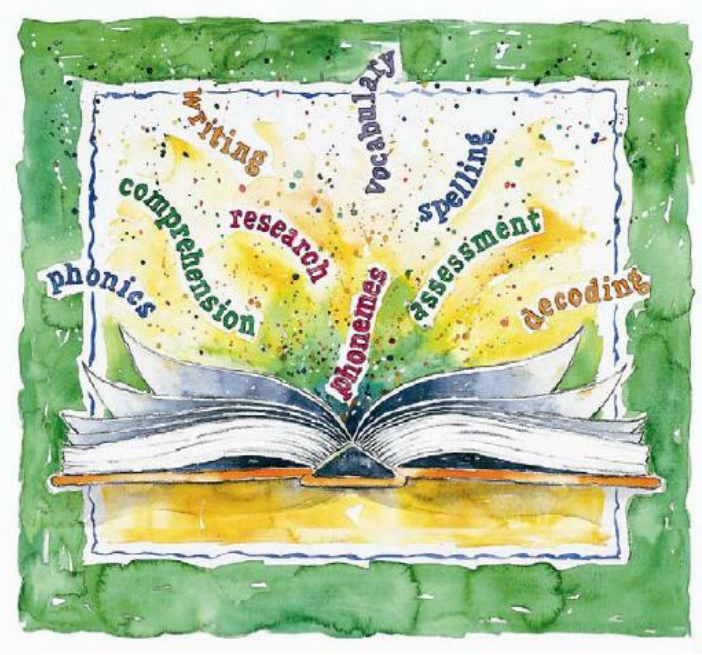
# The California Reading First Year 3 Evaluation Report

**November 2005**




**Prepared by:**

**Diane Haager, Ph.D.**
Principal Investigator

and

**Educational Data Systems**
Renuka Dhar
Mark Moulton, Ph.D.
Seema Varma, Ph.D.

EDUCATIONAL DATA SYSTEMS

# Executive Summary

Reading First is a federal initiative aimed at improving reading instruction in America. Authorized in 2001 as part of the No *Child Left Behind Act*, Reading First promotes the use of scientifically based reading practices in Grades K-3. This initiative provides an unprecedented amount of federal funding for improving reading instruction in schools with large proportions of students experiencing academic difficulty and socio-economic disadvantage. This report evaluates California's progress in implementation and achievement during the first three years of funding and provides information regarding program efficacy.

## General Findings

### Effect of Reading First on Achievement

***What is the impact of the Reading First program on K-3 students in participating schools and districts?***

There is strong evidence that California's Reading First program is having a positive impact on student achievement as measured using the California Standards Test (CST) and the California Achievement Test, 6th Edition (CAT/6), both components of the California Standardized Testing and Reporting program, or STAR; the curriculum-embedded End of Year test (EOY); and the Reading First Achievement Index (RFAI).[1] To support this conclusion, we examine trends in achievement data from 2002 to 2005 using three cohorts of schools. We then compare Reading First schools to non-Reading First "comparison group" schools to examine the differential impact of Reading First. Finally, we compare the achievement gains of High Implementation and Low Implementation Reading First schools to determine the degree to which the program is effective when properly implemented.

<u>Reading First Schools Compared to Themselves Over Time</u>

- Reading First schools consistently show strong upward trends in the percentage of students in the Proficient and Advanced categories of the CST performance scale for Grade 2.[2] These trends are apparent for all three Reading First cohorts. (Cohort 1 entered the program in the 2002-2003 school

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

[2] The California Standards Test classifies students into one of five performance levels: Advanced, Proficient, Basic, Below Basic, and Far Below Basic. In this evaluation, for purposes of analysis, the Proficient and Advanced categories have been combined into a single category called "Proficient and Above." The bottom two categories, Below Basic and Far Below Basic, have also been combined and are sometimes referred to simply as "the bottom" categories.

year, Cohort 2 in the 2003-2004 school year, and Cohort 3 in the 2004-2005 school year.) On the Grade 3 CST metric the trend lines are flat, reflecting a statewide trend.

- There is evidence that the academic advantage of participating in the Reading First program increases with time in the program. In school-level regression models of CST achievement, Years in Program is found to be a statistically significant predictor of 2005 CST scores for Grade 2 (but not Grade 3) after controlling for starting point and percent of English Learners.
- Reading First schools show substantial percentages of students moving out of the bottom CST performance categories (Below Basic and Far Below Basic). Similar trends are observed with regard to the California English Language Development Test (CELDT) Beginning and Early Intermediate categories of English language development.

Reading First Schools Compared to Non-Reading First Schools

- When compared to a demographically matched sample of non-Reading First schools called the "Comparison Group," all three cohorts of Reading First schools show somewhat larger achievement gains than the Comparison Group over time, though the differences are often not significant. Why the differences between Reading First schools and Comparison Group schools are not more significant may, perhaps, be explained by a recent history of statewide and district reading initiatives that may have impacted Comparison Group schools. At present, however, such a history is not available, making Reading First / non-Reading First comparisons hard to interpret.
- High Implementation Reading First schools show higher gains and sharper growth curves than those of the non-Reading First Comparison Group.
- Reading First schools show larger percentages of students moving out of the bottom CST performance categories than Comparison Group schools do.
- These patterns are consistent across the Grade 2 and Grade 3 CSTs and the CAT/6.

High Implementation Reading First Schools Compared to Low Implementation Reading First Schools

- When student achievement is disaggregated by school implementation level, the gains made by High Implementation schools are significantly higher than those for other Reading First schools, especially Low Implementation Reading First schools. In school-level regression models of CST achievement, Reading First implementation was found to be a statistically significant predictor of 2005 CST scores for both Grades 2 and 3 after controlling for starting point and percent of English Learners. This

*supports* (but does not *prove*, given the limitations of our research design[3]) the hypothesis that implementation of Reading First causes achievement to rise.

- High Implementation schools not only move more students into the Proficient and Above category, but the rate at which this movement occurs increases dramatically in the second and third years of program implementation. The focus group interviews support this finding. It can be concluded that Reading First takes at least 1 to 2 years to become integrated into a school's teaching and learning environment. Once that happens, it may be expected to have a strong effect on student learning and achievement. As mentioned, Years in Program is a significant predictor of CST achievement gains for Grade 2, though not for Grade 3.

- High Implementation schools are more successful than Low Implementation schools in moving students out of the bottom performance categories of the CSTs.

- High Implementation schools are well-distributed along the socio-economic continuum of Reading First schools. High Implementation schools show higher gains in student achievement than Low Implementation schools regardless of where they are on that continuum.

Taken as a whole, these comparisons allow us to publish a reasonably confident finding that Reading First is working and that the effect is enhanced by higher degrees of implementation and length of time in the program.

**Program Implementation**

***How well has the Reading First program been implemented in participating schools and districts?***

- The basic elements of Reading First have been adequately implemented in Reading First schools. Ninety-six percent of schools in 2005 were rated "adequate" or better by teachers in Reading First schools.

- The Reading First Implementation Survey, completed by teachers, principals and coaches, yields a measure of each individual school's implementation called the Reading First Implementation Index (RFII). The mean RFII for 2005 was 36, the same as that for 2004.

- Nonetheless, Cohorts 1 and 2 have raised their levels of implementation since the 2003-2004 school year, supporting the finding that implementation takes at least one or two years to take hold.

[3] Because Reading First schools are not randomly assigned to Low Implementation and High Implementation groups as a true experimental design would require, no causal inferences can legitimately be made. We can only say that implementation significantly *predicts* achievement ($p<0.05$). Nonetheless, our findings are consistent with what would be expected if there were a causal relationship.

- Cohort 3 has begun at a lower level of implementation than Cohorts 1 and 2 did in their first years (though the RFII for Cohort 1 in its first year is not known) due mainly to lower levels of teacher and coach professional development. Because many Cohort 3 schools are in rural areas, it is possible that they have less access to high quality professional development, but the full reasons for Cohort 3's slow start are not yet known.

## Policy Recommendations

### Policy Recommendation 1

*Continue to focus on full implementation of Reading First.*

Review of school RFII measures suggests that Reading First is being adequately implemented across the state in essential respects, in particular professional development, program materials, coaching, and instructional practices. However, our findings show that high-implementing schools yield higher academic gains than moderate- or low-implementing schools. State and local Reading First personnel should focus extra efforts on schools with lower levels of implementation to ensure maximum benefit from participation in Reading First. Cohort 3 in particular may need extra monitoring and support from LEAs and the State and regional technical assistance services in this regard.

### Policy Recommendation 2

*Support participation in Reading First over multiple years. It takes at least two years of implementation to show significant achievement gains, even with extensive training and support. Continued support beyond the initial two years is essential to achieve significant and lasting results and to establish the long-term institutional changes needed for Reading First instructional practices to continue even after funding is discontinued.*

Quantitative and qualitative data support the notion that continued program participation leads to continued and lasting improvement in teaching practices and student outcomes. Extended time (e.g., 6 years) will allow school personnel to gain depth of knowledge, refine their skills, and integrate program principles into the fabric of their school operations. On the other hand, if individual schools show little evidence of benefiting from the program after three years of support, they should be dropped from the program as the Reading First NCLB legislation requires. Cohort 1 schools show steady academic gains over three years of participation, supporting the premise that extended support and participation leads to continued improvement.

## Table of Contents

# List of Tables

## List of Figures

# List of Appendices

# Chapter 1: Introduction

Reading First is a federal initiative aimed at improving reading instruction in America. Authorized in 2001 as part of the federal *No Child Left Behind Act*, Reading First promotes the use of scientifically based reading practices 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. California's successful application for federal Reading First funds met strict federal criteria by outlining a plan for assessments, instruction, materials, professional development, monitoring and oversight. This report evaluates California Reading First program implementation and academic achievement during the first three years of funding.

For the 2002-2003 school year (Year 1), California received approximately $133 million for its Reading First program. Funding totaled $146 million for the 2003-2004 academic year (Year 2) and $152 million for the 2004-2005 academic year (Year 3). With Reading First funds, California has established a support network to conduct training, to assist school districts in acquiring curricular materials, to monitor progress toward goals, and to provide technical assistance to participating schools and school districts.

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

School districts are required to use one of California's state-adopted reading programs: SRA/McGraw-Hill's *Open Court Reading 2000* or *2002* (OCR) or the Houghton Mifflin *Reading: A Legacy of Literacy 2003* (HM). California law (Proposition 227) mandates instruction in English for all students unless parents sign a waiver requesting bilingual instruction for their child. In this study, classes taught in Spanish are referred to as "waiver" classrooms. In 2004, the State adopted two Spanish language reading programs that reflect research-based instructional principles: SRA/McGraw Hill's *Foro abierto para la lectura* and Houghton Mifflin's *Lectura: Herencia y futuro* programs. Students receiving reading instruction in Spanish must participate in the statewide testing program in English and must transition out

of Spanish-instruction classes and take the English STAR test by the end of Grade 3. In the 2004-2005 school year, California's Reading First program began offering support and training for schools with waiver classrooms using the adopted Spanish language programs.

California's Reading First program provides for a system of state and regional technical assistance for implementation of Reading First. Reading First is designed to build state and local capacity in reading education. During the first three years of implementation, C-TAC, housed at the Sacramento County Office of Education Reading Lions Center, conducted regional meetings for school district-level Reading First administrators, including an annual "Superintendents' Summit" and professional development training for more than 900 Reading First coaches. During the first two years, C-TAC developed eight Regional Technical Assistance Centers (R-TACs) at geographically disbursed county offices of education throughout the state to facilitate Reading First technical assistance at the local school district level. C-TAC and the R-TACs have assisted local school districts with various aspects of implementation of reading instructional programs, beginning and advanced professional development programs, assessments and the use of data from these assessments, and coaching strategies connected to the OCR and HM programs. For further information on C-TAC assistance for local Reading First school districts in California, the reader may visit www.calread.net.

The 2004-2005 academic year marked the end of the first three years of funding for California. In October 2002, local educational agencies (LEAs) submitted applications in a competitive grant process. The SBE selected 13 school districts with a total of 283 schools in November 2002 to receive funding, in this report referred to as Cohort 1. The funding was not disbursed to the school districts until after February 2003; thus for this first year, characterized as a start-up year, Cohort 1 schools had approximately a half-year to implement Reading First. Schools in Cohort 2 were selected in a competitive grant process in the spring of 2003 when an additional 60 school districts, including just fewer than 400 schools, were selected for funding for the 2003-04 school year. A third round of applications was conducted in the spring of 2004 and an additional 37 school districts were selected involving roughly 135 schools for funding beginning in the 2004-05 school year (Cohort 3 schools). A total of 821 schools in 110 districts received Reading First funding in the 2004-05 academic year.

**Components of the California Reading First Plan**

The California Reading First Plan is based on a series of Assurances that are implemented by the LEAs. CDE, C-TAC, and the R-TACs monitor and assist LEAs in implementing the Assurances. The Reading First program is designed to ensure full implementation with fidelity to a comprehensive research-based reading program.

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

Instructional Program

Each LEA is required to implement fully the district-adopted reading program for 60 minutes per day in Kindergarten and 150 minutes per day in Grades 1-3 according to a district pacing plan that outlines when each daily lesson is taught at each grade level in an academic year. This plan not only assures that students will complete the grade-level curriculum but also that implementation occurs systematically throughout every classroom at each grade in every Reading First school.

Professional Development

LEAs must assure that all K-3 and special education K-12 teachers in Reading First schools participate in professional development focused on the adopted reading program during the first three years of the program. Teachers attend a state-approved 40-hour training as mandated in AB 466 during the first year of a school's participation in Reading First. In these trainings, teachers learn about the design, content and instructional strategies included in the program. Subsequent year professional development programs provide in-depth emphasis on research based instructional strategies embedded in the OCR and HM programs. Participant teachers must also annually complete 80 hours of district-sponsored follow-up professional development based on local needs.

Most of the Reading First LEAs use funding to support local experts (i.e., reading coaches, content experts, peer support teachers) to reinforce teacher use of the strategies acquired in the training. These site-based experts provide classroom-based training by going into classrooms and consulting with teachers, often demonstrating lessons or assisting with planning.

Coaching

The Reading First program encourages LEAs to utilize coaches or content experts solely to provide site-based support of the adopted reading program. Twenty days of extensive professional development for coaches was provided by the C-TAC in 2004-05. At the Coach's Institutes, delivered both in the southern and northern regions of California, coaches, content experts and district coach coordinators not only developed expertise in the adopted reading program at all K-3 grade levels, but also learned how to provide consultation and serve in a leadership role regarding reading instruction. C-TAC developed and provided the instruction for the Coaches Institutes (40 modules). LEAs may request technical assistance from the R-TACs on how to support their coaches.

Curriculum-Embedded Assessment

LEAs are required to conduct curriculum-based assessments for ongoing program monitoring. Teachers conduct assessments every 6 to 8 weeks to determine student progress and the efficacy of the delivery of instruction. The assessments are aligned with each unit/theme content in six areas: vocabulary, comprehension, oral fluency, spelling, usage, and writing. 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 tests (EOY)—the curriculum-embedded 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 4 of this report). The LEAs' Internal Evaluation Reports also use these End-of-Year assessments to provide information about students' reading performance and to evaluate the impact of Reading First at the school level.

Collaborative Teacher Meetings

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

Site Leadership

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

Program Coherence

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

State Leadership

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

Technical Assistance

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

Special Education Referral Reduction Program

In the fall of 2004, the State provided an opportunity for LEAs to apply for one-time-only additional Reading First funds to support activities aimed at reducing the number of K-3 students referred for special education. Criteria for funding included a multi-level plan that starts with improving instructional strategies using the core-adopted program. If needed, various interventions are introduced for addressing specific needs of struggling readers with supplemental materials, additional instructional time, and a stand-alone intensive intervention program (available for Grade 3 students only). Ninety-eight out of the 110 Reading First LEAs participated in this program. The funds were designated for purchase of the supplemental intervention materials or stand-alone materials; diagnostic assessments in the areas of phonemic awareness, decoding/phonics, oral fluency, vocabulary, and reading comprehension; and for release time or substitute time for additional training in the use of the assessments or materials. The SBE approved all supplemental intervention and stand-alone intensive intervention programs that were recommended through a two-phase review process. The CDE's Reading/Language Arts Leadership Office and the Office of Special Education conducted the first level of review based on programs listed

with the Florida State Reading Center and/or the University of Oregon Reading Center. The C-TAC conducted the second level of review. Between November 2004 and May 2005, SBE had approved materials from 36 publishers.

