality for beginning special education teachers: The role of domain expertise. *Exceptional Children.*

Brownell, M. T., Haager, D., Bishop, A., Klingner, J. K., Penfield, R., Dingle, M. & Menon, S. (in preparation). Teacher quality in special education and domain expertise for teaching reading.

Denton, C. A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle-school students with severe reading difficulties. Learning Disabilities Research and Practice, 23 (2), 79-89.

Moody, S. W., Vaughn, S., Hughes, M. T., & Fischer, M. (2000). Reading instruction in the resource room: Set up for failure. *Exceptional Children, 66* (3), 305-316.

President's Commission of Excellence in Special Education (2002). *A new era: Revitalizing special education for children and their families.* Washington, DC: U. S. Department of Education.

Scammacca, N., Roberts, G., Vaughn. S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007), Interventions for adolescent struggling readers: A meta-analysis with implications for practice. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive Remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. Journal of Learning Disabilities, 34(1), 33-58, 78.

Vaughn, S., Levy, S., Coleman, M. & Bos, C. S. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *Journal of Special Education, 36*, 2-14.

Vaughn, S., Moody, S. W., & Schumm, J. S. (1998). Broken promises: Reading instruction in the resource room. *Exceptional Children, 64* (2), 211-225.

Vaughn, S., Wanzek, J., Woodruff, A. L., & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Haager, J. K. Klingner & S. Vaughn (Eds.), *Evidence-based reading practices for response to intervention* (pp. 11-28). Baltimore, MD: Paul H. Brookes Publishing.