**California Reading First Evaluation Study**

The California Reading First Plan approved by the U. S. Department of Education (USDE), included an annual evaluation to study the implementation of the program and to evaluate program outcomes. A Request for Proposals (RFP) for prospective vendors for this study was issued in the spring of 2003 and at its June 2003 meeting, the SBE approved Educational Data Systems (EDS[4]) as the contractor for the Reading First evaluation study. The contract was finalized in November 2003. EDS completed a Year 1 Report in April 2004, which covered start-up activities and outcomes for the 2002-03 school year. The Year 2 Report, completed in May 2005, reported on activities and outcomes for the 2003-04 school year and provided a cumulative look over the first two years of implementation. This report serves as an evaluation of Year 3, reporting on activities and outcomes throughout the 2004-2005 academic year.

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

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

(b) 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? (See Chapter 3 and Appendices B, C, D, and E.)

Under program impact, the guiding questions ask:

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

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

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

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

**Evaluation Study Design**

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

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

2. Implementation Measures. Program implementation was measured using comprehensive surveys given to teachers, principals and coaches of all Reading First schools. A secondary, qualitative source of implementation data is a set of focus group interviews conducted throughout the state in 2005. The implementation data include responses of teachers, reading coaches and principals regarding their own and their school's implementation of the Reading First program, including use of scientifically-based instructional materials, involvement in professional development, use of 6-8 week unit assessment data, and use of coaching services. For the 2003-04 school year (Year 2), EDS designed a set of three surveys (one each for Reading First classroom teachers, coaches, and school principals) and collected data from roughly 14,000 Reading First teachers, coaches, and school principals during the spring of 2004. For the 2004-2005 school year (Year 3), adjustments were made to the survey based on feedback from the EAG. Questions were added to capture information on the use of the Spanish versions of the curricular materials. The surveys collected data from approximately 20,200 teachers, coaches and principals in the spring of 2005.

3. Outcome Measures. Student academic progress in basic reading skills is of primary importance in evaluating the impact of Reading First on children in grades K-3. As shown in Table 1.1 below, this report includes student outcome data from three sources over the three years of implementation. The STAR testing program includes the California Standards Test (CST), designed specifically to measure the challenging content standards adopted by the SBE in 1997,

and a Norm-Referenced Test (NRT) for Grade 3, designed to measure a broader range of reading achievement. An additional data source collected by C-TAC in their monitoring and assistance role includes the End-of-Year (EOY) assessment administered by classroom teachers, which measures oral fluency in Grades 1, 2 and 3 and phonological awareness and alphabetic knowledge in Kindergarten.

**Table 1.1: Student Assessments Used in Reading First Evaluation Study**

| | EOY | CST | NRT |
|---|---|---|---|
| Kindergarten | X | | |
| Grade 1 | X | | |
| Grade 2 | X | X | |
| Grade 3 | X | X | X |

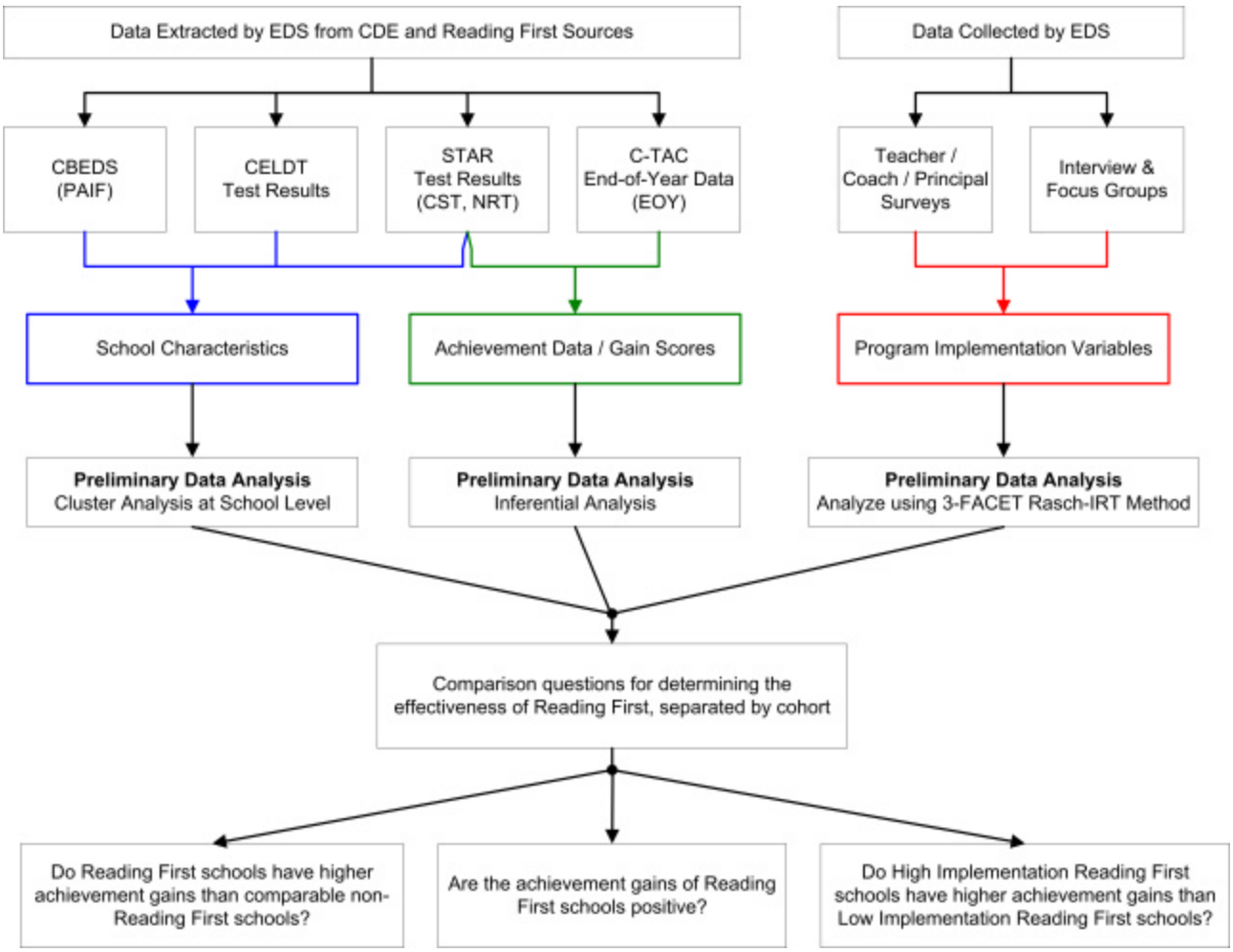
Note: The Kindergarten EOY test includes seven subtests: Consonants, Lower Case Letters, Phonics, Rhyming, Syllables, Upper Case Letters and Vowels. The EOY tests for Grades 1, 2 and 3 consist of timed oral reading fluency passages. The CSTs include data from the English Language Arts component. The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

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

The conceptual framework below illustrates the data sources for the Year 3 study and how they are grouped into demographic, implementation and student outcome variables. The framework then lists the type of statistical analyses utilized to address each set of variables and the three "comparison questions" used to determine whether the Reading First program is effective (see the bottom three boxes). This evaluation study does not employ a true experimental research design (which would fail to answer some guiding questions, and would not in any case have been technically feasible), but by relating school characteristics and program implementation to student outcomes it does employ three cross-validating

non-experimental designs that, taken together, provide defensible answers to the five guiding questions of this study.

**Figure 1.1: Conceptual Framework – Year 3**




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

Chapter 4 provides information on student achievement measures, including CST, NRT, EOY, and RFAI gain scores for Reading First schools, non-Reading First Comparison Group schools, and all K-3 schools in California, and compares the achievement gains of High Implementation and Low Implementation Reading First schools. It makes the case that Reading First schools have higher gains than non-Reading First schools and that program implementation is a strong predictor of achievement gains. Chapter 4 also describes the construction of the Reading First Achievement Index (RFAI) and its usefulness in determining significant progress.[5] Chapter 4 corresponds to the "Achievement Data/Gain Scores" box and to the three "comparison question" boxes at the bottom of the Conceptual Framework. Associated with Chapter 4, Appendix F provides results of a multivariate regression analysis of CST scores and reports disaggregated achievement gains. Appendix G describes how the RFAI is calculated. Appendix H provides a complete list of individual schools with selected achievement measures as required in Reading First reporting.

Chapter 5, corresponding to the "Program Implementation Variables" box of the conceptual framework, presents information from focus group interviews, including qualitative information on unintended consequences of the Reading First program in California. Finally, Chapter 6 presents the findings of three years of evaluation as relates to each of the original research questions, gives policy implications and recommendations, and makes suggestions for areas of further study.

[5] The use of the RFAI for determining whether schools have made significant progress and are eligible for continued funding is pending with the State Board of Education at the time of the writing of this report.

## Chapter 2: Demographics

Chapter 2 provides a description and analysis of demographic data for Reading First and non-Reading First schools in California. The data for Reading First schools is presented both for the group as a whole and broken out by cohort. The demographic data for non-Reading First schools is presented for two separate groups of non-Reading First schools and one group of All Elementary schools in the state. The goal of this chapter is: (1) to present demographic information for the three cohorts of Reading First schools across three years of the Reading First program; (2) to examine the comparability of Reading First and non-Reading First schools across various demographic dimensions; and (3) to demonstrate that of the two comparison groups to which Reading First schools have been compared in previous reports, one stands out as the most suitable comparison group.

### Background and Overview

The Reading First program started in California during the 2002-2003 academic year with 283 schools in 13 school districts. These 283 schools are referred to as Cohort 1 in this report. In the 2003-2004 academic year, 391 schools from 60 school districts were funded and are referred to as Cohort 2. In the 2004-2005 academic year, 158 schools from 37 school districts were added to the Reading First program and are referred to as Cohort 3. The first column in Table 2.1 provides the number of schools in each Reading First cohorts by funding year.

In Years 1 and 2 of the evaluation, Reading First schools were compared to two groups of non-Reading First schools labeled Comparison Group A and Comparison Group B in previous reports. Comparison Group A (labeled RF Eligible Schools in Table 2.1) was composed of schools that were randomly selected from districts eligible for Reading First funding but not funded in either round of applications. Comparison Group B schools (labeled Comparison Group in Table 2.1) were selected using a cluster methodology and randomly selected from a list of schools in the state that most closely matched the Reading First schools for Socio-Economically Disadvantaged (SED) and English Learner (EL)[6] school percentages.

As of this Year 3 Evaluation Report, what in previous years were referred to as Comparison Group A schools are now referred to as Reading First Eligible (RF Eligible) schools. What in previous years were referred to as Comparison Group B schools will now be referred to simply as the Comparison Group schools. The reasons for making these changes are elaborated in subsequent sections of this chapter.

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

**Table 2.1: Number of Reading First and Non-Reading First Comparison Schools**

| Academic Year | Number of Schools | | |
|---|---|---|---|
| | RF Schools | RF Eligible Schools (was Comparison Group A) | Comparison Group Schools (was Comparison Group B) |
| 2002-2003 | 283 (Cohort 1) | 283 | 283 |
| 2003-2004 | 391 (Cohort 2) | 400 | 400 |
| 2004-2005 | 158 (Cohort 3) | 393[2] | 392[3] |
| 2005 | 821[1] (Cohorts 1,2 and 3) | 393 | 392 |

[1]The total number of Reading First schools in 2005 was 821 and not 832 (the sum of schools in cohorts 1, 2 and 3) because by 2005, 11 RF schools closed or left the program.
[2]Seven schools from the RF Eligible group of schools closed, showed no student enrollment, or changed their status to higher grade schools, causing the N of schools to drop from 400 to 393 between 2003-2004 and 2004-2005.
[3]Eight Comparison Group schools closed, showed no student enrollment, or changed their status to higher grade schools, causing the N of schools to drop from 400 to 392 between 2003-2004 and 2004-2005.

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

The aggregated student descriptive data presented in this chapter consists of:

- Average school-level percentage of Socio-Economically Disadvantaged (SED) students
- Average school-level percentage of English Learner (EL) students
- Average school-level percentage of Students with Disabilities
- Average school-level percentages of students in various ethnic subgroups: African American, American Indian, Asian, Filipino, Hispanic, Pacific Islander, and White
- Average school-level percentages of students in Beginning and Early Intermediate performance levels as measured on the CELDT test

[7]The STAR research file used for the 2004-2005 data was the version released on September 16, 2005, referred to as "P2."

Two types of non-student demographic data are presented for Reading First districts, specifically:

- The urban-rural location breakdown of Reading First districts by cohort
- The number of waiver classrooms funded in the 2003-2004 and 2004-2005 academic years, by cohort

Teacher credential and experience data are presented for Reading First and non-Reading First schools, including:

- Percentage of teachers at varying educational degree categories, from Ph.D.s to less than Bachelors
- A weighted teacher qualification index ranging from 1 to 5
- The average percentage of fully credentialed teachers
- Average years of teaching

**Presentation and Discussion of Data**

Socio-Economically Disadvantage (SED)

The first row in Table 2.2 compares Reading First school cohorts to non-Reading First schools with regard to socio-economic status (SED). We see that Cohort 1 Reading First schools started with the highest percentages (90.4% in 2003, 92.1% in 2004) of students with socio-economic disadvantage of all the school groups. Cohort 2 Reading First schools had lower percentages of students with SED (84.3% in 2004) than Cohort 1 when they began in the Reading First program, but the percentage increased to 87.8% in 2005, just above that of Cohort 1 (87.7%). Cohort 3 is in the same range as Cohorts 1 and 2 with 85.7% of its students classified as SED. Thus, by 2005 all three cohorts were similar in terms of SED. The Comparison Group schools showed lower percentages of students with SED than the Reading First schools (about 82%) in both 2004 and 2005, whereas the RF Eligible schools were even more dissimilar to the Reading First schools when compared on SED. Their average SED percent was 78.8% in 2004 and 82.2% in 2005. When compared to the state average (53% in 2005), it is evident that the Reading First cohorts have a relatively high concentration of SED students, consistent with the federal guidelines that funds must go to schools with high levels of SED and academic underachievement.

English Learners (EL)

The demographic data show that on average 57% of the students in Reading First schools had limited English proficiency. Cohort 3 had the highest percentage of EL students (58.8%, compared to 57.6% for

Cohort 1 and 56.2% for Cohort 2). The Comparison Group schools were very similar to the Reading First schools with regard to English Learners, with an average EL of 57.5% in 2005. The RF Eligible school group, on the other hand, had much lower percentages of EL students, 47% in 2004 and 49.7% in 2005. This 8-10 percentage point difference may have a significant effect on achievement comparisons. Proficiency in English is highly correlated with success in academics and with achievement scores (August & Hakuta, 1997; National Research Council, 2002). Given the large difference in the percentage of EL students between the Reading First and RF Eligible schools, there is little justification for treating the two groups as comparable for the purpose of this evaluation study. That is one reason why the RF Eligible schools are no longer referred to as Comparison Group A or used in comparative analyses. We will return to this point later in the chapter.

Students with Disabilities

Students with disabilities are served by the Reading First program only if they receive their reading instruction in the general education classroom. Though special education teachers are included in Reading First professional development and may be using the core reading program, they are not direct recipients of Reading First support. The data presented in Table 2.2 show that, in general, Reading First schools had lower percentages of students with disabilities (average of 7.9%) compared to the Comparison Group schools (average of 8.9%) or the All Elementary schools (average of 10.6%). Among the three cohorts, although they are relatively similar, Cohort 1 had the highest percentage of students with disabilities (8.6% in 2005) compared to Cohort 2 (7.7% in 2005) or Cohort 3 (7.2%).

Ethnicity

When compared on Ethnicity, the three Reading First cohorts differ from each other in several respects. Looking at the most recent study year, 2005, Cohort 3 had the highest percentage of Hispanic students (78.2%) compared to Cohorts 1 and 2 with percentages of 73.2% and 74.5%, respectively. Cohort 3 also had the lowest percentage of African American students (6.4%) and the highest percentage of White students (10.4%). Cohort 1 had the highest percentage of African American students (15.4%) and the lowest percentage of White students (3.7%). These differences in ethnic group percentages by cohort may have occurred because of differences in the concentration of schools in rural and urban areas across the different cohorts. We study this idea a bit more in the next section.

The Comparison Group schools had a relatively high percentage of Hispanic students (71.6%) as compared to the state as a whole, but not as high as the Reading First schools, which averaged 75.3%. Moreover, the Comparison Group schools had a higher percentage of White (11.6%) and Asian (7.0%) students and a lower percentage of African American students (5.1%) than the Reading First schools.

The RF Eligible schools were quite disparate in their ethnic makeup when compared to the Reading First cohorts. In 2005, the RF Eligible schools had 69.4% Hispanic students (lower than the Reading First schools), 15.8% White students (much higher than Reading First schools), and 6.4% African American students (lower than Reading First schools, except for Cohort 3). The disparity in ethnic data between Reading First schools and RF Eligible schools further supports the decision to discontinue using the latter for comparison purposes.

The average elementary school in the state is composed of 7.6% African American students, 7.5% Asian students, 42.6% Hispanic students, and 33.9% White students.

The ethnicity data also support the conclusions drawn from the SED and EL data that the Reading First cohorts differ from each other demographically. Even though all three Reading First cohorts contain SED and EL students, their ethnic composition varies on important factors and does not justify treating them as one homogenous Reading First population. As of this Year 3 Evaluation Report, most of our analyses are broken out by cohort.

**Table 2.2: Student Demographic Data, 2002 to 2005**

| | Reading First Schools | | | | | | | | | Comparison Group Schools [3] | | | RF Eligible Schools [3] | | | All Elementary Schools [4] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cohort 1 | | | Cohort 2 [1] | | | Cohort 3 [2] | | | | | | | | | | | |
| | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 |
| *Number of Schools* | *283* | *282* | *276* | - | *391* | *386* | - | - | *152* | - | *400* | *392* | - | *400* | *393* | *5823* | *5919* | *5977* |
| SED (%) | 90.4 | 92.1 | 87.7 | - | 84.3 | 87.8 | - | - | 85.7 | - | 82.7 | 82.1 | - | 78.8 | 82.2 | 51 | 51.6 | 53.3 |
| EL (%) | 57.1 | 57.1 | 57.6 | - | 54.6 | 56.2 | - | - | 58.8 | - | 57.0 | 57.5 | - | 47.4 | 49.7 | 27.1 | 28.2 | 29.3 |
| Students with Disabilities (%) | 7.7 | 8.7 | 8.6 | - | 7.9 | 7.7 | - | - | 7.2 | - | 9.4 | 8.5 | - | 7.9 | 7.8 | 9.8 | 11.0 | 11.1 |
| African American (%) | 17.1 | 16.7 | 15.4 | - | 9.8 | 9.2 | - | - | 6.4 | - | 5.2 | 5.1 | | 6.7 | 6.4 | 7.8 | 7.8 | 7.6 |
| American Indian (%) | 0.3 | 0.3 | 0.3 | - | 0.9 | 0.9 | - | - | 0.7 | - | 0.8 | 0.9 | - | 1.5 | 1.6 | 1.3 | 1.3 | 1.3 |
| Asian (%) | 4.5 | 4.2 | 4.4 | - | 4.1 | 3.8 | - | - | 1.1 | - | 7.1 | 7.0 | - | 3.9 | 3.9 | 7.3 | 7.3 | 7.5 |
| Filipino (%) | 1.1 | 1.0 | 1.2 | - | 1.5 | 1.5 | - | - | 1.2 | - | 1.5 | 1.6 | - | 1.2 | 1.2 | 2.2 | 2.2 | 2.3 |
| Hispanic (%) | 70.5 | 72 | 73.2 | - | 73.1 | 74.5 | - | - | 78.2 | - | 71.5 | 71.6 | - | 67.9 | 69.4 | 40.2 | 41.5 | 42.6 |
| Pacific Islander (%) | 0.6 | 0.6 | 0.5 | - | 0.7 | 0.7 | - | - | 0.5 | - | 0.5 | 0.6 | - | 0.4 | 0.5 | 0.6 | 0.6 | 0.7 |
| White (%) | 4.1 | 3.9 | 3.7 | - | 8.5 | 8.0 | - | - | 10.4 | - | 12.1 | 11.6 | - | 17.2 | 15.8 | 36.5 | 35.2 | 33.9 |

[1]Cohort 2 demographics are provided beginning in 2004 because 2003-2004 was the first year of Reading First Implementation in those schools.
[2]Cohort 3 demographics are provided beginning in 2005 because 2004-2005 was the first year of Reading First Implementation in those schools.
[3]Demographics for the Comparison Group Schools and the RF Eligible Schools are presented only from 2004 and 2005 because these groups were formed in the second year of the study.
[4]The group "All Elementary Schools" *includes* Reading First schools in this chapter. Whereas in Chapter 4, "All Elementary Schools" *excludes* Reading First schools.

Urban-Rural Distribution

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

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

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

| District Location | Cohort 1 | | Cohort 2 | | Cohort 3 | | All Cohorts | |
|---|---|---|---|---|---|---|---|---|
| | N of Districts | Percent of Districts[2] | N of Districts | Percent of Districts | N of Districts | Percent of Districts | N of Districts | Percent of Districts |
| Large City | 6 | 46.2 | 10 | 16.7 | 2 | 5.4 | 18 | 16.4 |
| Mid-size City | 4 | 30.8 | 11 | 18.3 | 7 | 18.9 | 22 | 20.0 |
| Urban Fringe of Large City | 1 | 7.7 | 16 | 26.7 | 10 | 27.0 | 27 | 24.5 |
| Urban Fringe of Mid-size City | 1 | 7.7 | 14 | 23.3 | 9 | 24.3 | 24 | 21.8 |
| Small Town | 0 | 0.0 | 1 | 1.7 | 1 | 2.7 | 2 | 1.8 |
| Rural | 1 | 7.7 | 7 | 11.7 | 7 | 18.9 | 15 | 13.6 |
| Unknown[1] | 0 | 0.0 | 1 | 1.7 | 1 | 2.7 | 2 | 1.8 |
| Total | 13 | 100.0 | 60 | 100.0 | 37 | 100.0 | 110 | 100.0 |

[1]Location information is not available for the Santa Monica Boulevard Community Charter School in Los Angeles and the East Palo Alto Charter School in San Mateo.

[2] The percent of the districts in that cohort in a particular type of location.

California English Language Development Test (CELDT)

The CELDT is a statewide test administered to all English Learners in California for the purpose of assessing English language development. Because more than 57% of students (on average across all three cohorts in 2005) in Reading First schools are English Learners (from Table 2.2), it is important to examine what effects Reading First may have on the EL population.

Table 2.4 provides the percentages of students at the Early and Beginning Intermediate levels on the CELDT (indicating limited English proficiency) for each Reading First cohort, the Comparison Group schools, the RF Eligible schools, and the state as a whole. Based on the 2005 CELDT data, Cohort 3 had the highest percentage of students at the Early and Beginning Intermediate levels, ranging from an average of 28.3% in Grade 1 to a high of 48.8% in Grade 3. In 2005, Cohort 1 had the lowest percentages of the three cohorts in these categories, from 20.6% in Grade 1 to 39.5% in Grade 3.

Of interest is that when Cohort 1 received Reading First funding in 2002-2003, the percentage of students scoring on CELDT at the Early and Beginning Intermediate levels was similar to that of Cohort 3 in 2005. However, the percentages dropped steadily in Cohort 1 schools over three years to the lower numbers we see in 2005. The same pattern is observed with Cohort 2 over two years of implementation. The CELDT percentages dropped between the first year of program implementation (2003-2004) to the present. This implies that the students scoring in the Beginning and Early Intermediate levels are moving into the more proficient CELDT performance levels or are being reclassified as English proficient. Notably, neither the Comparison Group schools nor the RF Eligible schools showed a similar positive trend in the movement out of Beginning and Early Intermediate levels. For both 2004 and 2005, the CELDT percentages remain unchanged for both non-Reading First groups. Statewide, CELDT percentages of students in the lowest two performance levels dropped from 2003 to 2004, but increased slightly from 2004 to 2005.

The CELDT data raise the hypothesis—and it cannot be more than that yet—that the percentage of Early and Beginning Intermediate students in Reading First schools is declining over time as a result of the Reading First program. This may reflect a stronger English language development effort in Kindergarten and Grade 1 in recent years in Reading First schools. The trend warrants further study.

**Table 2.4: CELDT Average Percentage of Beginning and Early Intermediate Students**

| | Reading First Schools | | | | | | | | | Comparison Group Schools [3] | | | RF Eligible Schools [3] | | | All Elementary Schools [4] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cohort 1 | | | Cohort 2 [1] | | | Cohort 3 [2] | | | | | | | | | | | |
| | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 |
| *Number of Schools* | *283* | *282* | *276* | - | *391* | *386* | - | - | *152* | - | *400* | *388* | - | *400* | *393* | *5823* | *5919* | *5744* |
| Grade 1 (%) | 27.6 | 22.3 | 20.6 | - | 23.4 | 23.3 | - | - | 28.3 | - | 19.4 | 20.8 | - | 19.7 | 19.8 | 16.9 | 12.4 | 12.8 |
| Grade 2 (%) | 38.8 | 32.3 | 28.8 | - | 33.9 | 31.8 | - | - | 38.9 | - | 29.9 | 29.6 | - | 31.1 | 29.8 | 24.6 | 19.3 | 19.1 |
| Grade 3 (%) | 47.6 | 40.2 | 39.5 | - | 43.6 | 43.3 | - | - | 48.8 | - | 39.3 | 40.2 | - | 41.4 | 40.6 | 29.6 | 25.7 | 26.8 |
| Grades 1-3 (%) | 38.0 | 31.5 | 29.6 | - | 33.6 | 32.8 | - | - | 38.7 | - | 29.5 | 30.2 | - | 30.7 | 30.1 | 23.7 | 19.1 | 19.6 |

[1]Cohort 2 demographics are provided beginning in 2004 because 2003-2004 was the first year of Reading First Implementation in those schools.
[2]Cohort 3 demographics are provided beginning in 2005 because 2004-2005 was the first year of Reading First Implementation in those schools.
[3]Demographics for the Comparison Group Schools and the RF Eligible Schools are presented only from 2004 and 2005 because these groups were formed in the second year of the study.
[4]The group "All Elementary Schools" *includes* Reading First schools in this chapter. Whereas in Chapter 4, "All Elementary Schools" *excludes* Reading First schools.

Waiver Classrooms

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

Table 2.5 presents the number of waiver classrooms added to the Reading First program by cohort and year. As of this study there are 1734 waiver classrooms in Reading First schools across 51 districts. Cohort 1 has a lower percentage of waiver classrooms (23.8%) than Cohorts 2 and 3 (40.1% and 36.2%, respectively). Cohort 2 has the highest percentage of waiver classrooms of the three cohorts.

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

**Table 2.5.1: Number of Waiver Classrooms Added to Reading First Districts by Cohort, 2003-2004 and 2004-2005**

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

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

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

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

Teacher Qualifications

Table 2.6 provides teacher credential and experience information as available in the CBEDS teacher Professional Assignment Information Form (PAIF) research files. This database shows the percent of teachers falling into each educational degree category by cohort and year. A comparison of the Reading First cohorts in terms of teacher qualifications shows that, in general, the teachers in Cohort 1 had lower educational degrees than teachers in Cohorts 2 and 3. Cohort 1 had fewer teachers with higher levels of education (Bachelors plus 30 or more semester units, Masters, Masters plus 30 or more semester units, and Ph.D.s) than the other cohorts. To more easily compare cohorts to each other and to the other non-Reading First groups, a weighted index was computed based on CBEDS data sources relative to teacher qualifications. The weighted teacher qualification is an index ranging from a low teacher qualification of 1 to a high teacher qualification of 5. Refer to the note under Table 2.6 for a description of how this index was computed. This index shows that Cohort 1 Reading First schools had the lowest index (2.0 to 2.11) of the other Reading First cohorts (ranging from 2.2 to 2.31) and non-Reading First schools. Of the Reading First cohorts, Cohort 3 had the highest percent of teachers with higher education.

It is interesting to note that in the data presented previously in Table 2.3, the urban-rural distribution of Reading First schools, Cohort 1 had the highest concentration of urban schools, and in Table 2.6 we see that Cohort 1 also had the lowest number of teachers with advanced degrees and full credentials. This is consistent with national reports indicating that urban schools tend to have a greater number of teachers with lower levels of qualifications, including more teachers with less than full credentials and fewer teachers with advanced degrees (Darling-Hammond, 1999; National Commission on Teaching and America's Future, 1996).

There are several indications of positive change in teacher qualifications in Reading First schools in Table 2.6. In Cohorts 1 and 2, the only cohorts with multiple years of data, we see an increase over time in the percents of teachers holding a Bachelors degree plus 30 semester hours. In contrast, the Comparison Group schools and the RF Eligible schools show a slight decrease, and the All Elementary schools show almost no change in this category across time. The number of fully credentialed teachers also increased over time in Cohorts 1 and 2, though increases also occurred in the Comparison Group schools, the RF Eligible schools and in the All Elementary schools groups.

Though the statewide increase in credentialed teachers may be a result of the pressure on districts to employ "highly qualified" teachers as a result of the *No Child Left Behind Act*, it is possible that

participation in Reading First can explain the increases in the "Bachelors plus 30" category for Cohorts 1 and 2 of Reading First schools. Reading First has provided opportunities for teachers to earn course credit for some aspects of advanced training. This is an important factor that will be monitored in future years.

**Table 2.6: Elementary Teacher Credential and Experience**

| | Reading First Schools | | | | | | | | | Comparison Group Schools [3] | | | RF Eligible Schools [3] | | | All Elementary Schools [5] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cohort 1 | | | Cohort 2 [1] | | | Cohort 3 [2] | | | | | | | | | | | |
| | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 |
| *Number of Schools* | *283* | *282* | *276* | - | *391* | *377* | - | - | *152* | - | *400* | *388* | - | *400* | *393* | *5647* | *5694* | *5720* |
| **Highest Degree held** | | | | | | | | | | | | | | | | | | |
| PhDs (%) | 0.5 | 0.7 | 0.7 | - | 0.7 | 0.6 | - | - | 0.7 | - | *0.9* | 0.6 | - | 0.9 | 0.6 | 0.9 | 0.8 | 0.8 |
| Masters plus 30 or more semester units (%) | 9.3 | 11.7 | 12.6 | - | 13.7 | 13.6 | - | - | 15.6 | - | 11.7 | 11.2 | - | 12.3 | 11.8 | 14.0 | 14.5 | 14.3 |
| Masters (%) | 10.8 | 11.6 | 12.1 | - | 15.9 | 17.3 | - | - | 18.1 | - | 18.3 | 20 | - | 17.2 | 19.0 | 15.5 | 16.9 | 18.1 |
| Bachelors plus 30 or more semester units (%) | 43.4 | 46 | 46.6 | - | 47.9 | 48.6 | - | - | 45.8 | - | 52.8 | 51.3 | - | 53.1 | 51.8 | 51.3 | 51.7 | 50.6 |
| **Total Advanced Degrees** | **64.0** | **70.0** | **72.0** | **-** | **78.2** | **80.1** | **-** | **-** | **80.2** | **-** | **83.7** | **83.1** | **-** | **83.5** | **83.2** | **81.7** | **83.9** | **83.8** |
| Bachelors (%) | 33.3 | 29.3 | 27.1 | - | 21.6 | 19.0 | - | - | 19.7 | - | 16.2 | 16.8 | - | 16.2 | 16.7 | 16.4 | 15.8 | 15.9 |
| Less than Bachelors (%) | 0.6 | 0.6 | 0.9 | - | 0.1 | 0.8 | - | - | 0.3 | - | 0.2 | 0.2 | - | 0.2 | 0.2 | 0.2 | 0.2 | 0.4 |
| **Total Bachelors or less** | **33.9** | **29.9** | **28.0** | **-** | **21.7** | **19.8** | **-** | **-** | **20.0** | **-** | **16.4** | **17.0** | **-** | **16.4** | **16.9** | **16.6** | **16.0** | **16.3** |
| Fully Credentialed Teachers (%) | 78.7 | 83.3 | 91.4 | - | 91.5 | 95.4 | - | - | 92.3 | - | 93.4 | 95.1 | - | 93.6 | 95.4 | 90.9 | 93.7 | 95.8 |
| Weighted Teacher Qualification [4] | 2.0 | 2.1 | 2.11 | - | 2.2 | 2.26 | - | - | 2.31 | - | 2.3 | 2.27 | - | 2.3 | 2.28 | 2.2 | 2.3 | 2.32 |
| Average years teaching | 10.9 | 11.1 | 11.4 | - | 11.1 | 11.4 | - | - | 11.4 | - | 11.7 | 11.8 | - | 11.8 | 12.0 | 12.7 | 12.8 | 12.8 |

[1] Cohort 2 demographics are provided beginning in 2004 because 2003-2004 was the first year of Reading First Implementation in those schools.
[2] Cohort 3 demographics are provided beginning in 2005 because 2004-2005 was the first year of Reading First Implementation in those schools.
[3] Demographics for the Comparison Group Schools and the RF Eligible Schools are presented only from 2004 and 2005 because these groups were formed in the second year of the study.
[4] The Weighted Teacher Qualification is computed as follows: The percentage of teachers with PhDs is given a weight of 5, the percent of teachers with Masters plus 30 or more semester units is given a weight of 4, the percent of teachers with Masters is given a weight of 3, the percent of teachers with Bachelors plus 30 or more semester units is given a weight of 2 and teachers with Bachelors are given a weight of 1. The weighted degree percents are summed to reach the Weighted Teacher Qualification. This index spans from 1 (lowest qualification) to 5 (highest qualification).
[5] The group "All Elementary Schools" *includes* Reading First schools in this chapter. Whereas in Chapter 4, "All Elementary Schools" *excludes* Reading First schools.

**Non-Reading First Schools**

As pointed out in several sections in this chapter, the Reading First Eligible schools are no longer used as a comparison group in the analysis of Reading First schools, nor referred to as Comparison Group A. The objective in examining demographic information in this chapter is not only to understand the nature of the Reading First schools, but also to justify the use of a specific set of schools for purposes of comparisons in outcome and impact analysis. In general, comparison groups are expected to fill a role similar to a placebo in clinical studies. Control groups are intended to match the treatment group in all respects but the treatment itself. In educational research, while it is generally impossible to create a true control group, it is often possible to create a group that controls for the presence of "predictor" variables that have been found to impact outcomes. The simplest way to control for predictor variables in this study is to identify a group of schools (or students) that most closely resemble the treatment group, the Reading First schools.

Using the demographic data presented earlier in this chapter, we evaluated the use of the two groups as comparisons. That which was called Comparison Group B in previous years' reports—now referred to simply as the Comparison Group—best fills the role of a control group. This group of schools was selected from the population of elementary schools using a two-step process that involved k-means Cluster Analysis and random selection. In the Comparison Group schools, the two demographic variables controlled were percentages of SED and EL students per school. That which was called Comparison Group A in previous reports was randomly selected from the list of schools that were eligible for funding but resided in school districts that had not received any Reading First funding. Had Reading First funding been randomly assigned to the list of eligible schools, it is possible that the RF Eligible (Comparison Group A) schools could have served as comparison schools for the Reading First cohorts. However, the demographic evidence presented in this chapter shows that funding in the first three years of the program went to schools with higher percentages of SED and EL students. As a result, the eligible schools that have not received Reading First funding as of 2005 appear to be quite different from the funded schools in terms of demographic characteristics. Therefore, from the perspective of research design, it is not advisable to compare Reading First schools to the RF Eligible group of schools. In this report, we continue to provide demographic data for the RF Eligible schools but no longer use them in achievement and implementation analyses.

**Conclusions**

The demographic data presented in this chapter yield the following findings:

- The Reading First cohorts had inherent differences in their demographic composition. Cohort 1 had the highest percentages of SED and African American students. Cohort 3 had the highest percentages of EL and Hispanic students.
- Cohort 1 schools largely represented urban city districts. The percentage of rural districts was relatively small in all Reading First cohorts, but when compared among cohorts, Cohort 3 had the highest concentration of rural districts.
- Waiver classrooms (Spanish language instruction) were fewest among Cohort 1 districts and more prevalent among Cohorts 2 and 3.
- Due to pronounced differences in demographics between Reading First cohorts, it is important to study outcome measures by cohort and not combine all three cohorts together.
- That which was previously labeled Comparison Group B was closest, demographically, to the Reading First schools. That which was previously labeled Comparison Group A was quite dissimilar to the Reading First schools. The labels have been modified to reflect this observation. Comparison Group B is now referred to as the "Comparison Group." Comparison Group A schools are now referred to as "Reading First Eligible" schools and are no longer used in comparative analysis.
- Participation in Reading First may assist schools in reducing the number of EL students at low levels of English proficiency. Reading First schools had high percentages of Beginning and Early Intermediate students on the CELDT in the early years of participation in Reading First, but over time moved students out of those two groups. This phenomenon does not seem to be as pronounced in non-Reading First schools. This pattern warrants further analysis.
- Reading First schools showed increases in teachers with advanced education, a trend that was not evident in non-Reading First groups. There was an increase statewide in the number of fully credentialed teachers and this trend was also evident in Reading First schools. Cohort 1 schools had the lowest percentages of well-qualified teachers of Reading First schools.

## References

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

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

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

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

## Chapter 3: Measurement of School Implementation

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

To evaluate the implementation of Reading First in California, it is necessary to construct implementation measures from data gathered in the field specifically for this purpose; no pre-existing data source is currently available. For this effort, Educational Data Systems (EDS) developed three surveys (paper and pencil in 2004, online in 2005) – one each for Reading First teachers, coaches, and principals – and administered them in spring 2004 and 2005. In 2004, approximately 14,000 surveys were returned to yield a response rate between 73% and 82%. The Spring 2005 administration yielded approximately 20,200 surveys, a response rate of 86%. This chapter discusses the organization, scoring, and analysis of the survey data to compute a Reading First Implementation Index (RFII) for each school. It also discusses the role of such information in assessing the efficacy of the Reading First program. Further technical details are provided in Appendix E. Information on program implementation was also collected using focus groups of school officials from randomly selected schools, discussed in Chapter 5.

### Rationale for Measuring Implementation

Why should implementation be measured at all? Is it sufficient to assess Reading First implementation in terms of school-level achievement gains? The answer is no. When assessing the effectiveness of an educational program, two pieces of information are required: a) the effect of the program (e.g., achievement gains); and b) the *presence* or *absence* or *degree of implementation* of the program in question. If it is found that the more a program is implemented, the higher on average the achievement gains are of schools in that program, evidence exists that the program is working. If achievement gains bear no relation to program implementation, no evidence of program efficacy can be claimed (Schiller, 2001).

Unfortunately, there is no easy or obvious way to directly measure the presence, absence, or degree of implementation of Reading First in participant schools and districts, i.e., the degree to which Reading First funding is being applied in the prescribed manner by school officials and teachers. There is no statewide database on school Reading First implementation.

Much of the analysis in Chapter 4 involves comparing the achievement gains of Reading First schools to non-Reading First schools. Unfortunately, such comparisons allow only limited conclusions. Suppose comparable non-Reading First schools receive funding from other programs, employ equivalent instructional materials, and demonstrate that they match or outperform Reading First schools. It would be unfair to claim that this constitutes evidence that the Reading First program is not working; it may in fact be working quite well. The finding would tell us only that there exist other programs that may be working as well or better.

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

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

By the same token, if we were to find that higher implementation corresponds to higher achievement gains, the important policy issue then becomes "How can we ensure that Reading First is systematically implemented in all Reading First schools?" That is why we include a research component to measure the degree of program implementation in each Reading First school. Without such information, in combination with achievement gain scores, it is difficult to truly understand the impact of Reading First.

**Rationale for Using a Survey**

In 2005 there were over 800 Reading First schools in California. To measure implementation in each school, the external evaluator would ideally send trained auditors to observe each Reading First classroom

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

The alternative is to administer a survey. This was done in the spring of 2004 and 2005. The advantage of using a survey is that it is easy to administer and analyze and the respondents (teachers, coaches, principals) are the most knowledgeable regarding what is happening inside their schools and classrooms throughout the school year. Nonetheless, there are limitations that must be addressed in constructing and implementing a survey for this use:

1. The respondents are, to a certain extent, reporting on themselves. This could lead to a substantial upward bias in estimations of school implementation.
2. Similarly, if school officials believe that survey results could be used to reduce or deny funding, there would be a strong incentive to "cheat" on the survey, 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, which would undermine any analysis of program efficacy.
4. In order for a survey to be specific enough to be useful, it needs to tailor its questions to particular types of respondents. For instance, there need to be questions tailored specifically to teachers, coaches, and principals, and to users of Open Court and Houghton Mifflin in the Spanish and English versions. This would seem to impair our ability to compare schools when they have different proportions of each respondent type.
5. To the degree the survey instrument is changed from year to year, results could lose their cross-year comparability.
6. Each question, taken on its own, inevitably carries ambiguities and imprecision. It is often difficult to be clear exactly what dimensional construct is being measured by a question, and whether it is indeed "implementation."

These are legitimate issues and as such, we employed accepted survey analysis models to ameliorate the effects of each issue where possible. Issues 1 and 2 (upward reporting bias) are addressed by measuring schools relative to each other, not in absolute terms. Thus, to the degree all schools suffer the same

upward bias, this bias has no effect when assessing program efficacy. Issue 3 ("cheating"), which causes differential bias, is certainly the most difficult to solve. We address it in part by making it possible for teachers, who are presumably less influenced by funding concerns than principals, to take the survey anonymously. We are careful to include questions whose "correct" answers are not immediately obvious. We also provide numerous opportunities for respondents to rate other respondent types. Coaches rate teachers. Teachers rate coaches. Both rate principals. Nonetheless, this issue is most effectively addressed at the policy level by refraining from using the Reading First Implementation Index (RFII) as a criterion for determining "significant progress." So long as schools and districts realize that the surveys are used for monitoring and research purposes only, and will not be used to deny funding, it is hoped there will be less of an incentive to manipulate the data.

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

The statistical reliability (Cronbach-alpha) of the Reading First Implementation Index is 0.90. (A reliability of 0.85 is widely considered "sufficient" in the field of educational measurement.) The correlation of the RFII across schools between 2004 and 2005 is 0.42. While not large, this correlation suggests, when taking into account changes in respondent samples per school and likely changes in implementation practices from year to year, that the RFII is measuring an objective school-level characteristic. In 2005, a large Spanish-based curriculum section was added to the questionnaire. Yet the mean school RFII of 36 was unchanged across years, lending credence to the claim that our equating procedure has been effective in adjusting for different types of respondents. It is frequently stated that survey instruments are "too subjective" to be used for measurement purposes. In this case, the statistical characteristics of the RFII persuade us that it is approximately as reliable as a typical large-scale student assessment, bearing in mind that it is more vulnerable to "cheating."

**Converting Survey Data into Implementation Measures: An Overview**

Figure 3.1 displays graphically the process by which the RFII was designed and is calculated annually. Described in greater detail in Appendix E, the following steps were followed.

1. Survey Construction. Surveys were written to be administered to Reading First teachers, coaches, and principals. The teacher survey included sections specifically tailored to grade level (K-3) and type of program (Open Court, Houghton Mifflin, Spanish or English).

2. Questions Keyed to Dimensions. The questions were assigned by a subcommittee of the EAG to 17 dimensions or categories that reflect different aspects of program implementation. A given question might appear in multiple dimensions. Each option in the question was rated as to the degree of the dimension in question it signifies. This process is analogous to deciding which option in a multiple-choice question on a student exam corresponds to "correct."

3. Survey Administration. In spring of the second and third years, the surveys were administered (online as of 2005) to all K-3 teachers in Reading First schools, plus Reading First coaches and school principals.

4. Raw Survey Results Reported. In summer of each year, the data were compiled and reported back to schools and districts as raw percentages for each question option. They were reported at the school, district, and state levels. However, results for questions that might be considered "evaluative" were suppressed at the school level and only presented at the district level.

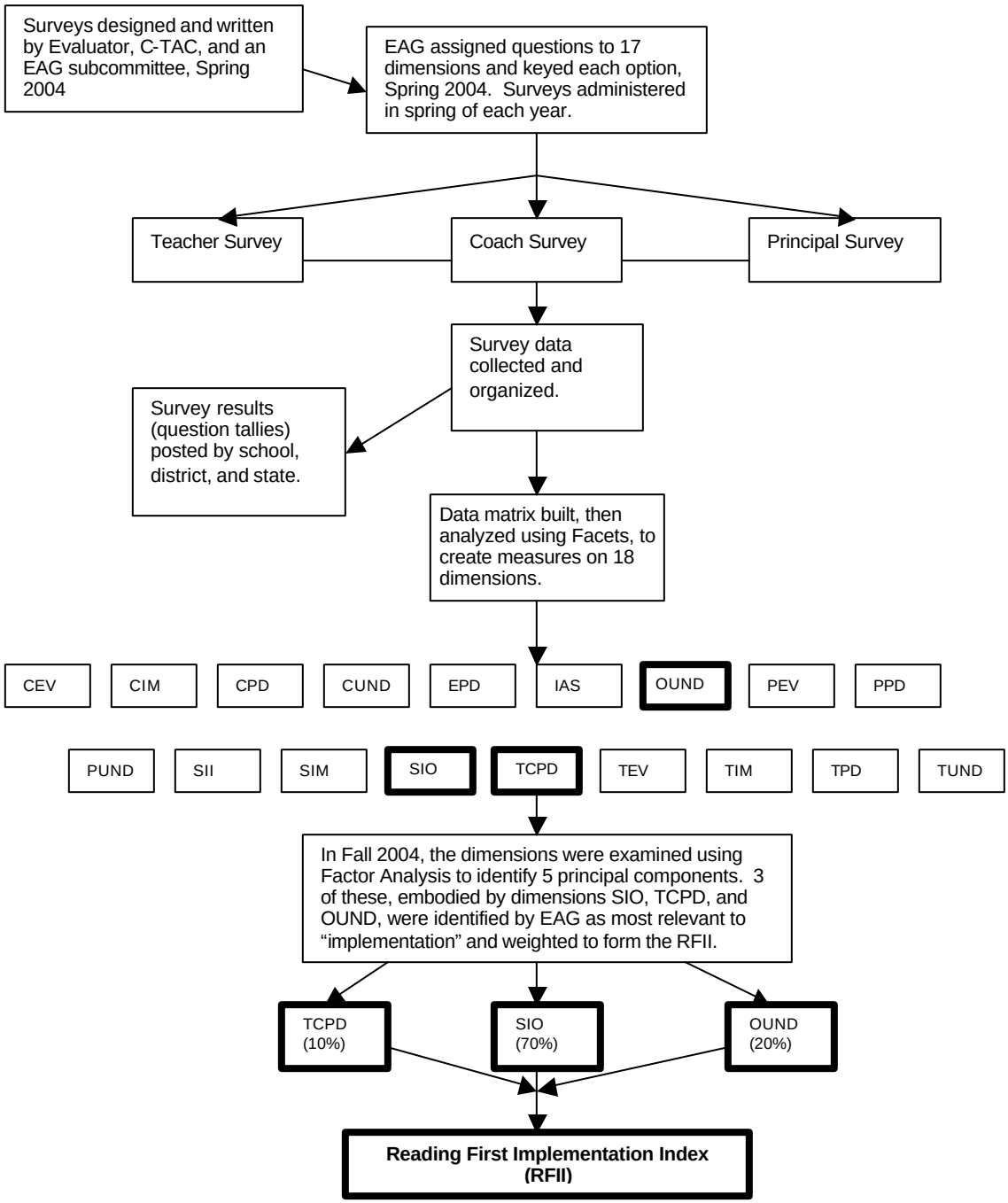
5. Construction of Data Matrix. Concurrently, the surveys were scored and the results collapsed into a matrix suitable for analysis using the Facets statistical program. In Fall 2004 the Facets methodology, described below, was applied to the survey data to verify that all questions keyed to a given dimension do in fact measure along a common construct. Aberrant questions were flagged to be ignored for purposes of measurement, though retained on the survey for informational purposes.

6. Facets Analysis. The data matrix was analyzed using a methodology called Facets, a variant of the Rasch Model which is frequently used to analyze data involving raters or judges. The Facets program was applied to generate measures on 18 dimensions (the original 17 plus a composite teacher/coach professional development dimension).

7. Calculating Each School's RFII Statistic. In Fall 2004, based on the empirical results of a factor analysis performed on the data, the external evaluator and a subcommittee of the EAG identified three of the 18 dmensions as indicators of "implementation" and assigned weights to them. These three dimensions are labeled School Implementation Overall (SIO), Overall Understanding (OUND), and Teacher Coach Professional Development (TCPD), weighted 70%, 20%, and 10%,

respectively. In September, each school RFII statistic was computed by combining its measures on these three dimensions.

**Figure 3.1: Flowchart for the annual computation of the Reading First Implementation Index (RFII)**



**Description of Dimensions**

Table 3.1 lists the dimensions that were identified and keyed by the EAG in May 2004, along with examples of questions that correlate highly to those dimensions. The reader should note that the number of dimensions listed below (18) differs sometimes from those given in other tables in this report and from those given in the Year 2 Report. The variation in number of dimensions reflects the inclusion or omission of "composite" dimensions that are combinations of two or more of the original 17 dimensions identified by the EAG when keying the data. For instance, the dimension listed as Teacher Coach Professional Development combines the Teacher and the Coach professional development. Some of the 2004 composite dimensions proved to be unnecessary and confusing. (A complete listing of each dimension, the abbreviations, and the survey sections that contain the questions keyed to that dimension is found in Appendix E.)

Composite dimensions are created by pooling their survey questions together when analyzing them using the Facets program. This is valid so long as the questions are reasonably correlated with each other, as is the case with teacher professional development and coach professional development. Some dimensions are sufficiently dissimilar that they cannot be combined in this way.

In interpreting the table, the number of questions per dimension is generally close to the number of questions in the relevant survey sections, but is not necessarily the same.

**Table 3.1: List of Dimensions, 2005 Survey, with Question Examples**

| Dimension Description (Number of questions per dimension in parentheses) | Abbre-viations | Sample Questions |
|---|---|---|
| Informational questions (6) | INF | How many years have you been teaching your district's adopted reading/language arts program? |
| Teacher Professional Development (9) | TPD | Which grade level Reading Professional Development Institute did you complete this academic year, 2004 - 05, if any? Select all that apply.<br>What percentage of Reading First teachers (K-3) in your school will have completed the 80-hour follow-up to AB 466 by the end of this school year? |
| Coach Professional Development (6) | CPD | 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?<br>How many hours of follow-up C-TAC Reading First Coach training have you completed this school year? |
| Principal Professional Development (3) | PPD | What training in your district's adopted reading/language arts program have you completed? Select all that apply.<br>How many hours of the 40-hour follow-up to the AB 75 Principal Training Program, Module 1, will you have completed by the end of the school year? |

| Dimension Description (Number of questions per dimension in parentheses) | Abbre-viations | Sample Questions |
|---|---|---|
| Teacher and Coach Professional Development (10) (Combines TPD , CPD) | TCPD | |
| Evaluation of Professional Development (5) | EPD | How well did it prepare you to teach the district's adopted reading/language arts program?<br>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? |
| School Implementation, Assurances (12) | IAS | Has your school established a well-defined vision with goals and objectives for student achievement?<br>Does your school promote the belief that all students can read at grade level if adequately taught? |
| School Implementation, Materials (170) | SIM | Level 2 Themes 1-6 Teachers Editions (Indicate whether the curricular material was: a) Received, b) Used, c) Deemed Effective)<br>Universal Access Handbooks Set Level 2 (Extra Support, Challenge, Classroom Management, Handbook for English Learners) |
| School Implementation, Instruction (Instructional Resources) (28) | SII | How involved is your school principal with the 6-8 week skill assessments?<br>About how frequently do teachers at your grade level have grade-level meetings related to your adopted program? |
| School Implementation Overall (205) | SIO | Open Court Reading Level 1, Books 1A, 1B, 1C, Books 1 and 2 (2000) Level 1, Units 1-10 (2002). (Indicate whether it was: Received, Used, Deemed Effective.)<br>In general, what level of support are you getting from your principal related to your teaching of the adopted reading/language arts program?<br>What is your access to a reading coach?<br>Is your coach helpful in answering questions about how to teach the program?<br>What options do you find to be most effective when students do poorly on the assessments? Select all that apply. |
| Coaching Implementation (29) | CIM | Is your coach helpful in answering questions about how to teach the program?<br>If the coach has conducted one or more demonstration lessons for you, how helpful were they? |
| Teacher Implementation (31) | TIM | To what degree do you follow your school's pacing schedule for reading/language arts?<br>When introducing a decodable book, I have my students: |
| Teacher RF Understanding (Instructional Practices) (15) | TUND | Most of my spelling instruction is focused on:<br>When introducing a decodable book, I have my students: |
| Coach RF Understanding (Instructional Practices) (15) | CUND | Most spelling instruction should be focused on:<br>Vocabulary instruction should focus mainly on: |
| Principal RF Understanding (Instructional Practices) (15) | PUND | Most spelling instruction should be focused on:<br>Vocabulary instruction should focus mainly on: |

| Dimension Description (Number of questions per dimension in parentheses) | Abbre-viations | Sample Questions |
|---|---|---|
| Overall RF Understanding (15) (Combines TUND, CUND, PUND) | OUND | |
| Teacher RF Evaluation (4) | TEV | Overall, how would you rate the effectiveness of your district's adopted reading/language arts program in your school?<br><br>In general, our school is satisfied with the student results we are getting with the district's adopted reading/language arts program. |
| Coach RF Evaluation (6) | CEV | Overall, how would you rate the effectiveness of your district's adopted reading/language arts program in your school?<br><br>In general, our school is satisfied with the student results we are getting with the district's adopted reading/language arts program. |
| Principal RF Evaluation (6) | PEV | Overall, how would you rate the effectiveness of your district's adopted reading/language arts program in your school?<br><br>In general, our school is satisfied with the student results we are getting with the district's adopted reading/language arts program. |

**Changes to the Survey in 2005**

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

The only other major change was the removal of Section H from the teacher survey. This section asked teachers to report their average classroom oral fluency scores, making it possible to examine the relationship between implementation and achievement at the classroom level as well as the school level. However, this method of gathering achievement information was perhaps unreliable and Section H was also burdensome for teachers to fill out.

The remaining changes to the survey were minor editorial corrections and clarifications of the instructions to respondents. Although the addition of the Spanish curricular questions made the survey appear larger, the actual number of questions that any single respondent faced decreased between 2004 and 2005 due to the removal of Section H. Anecdotal information received from teachers and coaches indicated that it took approximately 20 to 30 minutes to complete the survey.

**The Facets Methodology**

There are a number of methods for analyzing survey data. The method used here, the Many-Facet Rasch Model or Facets, is particularly well-suited to judging and equating designs in which there may be 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.

In this case, the surveys have a wide variety of respondent types, each of whom is exposed to a somewhat different version of the survey. Coaches and principals answer different questions than teachers. An Open Court teacher answers different questions than a Houghton Mifflin teacher. A kindergarten teacher answers different questions than a first grade teacher. Respondents in 2005 do not see exactly the same questions as respondents in 2004.

In order to put all schools on the same scale, even though they may consist of different respondent types who take different versions of the survey, it is necessary to use an equating model such as Facets. Equating involves computing the difficulty of each question, the leniency of each rater type, and the implementation level of each school (which is what we are trying to find). This, in turn, is made possible because of common questions that link all the versions of the survey together. The end result is an implementation measure for each school which is adjusted automatically for differences in surveys and respondent types. This allows for "apples to apples" comparisons between schools.

The Facets implementation measures are on a linear scale much like the scale scores used in standardized testing, the preferred metric for measuring growth and performing statistical analysis. For reporting purposes the RFII measures have been converted to a 0-100 metric which can be conceived of as a percentage. This is a little more tangible than a scale score, but a percentage of what? Let us state it simply: The RFII is a (theoretical) percentage of items for which teachers in the school rated their school "more than adequate." If we see that a school gets a "40" on School Implementation Overall (SIO), that loosely means that its teachers rated the school "more than adequate" 40% of the time, that is, on 40% of the questions.

The important thing to note is that Facets provides a range of options on how to define that percentage. The percentage could be the percent of times that teachers rate the school "adequate or better," or "adequate" and "more than adequate." It could be the percent of times that *school principals* rate the school "adequate or better." The size of the percentage will be different in each case. "Adequate or

better" is a more lenient standard than "more than adequate", so it will yield higher percentages. Principals are more lenient in how they rate their school's implementation than teachers, so using their ratings to define the percentage instead of teachers' ratings would also yield higher percentages. Calculating the percentage is a purely mathematical operation using parameters computed by the Facets program. Deciding which parameters should be used to define the percentage is a matter of human judgment.

Therefore, the external evaluator consulted the EAG for assistance in determining how best to set this percentage for this study. The EAG recommended in the 2003-04 study, and the external evaluator agreed, to define the percentage in terms of teachers and the "more than adequate" standard. This procedure was also used in this 2004-05 study. The recommendation was made in part to yield a Reading First Implementation Index (RFII) distribution that would fall in the same range on the 0-100 scale as the Reading First Achievement Index (RFAI) – in the 30s. The EAG also wanted the RFII to hold schools to the higher standard imposed by teachers and by the "more than adequate" criterion. As it happens, the mean RFII across all Reading First schools in California, both in 2004 and 2005, is 36. This can be interpreted to mean that teachers rated their schools "more than adequate" 36% of the time.

Note that the terms "adequate" and "more than adequate" (as well as "less than adequate" and "poor") appear rarely in the survey. This terminology was determined in consultation with the EAG, and it was the Facets program that quantified these attributions across all items on a common scale.

Note also that for those dimensions that combine teacher, coach, and principal data, the same method of defining the percentage was used. Such measures can be interpreted as the percent of items for which teachers found their schools "more than adequate" on the dimension in question. (Coach- or principal-specific dimensions produce measures that can be interpreted as the percent of items for which coaches or principals found their schools "more than adequate" on that dimension.) This chapter and Appendix E report these measures for each dimension, both at the "more than adequate" level and at the "adequate" or better level.

**Factor Analysis as a Step to the RFII**

In order to decide how to distill a single RFII scale from the 17 dimension measures, Factor Analysis was applied to the 17 dimension measures across schools (see Appendix E for details). This revealed which dimensions "hang together" and which do not. It was found that the 17 dimensions could be summarized by five principal components, also known as "factors." In order of their statistical importance in explaining differences between schools, i.e., in order of the amount of variance they explain (see Table E.8, Appendix E), they were:

- Implementation (teachers, coaches, curricular materials, school implementation overall)
- Evaluation (attitudes toward Reading First)
- Understanding (grasp of Reading First pedagogical principles)
- School Principals (training of principals, how they implement the Assurances)
- Professional Development (training of coaches and teachers, but not principals)

These five principal components were sufficient to explain most of the variation between schools. It was gratifying that the most important principal component was associated with the Implementation dimensions, i.e., Teacher Implementation, Coach Implementation, School Implementation Materials, and School Implementation Overall. This supports the expectation that the surveys are indeed capturing implementation differences between schools, as was intended.

The EAG, as experts in the field, again provided advice regarding which principal components best summarized what is meant by "school implementation." It was noted that the evaluation principal component does not involve implementation *per se*, just how much respondents "like" Reading First. The component relating to school principals also did not seem related to implementation. Indeed, the fact that school principals could only be explained by a separate component of their own suggested that they have a highly idiosyncratic view of their role in Reading First which does not correlate with the views of teachers and coaches, and therefore may be unreliable.

The remaining three components, labeled implementation, understanding, and professional development, all seemed very relevant to the EAG's understanding of implementation. Facets dimensions were identified that best summarize these principal components: School Implementation Overall, Overall Understanding, and Teacher Coach Professional Development. Facets dimensions were chosen instead of principal component "factor scores" because of their higher statistical stability; they are less prone to cause anomalous shifts as the respondent sample changes over time.

**Computation of the RFII**

To summarize, we embody the School Implementation *component* with the SIO *dimension* (School Implementation Overall, which combines implementation as it relates to Materials, Instruction, Teacher Implementation, and Coach Implementation). We embody the Reading First Understanding *component* with a composite *dimension* called OUND (Overall Reading First Understanding, which combines Principal Understanding, Coach Understanding, and Teacher Understanding). We embody the Professional Development *component* with a composite *dimension* called TCPD (Teacher Coach

Professional Development, which combines Teacher Professional Development and Coach Professional Development).

The external evaluator, with approval of the EAG, then assigned weights to the three dimensions with which to compute the Reading First Implementation Index. The weights recommended by the EAG were:

School Implementation Overall (SIO) = 70%

Overall Reading First Understanding (OUND) = 20%

Teacher Coach Professional Development (TCPD) = 10%

The formula for computing each school's RFII is therefore:

$$RFII_{schoolX} = 0.70*SIO_{schoolX} + 0.20*OUND_{schoolX} + 0.10*TCPD_{schoolX}$$

The school RFII is the weighted average of its School Implementation Overall (SIO) measure, its Overall Understanding (OUND) measure, and its Teacher Coach Professional Development (TCPD) measure.

**Distribution and Interpretation of the RFII**

Figure 3.2.1 shows how the RFII was distributed across schools in all three cohorts in 2005 (the 2004-2005 school year). The mean RFII was 36; the standard deviation around the mean was 5. The 2005 distribution is very similar to that of 2004, whose mean and standard deviation were 36 and 6, respectively. As stated above, this can be interpreted as follows: California teachers on average found their schools to be "more than adequate" 36% of the time, i.e., on 36% of the relevant items. For coach-specific and principal-specific dimensions (indicated in the footnotes to Table 3.2), the points of reference are not teachers but coaches and principals. The measures on the far right tail of the distribution (above 55) should be viewed with skepticism; a perusal of the surveys for those schools suggests that respondents were coached to provide the "right" answers to the survey questions. This underlines the importance of reducing incentives to "cheat."

Interpreting the RFII as a percentage of items has its pitfalls. Strictly speaking, the RFII is a probability that teachers in a school will rate their school "more than adequate" on an item of average difficulty. It is a theoretical abstraction. Interpreting it as a percentage of items makes it more familiar, but it should not be interpreted literally.

Figures 3.2.2 through 3.2.4 break out the school implementation distribution by cohort. Cohorts 1 and 2 have similar distributions. Cohort 3 shows a bimodal distribution, with a second peak to the left. This shows that Cohort 3 is having implementation problems for a subset of schools as discussed below.

**Figure 3.2.1: All Cohorts - 2005 Reading First Implementation Index (RFII), distribution of schools**

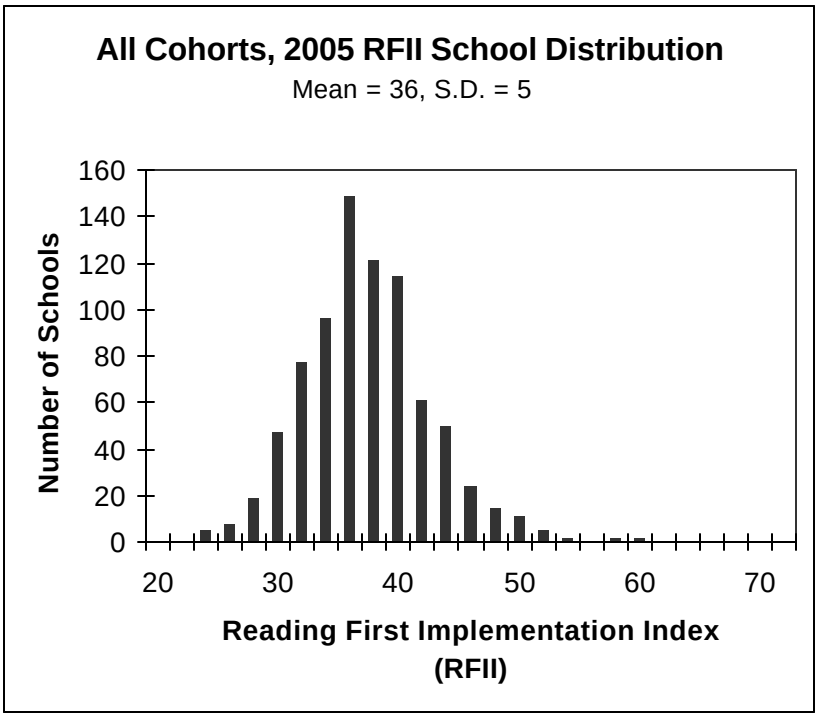



**Figure 3.2.2: Cohort 1 – 2005 Reading First Implementation Index (RFII), distribution of schools**

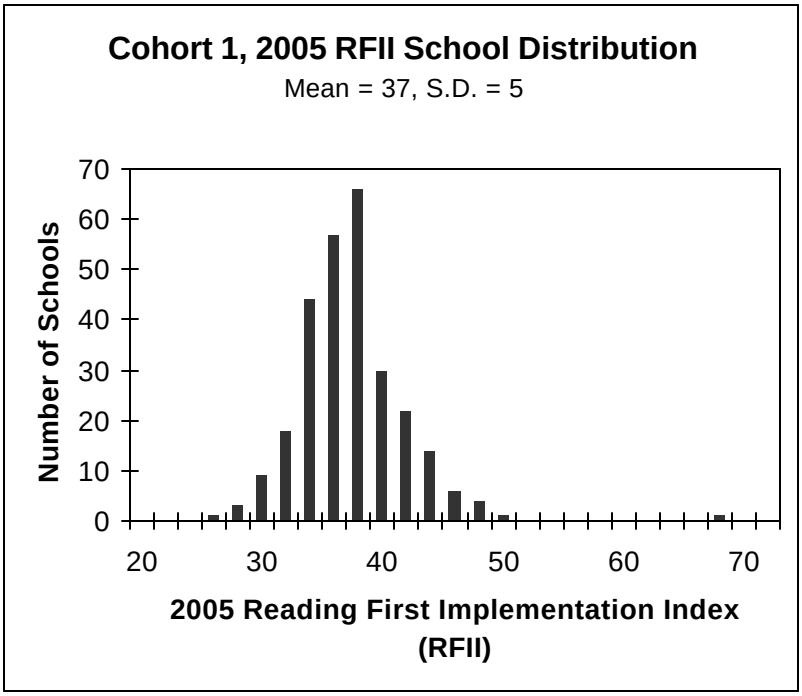


**Figure 3.2.3: Cohort 2 – 2005 Reading First Implementation Index (RFII), distribution of schools**

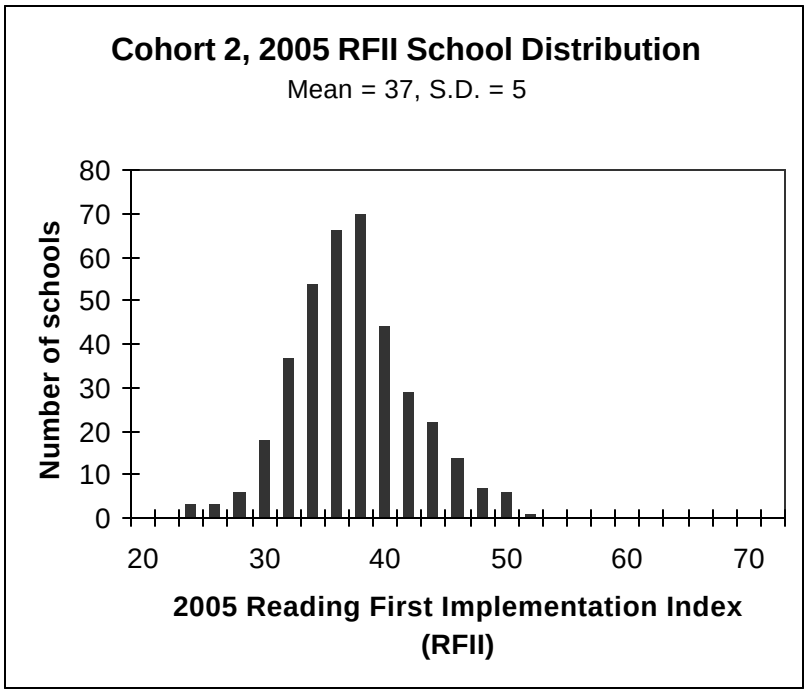



**Figure 3.2.4: Cohort 3 – 2005 Reading First Implementation Index (RFII), distribution of schools**

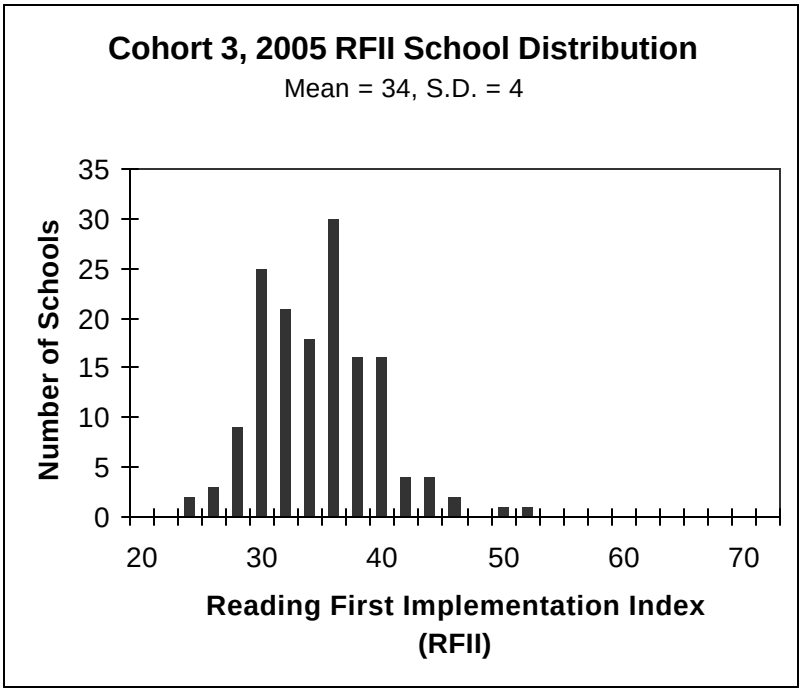


Table 3.2 reports the mean school RFII in the bottom right row for 2004 and 2005, with standard deviations. It also reports the means and standard deviations for the 18 remaining dimensions for both years. The columns on the left, for those dimensions rated by teachers, may be interpreted as the percent of times (items) teachers rated their school "adequate" or better on that dimension, averaged across schools. The columns on the right employ the tougher standard used in the RFII: the percent of times (items) teachers rated their school "more than adequate" on that dimension, averaged across schools.

Table 3.2 shows that implementation was virtually unchanged between 2004 and 2005, not only for the RFII (Dimension 18) but for all the dimensions. The stability of the measures, despite significant changes to the survey and to the rater population, supports the claim that our equating methodology has so far been effective, at least at the state level, in protecting the RFII and the other survey dimensions from irrelevant disturbances. The error at the level of the individual school is, however, much greater. It is possible that RFII stability will diminish over time as respondents (especially coaches and principals) are exposed to the survey multiple times.

We see that the coaching dimensions (Coach Professional Development, Coach Implementation, Coach Understanding, and the Coach Evaluation of Reading First) are very strong. This is unlikely to be an artifact of self-reporting since coach scores are largely derived from teachers. Given the central role of coaches in Reading First implementation, this is encouraging. We also see that the Implementation of Assurances dimension is high, but this is more likely to be an artifact of self-reporting as teachers do not answer these questions.

There is an interesting pattern in the Principal RF Understanding dimension. School principals are on par with teachers and coaches when held against the "adequate" or better standard. When held to the "more than adequate" standard, principals perform quite poorly (17 in 2004, 19 in 2005). This dimension comes from the section of the survey in which teachers, coaches, and principals are tested on their knowledge of specific pedagogical techniques. It appears that principals do not have the level of detailed pedagogical knowledge that would allow them to perform "more than adequately." Nonetheless, they do show growth from 2004 to 2005. While not as weak as principals, teachers also show weakness in the RF Understanding dimension at the "more than adequate" level.

**Table 3.2: All Cohorts, Mean and Standard Deviation of Each Dimension, 2004 - 2005**

| | | | % of the time teachers rated their school "adequate" or better, averaged across schools | | | | % of the time teachers rated their school "more than adequate", averaged across schools | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2004 | | 2005 | | 2004 | | 2005 | |
| | Dimension[1] | # Items, 2005 | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | Teacher Professional Development | 9 | 63 | 14 | 62 | 10 | 38 | 14 | 36 | 10 |
| 2 | Coach Professional Development | 6 | 74 | 19 | 72 | 19 | 58 | 22 | 56 | 21 |
| 3 | Principal Professional Development | 3 | 64 | 29 | 61 | 30 | 48 | 30 | 46 | 30 |
| **4** | **Teacher Coach Professional Development** | 10 | 64 | 15 | 63 | 11 | **40** | **16** | **37** | **11** |
| 5 | Evaluation of Professional Development | 5 | 76 | 12 | 80 | 9 | 11 | 6 | 14 | 7 |
| 6 | Implementation, Assurances | 12 | 69 | 15 | 73 | 13 | 44 | 18 | 48 | 16 |
| 7 | School Implementation, Materials | 170 | 57 | 11 | 58 | 9 | 36 | 10 | 37 | 9 |
| 8 | School Implementation, Instruction | 28 | 58 | 6 | 59 | 6 | 34 | 6 | 36 | 6 |
| **9** | **School Implementation Overall** | 205 | 58 | 7 | 58 | 6 | **39** | **7** | **40** | **6** |
| 10 | Coaching Implementation | 29 | 70 | 14 | 73 | 12 | 46 | 16 | 48 | 14 |
| 11 | Teacher Implementation | 31 | 70 | 4 | 72 | 4 | 48 | 5 | 50 | 5 |
| 12 | Teacher RF Understanding | 15 | 56 | 7 | 58 | 6 | 27 | 6 | 29 | 5 |
| 13 | Coach RF Understanding | 15 | 69 | 14 | 72 | 12 | 36 | 15 | 39 | 14 |
| 14 | Principal RF Understanding | 15 | 59 | 15 | 64 | 14 | 17 | 9 | 19 | 10 |
| **15** | **Overall RF Understanding** | 15 | 57 | 7 | 59 | 7 | **23** | **5** | **25** | **5** |
| 16 | Teacher RF Evaluation | 4 | 59 | 13 | 61 | 12 | 14 | 7 | 14 | 7 |
| 17 | Coach RF Evaluation | 6 | 74 | 18 | 75 | 18 | 20 | 18 | 19 | 18 |
| 18 | Principal RF Evaluation | 6 | 83 | 17 | 84 | 17 | 23 | 24 | 24 | 24 |
| **19** | **Composite Implementation Official RFIIs are in right-hand columns**[2] | 230 | 58 | 6 | 59 | 5 | **36** | **6** | **36** | **5** |

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

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

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

There has been no official effort to set an "adequate" or "more than adequate" cut-point on the RFII scale. Nonetheless, the initial keying of the rating scale categories for each question makes it possible for Facets to provide an empirical way to define "adequate." A school is considered "adequate" or better if its dimension measure falls above the "adequate" cut-point as judged by teachers. A school is considered

"more than adequate" if its dimension measure falls above the "more than adequate" cut-point as judged by teachers. Conveniently, Facets sets the dividing line automatically at a dimension measure of 50 (signifying a 50% chance or greater of teachers rating a school as "adequate" or "more than adequate"). Looking at Table 3.2 and Table 3.3 together, we see first in Table 3.2 that schools measured on Teacher Professional Development received an average of 62 in 2005 (left-hand columns, first row) and that these schools spread out above and below this mean as indicated by the standard deviation. Then in Table 3.3 we see that 92% of those schools fall above the 50 mark, the rest falling below (left-hand columns, first row). Held to the higher standard of "more than adequate," the average Teacher Professional Development measure is 36 in 2005 (right-hand columns of Table 3.2, first row). By this standard, only 9% of schools fall above the 50 mark according to Table 3.3 (right-hand columns, first row).

Table 3.3 reports, for each year and dimension, the percentage of schools that are "adequate" or better (left two columns), and "more than adequate" (right two columns). Note that the two percentages can sum to more than 100% since a school can be both "adequate" or better and "more than adequate." In general, as was true in 2004, we can say that teachers perceive the great majority of schools to be at least "adequate" on the various dimensions. In fact, the percentage of schools at the "adequate" or better standard on the composite RFII dimension rose from 92% to 96% from 2004 to 2005. Also of particular interest, very few schools (1%) register as "more than adequate," and the Professional Development dimensions and the Coaching/Teacher Implementation dimensions are registered as the strongest dimensions.

We voice an important caution about this data. Table 3.3 and the other tables that report how many schools fall above the "adequate" and "more than adequate" cut-points are heavily influenced by the error of the measures in each dimension, which is driven by the number of items it contains. When the number of items is small the error tends to be large, which spreads the schools more widely across the scale. This will cause more schools to fall above or below the 50% cut-point than might otherwise be the case, just by chance. In general, percentages on dimensions with fewer than ten items should be interpreted with caution.

By the same token, measures on dimensions that have numerous items (such as the RFII) will be more precise, and thus tend to cluster more closely together on the scale. Random error is less likely to spread the items out on the scale. That is part of the reason why only 1% of schools meet the "more than adequate" criterion. The high precision of the dimension means that fewer schools cross into the "more than adequate" territory by chance. The school RFII distribution is bracketed squarely within the "adequate" section of the survey rating scale.

**Table 3.3: All Cohorts, % of Schools Considered "Adequate" or better, and "More than Adequate", by Teachers, 2004 - 2005**

| | Dimension[1] | # Items, 2005 | % of Schools Considered "adequate" or better by Teachers | | % of Schools Considered "more than adequate" by Teachers | |
|---|---|---|---|---|---|---|
| | | | 2004[2] | 2005 | 2004 | 2005 |
| 1 | Teacher Professional Development | 9 | 87% | 92% | 17% | 9% |
| 2 | Coach Professional Development | 6 | 90% | 90% | 68% | 67% |
| 3 | Principal Professional Development | 3 | 67% | 69% | 42% | 41% |
| **4** | **Teacher Coach Professional Development** | **10** | **89%** | **92%** | **20%** | **12%** |
| 5 | Evaluation of Professional Development | 5 | 97% | 99% | 0% | 1% |
| 6 | Implementation Assurances | 12 | 88% | 94% | 32% | 41% |
| 7 | School Implementation Material | 170 | 76% | 83% | 10% | 7% |
| 8 | School Implementation Instruction | 28 | 90% | 93% | 0% | 1% |
| **9** | **School Implementation Overall** | **205** | **88%** | **92%** | **6%** | **5%** |
| 10 | Coaching Implementation | 29 | 92% | 96% | 39% | 46% |
| 11 | Teacher Implementation | 31 | 100% | 100% | 35% | 50% |
| 12 | Teacher RF Understanding | 15 | 79% | 89% | 0% | 0% |
| 13 | Coach RF Understanding | 15 | 87% | 94% | 18% | 20% |
| 14 | Principal RF Understanding | 15 | 73% | 80% | 1% | 0% |
| **15** | **Overall RF Understanding** | **15** | **83%** | **92%** | **0%** | **0%** |
| 16 | Teacher RF Evaluation | 4 | 77% | 82% | 0% | 0% |
| 17 | Coach RF Evaluation | 6 | 90% | 91% | 5% | 6% |
| 18 | Principal RF Evaluation | 6 | 93% | 94% | 19% | 20% |
| **19** | **Composite Implementation** | **230** | **92%** | **96%** | **1%** | **1%** |

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

### Implementation Broken Out by Cohort

Tables 3.4 through 3.9 break out Tables 3.2 and 3.3 by cohort, i.e., the year in which Reading First schools and districts first received Reading First funding. As noted in Chapters 2 and 4, the three Reading First cohorts (2003, 2004, and 2005) are by no means the same demographically. Tables 3.4 through 3.9 show how the cohorts also differ in how they implement the program. In particular (see Tables 3.8 and 3.9), Cohort 3 has a notably lower RFII statistic for its first year (34) than Cohorts 1 or 2 did in their first years (36). This is especially evident in the reduced levels of Teacher and Coach Professional Development (Dimension 4) that Cohort 3 respondents are reporting. Given the relationship between implementation and achievement gains found in Chapter 4, this suggests that Cohort 3 may be expected to suffer lower achievement gains in their second and third years than Cohorts 1 and 2 have. It

is unclear why Cohort 3 is having trouble, but the remedy seems straightforward. Cohort 3 teachers and coaches need extra support and encouragement from their schools, districts, and R-TAC's in obtaining professional development.

**Table 3.4: Cohort 1 (2003), Mean and Standard Deviation of Each Dimension, 2004 - 2005**

| | | | % of the time teachers rated their school "adequate" or better, averaged across schools | | | | % of the time teachers rated their school "more than adequate", averaged across schools | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2004 | | 2005 | | 2004 | | 2005 | |
| | Dimension[1] | # Items | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | Teacher Professional Development | 9 | 62 | 9 | 63 | 9 | 36 | 9 | 37 | 9 |
| 2 | Coach Professional Development | 6 | 75 | 20 | 72 | 21 | 60 | 23 | 57 | 23 |
| 3 | Principal Professional Development | 3 | 61 | 28 | 55 | 31 | 44 | 28 | 40 | 28 |
| **4** | **Teacher Coach Prof. Dev.** | 10 | 64 | 9 | 64 | 9 | **37** | **10** | **37** | **10** |
| 5 | Evaluation of Prof. Dev. | 5 | 79 | 10 | 81 | 8 | 13 | 6 | 14 | 7 |
| 6 | Implementation, Assurances | 12 | 70 | 15 | 73 | 12 | 44 | 17 | 47 | 14 |
| 7 | School Implementation, Materials | 170 | 54 | 8 | 55 | 8 | 33 | 8 | 33 | 8 |
| 8 | School Implementation, Instruction | 28 | 59 | 5 | 60 | 5 | 36 | 5 | 36 | 5 |
| **9** | **School Implementation Overall** | 205 | 58 | 6 | 58 | 6 | **39** | **6** | **39** | **6** |
| 10 | Coaching Implementation | 29 | 72 | 13 | 74 | 11 | 47 | 14 | 50 | 13 |
| 11 | Teacher Implementation | 31 | 72 | 4 | 73 | 4 | 50 | 5 | 51 | 5 |
| 12 | Teacher RF Understanding | 15 | 59 | 6 | 60 | 6 | 29 | 5 | 31 | 5 |
| 13 | Coach RF Understanding | 15 | 73 | 13 | 74 | 11 | 40 | 15 | 41 | 14 |
| 14 | Principal RF Understanding | 15 | 65 | 14 | 67 | 14 | 20 | 10 | 22 | 10 |
| **15** | **Overall RF Understanding** | 15 | 61 | 6 | 62 | 6 | **26** | **5** | **27** | **5** |
| 16 | Teacher RF Evaluation | 4 | 61 | 13 | 60 | 13 | 15 | 7 | 14 | 8 |
| 17 | Coach RF Evaluation | 6 | 78 | 17 | 76 | 18 | 22 | 20 | 21 | 19 |
| 18 | Principal RF Evaluation | 6 | 85 | 16 | 83 | 18 | 26 | 25 | 24 | 25 |
| **19** | **Composite Implementation Official RFIIs are in right-hand columns**[2] | 230 | 59 | 5 | 59 | 5 | **36** | **5** | **37** | **5** |

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

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

**Table 3.5: Cohort 1 (2003), % of Schools "Adequate" or better and "More than Adequate," 2004 - 2005**

| | | | % of Schools Considered "adequate" or better by Teachers | | % of Schools Considered "more than adequate" by Teachers | |
|---|---|---|---|---|---|---|
| | Dimension | # Items | 2004 | 2005 | 2004 | 2005 |
| 1 | Teacher Professional Development | 9 | 91% | 94% | 7% | 9% |
| 2 | Coach Professional Development | 6 | 90% | 89% | 68% | 68% |
| 3 | Principal Professional Development | 3 | 63% | 60% | 35% | 33% |
| **4** | **Teacher Coach Prof. Dev.** | **10** | **95%** | **96%** | **10%** | **12%** |
| 5 | Evaluation of Prof. Dev. | 5 | 99% | 100% | 0% | 0% |
| 6 | Implementation Assurances | 12 | 88% | 97% | 36% | 38% |
| 7 | School Implementation Material | 170 | 71% | 71% | 3% | 4% |
| 8 | School Implementation Instruction | 28 | 96% | 96% | 0% | 1% |
| **9** | **School Implementation Overall** | **205** | **92%** | **92%** | **3%** | **6%** |
| 10 | Coaching Implementation | 29 | 96% | 97% | 41% | 52% |
| 11 | Teacher Implementation | 31 | 100% | 100% | 47% | 60% |
| 12 | Teacher RF Understanding | 15 | 90% | 94% | 0% | 0% |
| 13 | Coach RF Understanding | 15 | 92% | 97% | 24% | 25% |
| 14 | Principal RF Understanding | 15 | 84% | 88% | 2% | 0% |
| **15** | **Overall RF Understanding** | **15** | **92%** | **96%** | **0%** | **0%** |
| 16 | Teacher RF Evaluation | 4 | 81% | 80% | 0% | 1% |
| 17 | Coach RF Evaluation | 6 | 95% | 92% | 8% | 7% |
| 18 | Principal RF Evaluation | 6 | 94% | 92% | 24% | 21% |
| **19** | **Composite Implementation** | **230** | **97%** | **97%** | **0%** | **1%** |

**Table 3.6: Cohort 2 (2004), Mean and Standard Deviation of Each Dimension, 2004 - 2005**

| | | | % of the time teachers rated their school "adequate" or better, averaged across schools | | | | % of the time teachers rated their school "more than adequate", averaged across schools | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2004 | | 2005 | | 2004 | | 2005 | |
| | Dimension[1] | # Items | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | Teacher Professional Development | 9 | 63 | 17 | 64 | 9 | 39 | 17 | 38 | 10 |
| 2 | Coach Professional Development | 6 | 73 | 18 | 73 | 17 | 57 | 21 | 56 | 19 |
| 3 | Principal Professional Development | 3 | 65 | 30 | 67 | 28 | 50 | 31 | 52 | 29 |
| **4** | **Teacher Coach Prof. Dev.** | 10 | 65 | 18 | 65 | 9 | **41** | **20** | **38** | **10** |
| 5 | Evaluation of Prof. Dev. | 5 | 74 | 13 | 80 | 8 | 11 | 6 | 14 | 7 |
| 6 | Implementation, Assurances | 12 | 68 | 15 | 76 | 13 | 43 | 18 | 53 | 17 |
| 7 | School Implementation, Materials | 170 | 60 | 12 | 62 | 8 | 38 | 12 | 40 | 9 |
| 8 | School Implementation, Instruction | 28 | 57 | 7 | 59 | 6 | 34 | 6 | 35 | 6 |
| **9** | **School Implementation Overall** | 205 | 57 | 8 | 60 | 6 | **39** | **8** | **41** | **7** |
| 10 | Coaching Implementation | 29 | 69 | 15 | 73 | 12 | 45 | 17 | 48 | 15 |
| 11 | Teacher Implementation | 31 | 69 | 5 | 71 | 4 | 47 | 5 | 50 | 5 |
| 12 | Teacher RF Understanding | 15 | 53 | 6 | 56 | 6 | 25 | 5 | 27 | 5 |
| 13 | Coach RF Understanding | 15 | 66 | 14 | 71 | 12 | 33 | 15 | 38 | 13 |
| 14 | Principal RF Understanding | 15 | 57 | 15 | 61 | 14 | 15 | 8 | 18 | 9 |
| **15** | **Overall RF Understanding** | 15 | 54 | 7 | 58 | 6 | **21** | **5** | **23** | **4** |
| 16 | Teacher RF Evaluation | 4 | 58 | 13 | 62 | 12 | 13 | 6 | 15 | 7 |
| 17 | Coach RF Evaluation | 6 | 72 | 19 | 76 | 16 | 17 | 16 | 20 | 19 |
| 18 | Principal RF Evaluation | 6 | 82 | 18 | 85 | 16 | 22 | 24 | 24 | 24 |
| **19** | **Composite Implementation Official RFIIs are in right-hand columns**[2] | 230 | 57 | 7 | 60 | 5 | **36** | **6** | **37** | **5** |

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

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

**Table 3.7: Cohort 2 (2004), % of Schools "Adequate" or better and "More than Adequate," 2004 - 2005**

| | | | % of Schools Considered "adequate" or better by Teachers | | % of Schools Considered "more than adequate" by Teachers | |
|---|---|---|---|---|---|---|
| | Dimension | # Items | 2004 | 2005 | 2004 | 2005 |
| 1 | Teacher Professional Development | 9 | 84% | 94% | 24% | 11% |
| 2 | Coach Professional Development | 6 | 89% | 93% | 68% | 70% |
| 3 | Principal Professional Development | 3 | 69% | 78% | 46% | 49% |
| **4** | **Teacher Coach Professional Development** | **10** | **85%** | **94%** | **28%** | **13%** |
| 5 | Evaluation of Professional Development | 5 | 95% | 99% | 0% | 1% |
| 6 | Implementation Assurances | 12 | 88% | 96% | 29% | 54% |
| 7 | School Implementation Material | 170 | 80% | 94% | 15% | 12% |
| 8 | School Implementation Instruction | 28 | 85% | 93% | 1% | 1% |
| **9** | **School Implementation Overall** | **205** | **84%** | **93%** | **9%** | **7%** |
| 10 | Coaching Implementation | 29 | 90% | 95% | 37% | 44% |
| 11 | Teacher Implementation | 31 | 99% | 100% | 26% | 47% |
| 12 | Teacher RF Understanding | 15 | 69% | 87% | 0% | 0% |
| 13 | Coach RF Understanding | 15 | 83% | 94% | 12% | 16% |
| 14 | Principal RF Understanding | 15 | 68% | 75% | 0% | 0% |
| **15** | **Overall RF Understanding** | **15** | **75%** | **92%** | **0%** | **0%** |
| 16 | Teacher RF Evaluation | 4 | 74% | 85% | 0% | 0% |
| 17 | Coach RF Evaluation | 6 | 86% | 93% | 3% | 7% |
| 18 | Principal RF Evaluation | 6 | 92% | 96% | 17% | 21% |
| **19** | **Composite Implementation** | **230** | **87%** | **97%** | **1%** | **2%** |

**Table 3.8: Cohort 3 (2005), Mean and Standard Deviation of Each Dimension, 2004 - 2005**

| | | | % of the time teachers rated their school "adequate" or better, averaged across schools | | | | % of the time teachers rated their school "more than adequate", averaged across schools | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2004 | | 2005 | | 2004 | | 2005 | |
| | Dimension[1] | # Items | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | Teacher Professional Development | 9 | . | . | 56 | 14 | . | . | 31 | 12 |
| 2 | Coach Professional Development | 6 | . | . | 69 | 18 | . | . | 52 | 21 |
| 3 | Principal Professional Development | 3 | . | . | 63 | 33 | . | . | 49 | 33 |
| **4** | **Teacher Coach Prof. Dev.** | 10 | . | . | 56 | 14 | . | . | **31** | **12** |
| 5 | Evaluation of Prof. Dev. | 5 | . | . | 77 | 10 | . | . | 12 | 9 |
| 6 | Implementation, Assurances | 12 | . | . | 64 | 15 | . | . | 37 | 15 |
| 7 | School Implementation, Materials | 170 | . | . | 59 | 7 | . | . | 37 | 7 |
| 8 | School Implementation, Instruction | 28 | . | . | 57 | 7 | . | . | 33 | 6 |
| **9** | **School Implementation Overall** | 205 | . | . | 56 | 5 | . | . | **38** | **5** |
| 10 | Coaching Implementation | 29 | . | . | 69 | 12 | . | . | 44 | 14 |
| 11 | Teacher Implementation | 31 | . | . | 69 | 5 | . | . | 47 | 5 |
| 12 | Teacher RF Understanding | 15 | . | . | 55 | 6 | . | . | 26 | 5 |
| 13 | Coach RF Understanding | 15 | . | . | 67 | 14 | . | . | 35 | 15 |
| 14 | Principal RF Understanding | 15 | . | . | 60 | 15 | . | . | 17 | 10 |
| **15** | **Overall RF Understanding** | 15 | . | . | 56 | 6 | . | . | **22** | **4** |
| 16 | Teacher RF Evaluation | 4 | . | . | 59 | 12 | . | . | 13 | 7 |
| 17 | Coach RF Evaluation | 6 | . | . | 66 | 19 | . | . | 12 | 12 |
| 18 | Principal RF Evaluation | 6 | . | . | 84 | 14 | . | . | 20 | 18 |
| **19** | **Composite Implementation Official RFIIs are in right-hand columns**[2] | 230 | . | . | 56 | 5 | . | . | **34** | **4** |

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

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

**Table 3.9: Cohort 3 (2005), % of Schools "Adequate" or better and "More than Adequate," 2004 - 2005**

| | | | % of Schools Considered "adequate" or better by Teachers | | % of Schools Considered "more than adequate" by Teachers | |
|---|---|---|---|---|---|---|
| | Dimension | # Items | 2004 | 2005 | 2004 | 2005 |
| 1 | Teacher Professional Development | 9 | . | 78% | . | 5% |
| 2 | Coach Professional Development | 6 | . | 82% | . | 55% |
| 3 | Principal Professional Development | 3 | . | 72% | . | 43% |
| **4** | **Teacher Coach Professional Development** | **10** | . | **76%** | . | **6%** |
| 5 | Evaluation of Professional Development | 5 | . | 98% | . | 1% |
| 6 | Implementation Assurances | 12 | . | 80% | . | 13% |
| 7 | School Implementation Material | 170 | . | 89% | . | 2% |
| 8 | School Implementation Instruction | 28 | . | 82% | . | 0% |
| **9** | **School Implementation Overall** | **205** | . | **84%** | . | **0%** |
| 10 | Coaching Implementation | 29 | . | 92% | . | 32% |
| 11 | Teacher Implementation | 31 | . | 100% | . | 31% |
| 12 | Teacher RF Understanding | 15 | . | 78% | . | 0% |
| 13 | Coach RF Understanding | 15 | . | 85% | . | 18% |
| 14 | Principal RF Understanding | 15 | . | 73% | . | 1% |
| **15** | **Overall RF Understanding** | **15** | . | **78%** | . | **0%** |
| 16 | Teacher RF Evaluation | 4 | . | 78% | . | 0% |
| 17 | Coach RF Evaluation | 6 | . | 80% | . | 1% |
| 18 | Principal RF Evaluation | 6 | . | 96% | . | 14% |
| **19** | **Composite Implementation** | **230** | . | **90%** | . | **0%** |

It is instructive to compare the dimension measures of Cohort 2 for 2004 (its first year) with the dimension measures for Cohort 3 for 2005 (Cohort 3's first year). Comparisons with Cohort 1 are problematic because survey data were only gathered after Cohort 1 had already been exposed to Reading First for a year, though this exposure was uneven due to variation regarding when the funding became available. We have already noted that Cohort 3 has begun its first year at a lower level of implementation (RFII = 34) than did Cohorts 1 and 2 in their early years (RFII = 36). Why is this? The evaluation dimensions (how respondents feel about Reading First, seen in Dimensions 16-18) show that Cohort 3 respondents are as favorably disposed to Reading First as Cohort 2 respondents. The Reading First Understanding dimensions (see Dimensions 12-15) are no lower for Cohort 3 than they were for Cohort 2 in 2004. The same can be said for the Teacher and Coach Implementation dimensions. Dimensions for which Cohort 3 falls lower include School Implementation Materials (Dimension 7), Implementation of Assurances (Dimension 6), and Teacher/Coach Professional Development (Dimension 4).

The Materials dimension, even though it shows a lag of only a point or so, is important because it is the most dominant and reliable of the RFII components. It addresses how teachers are actually using the curricular materials in their classroom. The Assurances dimension lag is larger but less reliable. It is intriguing because it suggests that Cohort 3 LEA's are not providing the same level of district support to their schools that Cohorts 1 and 2 did. However, the largest lag is clearly professional development – 10 points when compared to Cohort 2 in its first year. For some reason, Cohort 3 teachers and coaches are getting a slow start in obtaining the Reading First training they need. The demographic data in Chapter 2 show that there are more rural districts in Cohort 3, perhaps hampering access to training facilities. C-TAC also reports greater teacher reluctance to undergo AB 466 training. This is an issue that warrants further study in future evaluation reports, and it may dampen achievement gains in 2006.

In comparing the three cohorts together, we see that Cohort 1 began its second year at a fairly high level of implementation. Cohort 2 lagged Cohort 1 on several dimensions in its first year, but caught up impressively in its second year. Cohort 3 lags still further in its opening year. This suggests a trend which should be checked in subsequent reports: the later the cohort, the more difficulty it appears to have getting up to speed on professional development and use of curricular materials. This may be a demographic artifact. As Chapter 2 points out, the Cohort 1 schools are predominantly urban and reportedly more aggressive and practiced in pursuing grant money. Subsequent cohorts are more rural, perhaps less aggressive at the LEA level in implementing the Assurances, and they may have special logistical problems in accessing the professional development they need. However, based on Cohort 2's second year implementation surge (the RFII dimension at the "adequate and above" level moves from 57 to 60 from 2004 to 2005), we may hope to see Cohort 3 also surge in its second and third years as teachers and coaches become acclimated to the program and as LEA administrators implement the Assurances. If such a trend should occur, then based on the findings in Chapter 4 we may expect to see a concomitant rise in Grade 2 CST scores and, to a lesser extent, in the other achievement metrics. Otherwise, we will see a dampening of achievement growth.

Another finding is that Cohort 1 and Cohort 2 have both shown implementation gains from 2004 to 2005. Both increased their RFII scores from 36 to 37. Cohort 2, as mentioned, displayed an even greater increase, from 57 to 60, at the "adequate" or better standard of implementation. Cohort 2's gains are distributed across the dimensions, the only exception being Teacher Coach Professional Development which is flat (perhaps not unreasonable as teachers settle into the program). The Assurances Implementation dimension shows notable growth, supporting the link between LEA involvement and implementation at the school level. It is also notable that Cohort 2 teachers, coaches, and principals all gave stronger evaluations of Reading First in 2005.

Cohort 1 respondents, already in the program for three years, seem to have leveled off on most dimensions with the exception of Teacher and Coach Implementation and Implementation of the Assurances. The Cohort 1 evaluations of Reading First declined in 2005, but they end up around the same levels as Cohort 2. The Cohort 1 measures suggest that Reading First implementation tends to plateau at a reasonably high level by the end of the second year.

**Conclusions**

This chapter finds:

- Measuring implementation is an essential element in assessing program effectiveness, i.e., the potential of a program to produce achievement gains given a sufficient level of implementation.
- School Reading First implementation measures can be computed from survey data given proper attention to methodological issues. In particular, care needs to be taken to tailor survey forms to specific types of respondents, to equate their responses across forms and years with the appropriate measurement technology, to construct and empirically verify that questions hang together in their intended dimensions, and to reduce incentives or opportunity to "cheat" on the surveys. Given that these conditions are met—and it appears that they have been substantially met in this study so far—a Reading First Implementation Index (RFII) can be computed that is comparable in reliability to standardized achievement tests.
- The RFII can be interpreted as a (theoretical) percentage of times that teachers rate their schools "more than adequate" on relevant survey questions. Using the distribution of school RFII measures, it is possible to state how many schools in the state meet both the "adequate" or better standard and the "more than adequate" standard from the point of view of teachers on selected dimensions.
- 96% of schools in 2005 were rated "adequate" or better by their teachers, up from 92% in 2004.
- Cohorts 1 and 2 have both shown growth in implementation, especially Cohort 2. It appears that most of this growth is achieved by the end of the second year of program funding.
- Cohort 3 is beginning at a lower level of implementation (RFII = 34) than the previous cohorts, which may translate into lower achievement gains. It remains to be seen whether Cohort 3 implementation, like Cohort 2 before it, will surge in its second year.
- Cohort 3's biggest challenge is teacher and coach professional development, followed by LEA implementation of Asurances and teacher use of curricular materials. These findings may reflect

the preponderance of rural districts in Cohort 3, leading to less access to professional development and ongoing support. This is an issue warranting further study.

## References

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

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

## Chapter 4: Achievement Results

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

### The Argument

The objective of this evaluation is to determine whether or not, and to what degree, the Reading First program is effective. What is meant by "effective"? According to the federal guidelines for Reading First, the program is effective if it improves reading outcomes. There are three ways to examine reading outcomes in California's Reading First schools given the limitations of a non-experimental research design.

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

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

The second approach is straightforward once a valid comparison group has been identified. However, by itself it offers only a relative comparison, which is hard to interpret when not enough is known about the

comparison group. A true control group is invaluable in making inferences, but a comparison group with unknown features and less understood characteristics weakens the validity of comparisons, even if it is demographically matched to the Reading First group. A finding that Reading First schools perform better than the Comparison Group does not necessarily mean they are performing well; the Comparison Group schools may be unusually weak on the outcome variable. Likewise, a finding that Reading First schools perform worse than the Comparison Group does not necessarily mean they are performing poorly; the Comparison Group schools may be exceptionally strong. Unfortunately, in our case the Comparison Group is not clearly defined in terms of the degree to which its schools are implementing programs *like* Reading First.

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

Therefore, in order to combine the best aspects of all three approaches to measuring program efficacy, Reading First will be said to show evidence of being effective if:

1. Achievement gains in Reading First schools are positive;
2. Reading First schools show higher achievement gains than comparable non-Reading First schools;
3. High Implementing Reading First schools show higher achievement gains than Low Implementing Reading First schools.

The school achievement metrics are: (1) the California Standards Test or CST, (2) the norm-referenced California Achievement Test Mean Percentile Ranking (CAT/6 MeanPR), (3) the EOY test administered only by funded Reading First districts, and (4) the Reading First Achievement Index (RFAI), a weighted combination of the CST, CAT/6 and EOY data, also available only for Reading First schools. In keeping with our research design, change in achievement is measured over time in year increments, the unit of analysis being not the student but the school, thus the percentage of students within a school meeting some specified performance level or benchmark. The statistics in the charts and tables that follow are generally in this percentage metric and report the 2005 percentage of students meeting benchmark minus

the percentage for the year immediately preceding the program, which yields a gain score.[8] The CAT/6, EOY, and RFAI metrics and gain scores are handled a little differently, as described below. Thus, the reported gain scores can in all cases except the CAT/6 be interpreted as the subtraction of the percentage of students meeting some specified benchmark in one year from the corresponding percentage in a later year.

**Cohorts and Achievement Metrics**

What are the Reading First cohorts and why use cohorts?

It is often found in educational research that intervention program effects vary over time and across cohorts. For example, a change in state achievement tests can cause an anomaly in gain scores that affects the perceived growth curves of different cohorts quite differently. It may also turn out that cohorts have different demographic profiles, and thus respond to educational programs differently. Such cohort effects appear to be an important aspect of Reading First, which is why we have elected to analyze them separately.

Therefore, we report school achievement disaggregated by cohort. As explained in Chapters 2 and 3, there are three Reading First cohorts. Cohort 1 refers to schools that received the first round of funding in the 2002-2003 academic year, though this funding was not received until the middle of the school year. Cohort 2 refers to schools that received the second round of funding prior to the 2003-2004 academic year. Cohort 3 refers to schools that received the third round of funding prior to the 2004-2005 academic year. For each cohort, cross-year gains are reported in each achievement metric for Reading First and non-Reading First schools, with results broken out for High and Low Implementers of Reading First.

Measures of School Progress

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

1. The California Standards Test (CSTs). The CSTs are administered to students in Grades 2 and above toward the end of the school year. For purposes of this study, we use the English Language Arts (ELA) component of the CSTs for Grades 2 and 3. Within ELA, we study the

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

percentage of students per school that fall within each of the three following performance categories, which are a simplification of the five CST performance categories (Advanced, Proficient, Basic, Below Basic, Far Below Basic).

a. "Proficient and Above" means the percentage of students in a school that are in the Proficient and Advanced performance categories;

b. "Basic" means the percentage of students in a school that are in the Basic performance category;

c. "Below Basic and Far Below Basic" means the percentage of students in a school that are in the bottom two performance categories.

The CST gain score for each school is the 2005 percentage of students in a specified category minus the corresponding percentage in the year immediately preceding the first year of Reading First funding. These gain scores are averaged across schools within a cohort to yield the "Average School Gain Score" statistics reported in the tables below.

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

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

3. End-of-Year Test (EOY). As the name suggests, the EOY is a curriculum-based test administered by all Reading First schools to students in Grades K-3 at the end of the academic year. The Kindergarten EOY test consists of seven subtests: Consonants, Lower Case Letters, Phonics, Rhyming, Syllables, Upper Case Letters and Vowels. The EOY tests for Grades 1, 2 and 3 consist of timed oral reading fluency passages in which success is measured in terms of words correct per minute. The EOY is unique and valuable for this study because it is the only test that can be used to measure achievement in Grades K and 1. (The STAR metrics are available only for Grades 2 and above.) One limitation of the EOY is that it is not administered to non-Reading First schools, which prevents comparisons in this metric. Another is the lack of information about the psychometric properties of these particular tests – their reliability and internal consistency – though the validity of similar tests is established (Hasbrouck & Tindal, 2005, Yovanoff & Tindal, 2003). The correlation between EOY scores and CST scores has also been noted (Parker, 2003). The EOY score for each grade within a school consists of the percentage of students that meet the benchmark established for that grade based on national norms recommended by Hasbrouck & Tindal. The gain score for that grade is its 2005 EOY minus its EOY at the end of its *first* year of Reading First funding, which for Cohorts 1 and 2 is 2004. Cohort 3 has no EOY gain score since it was not in Reading First in 2004. Note that, unlike the CSTs and CAT/6, the base year is not the year *preceding* the first year of Reading First funding. Also note that the gain scores are averaged within a given grade across schools in the cohort to produce an average gain score for that grade.

4. Reading First Achievement Index (RFAI). It is a weighted combination of school-level percentages of students meeting various performance levels and benchmarks, drawn from the CSTs, the CAT/6 MeanPR, and the EOY, with the heaviest weights placed on the CSTs. Refer to Appendix G for a detailed explanation of how the RFAI is computed. The RFAI was first computed in 2004. As of this study there are two RFAI indices for Cohorts 1 and 2 (the 2004 and the 2005), and one RFAI for Cohort 3 (the 2005). Because it contains an EOY component, the RFAI can only be computed for Reading First schools. Like the CST, each school RFAI can be interpreted as a percentage of students meeting a set of combined benchmarks and performance levels. The gain score for that school is its 2005 RFAI minus its RFAI at the end of its *first* year of Reading First funding, which for Cohorts 1 and 2 is 2004. Cohort 3 has no RFAI gain score since it was not in Reading First in 2004.

## The Relationship between Reading First Cohorts and Achievement Gains

### *Achievement Gains on the STAR CSTs and CAT/6 MeanPR*

The earliest Reading First cohort (Cohort 1) was funded in the 2002-2003 academic year. As of this report, Cohort 1 has tested for STAR in 2003, 2004, and 2005 since the Reading First program began. However, we use STAR 2002 as the baseline year or pre-program year and compute student gains against that year. Therefore Cohort 1 has a maximum of three years of student gains data, from 2002-03 to 2004-05. Cohort 2 schools first implemented the program in the 2003-2004 academic year and have tested in 2004 and 2005 since program implementation. For Cohort 2, STAR 2003 is the baseline year and therefore this cohort has a maximum of two years of student gains data. Cohort 3, entering Reading First in 2004-05, has one year of gains data, 2004 (baseline year) to 2005. Note that Cohorts 1 and 2 have multiple opportunities to compute gains. Cohort 1 has the possibility of 1-year gain scores (2002 to 2003, 2003 to 2004, or 2004 to 2005) and multiple two-year gain scores (2002 to 2004 or 2003 to 2005). Cohort 2 has two opportunities to compute 1-year gains (2003 to 2004 and 2004 to 2005) and one opportunity for two-year gains (2003-2005). However, because we are most interested in measuring a cohort's total growth, we examine gains for each cohort across all available years following the baseline year.

### *Achievement Gains on the End-of-Year Test*

The EOY was first administered in 2003, the end of the first academic year in Reading First for Cohort 1. The 2003 EOY database proved too incomplete to use as a basis for computing 2-year gains. Therefore, for Cohort 1 we report only 1-year gains, 2004 to 2005. For Cohort 2 we report the same 1-year gain, 2004 to 2005. There is no EOY gain for Cohort 3. An issue in the use of the EOY test is that in 2005 schools with "waiver classrooms" (as defined by AB 1485) had the option of testing students in those classrooms on a Spanish version of the EOY test. Testing in Spanish could only be done in grades K-2, since schools are required by law to test all students in English at end of Grade 3. Therefore, in 2005, EOY data exist for Spanish-instruction and English-instruction students in Grades K-2. Since this is new as of 2005, there is no comparable Spanish-instruction group for 2004. Therefore, EOY gain scores are computed and presented only for non-waiver or English-instruction students for Grades K-3. For both Spanish-instruction and English-instruction students for Grades K-2 we report the percentage of students at benchmark.

*Achievement Gains on the RFAI*

As of this Year 3 Report the published RFAIs are for 2004 and 2005 for Cohorts 1 and 2, and for 2005 for Cohort 3. When reporting RFAI gains, we report only 1-year gains for Cohorts 1 and 2 and no gains for Cohort 3.

Comparison of Reading First to non-Reading First Schools

Chapter 2 provides a detailed description of the comparison schools. Reading First schools were initially compared to two comparison groups of non-Reading First schools—labeled Comparison Group schools and Reading First Eligible schools—and to All Elementary schools. Of these, only those in the Comparison Group are sufficiently similar in terms of percentages of EL and SED students to be comparable to Reading First schools. The pool of unfunded Reading First Eligible schools is demographically too dissimilar to Reading First schools to be used for comparisons. Therefore, we report results only for the Comparison Group and the All Elementary school group. The All Elementary school group does not include Reading First schools and is provided to give an overview of the rest of the state.

What are High Implementation and Low Implementation Reading First schools?

One of the features of this evaluation is that Reading First is studied not only in terms of student achievement but also in terms of observable program implementation at the school level. Chapter 3 describes the teacher, coach, and principal surveys that were administered to all Reading First schools and used to compute a Reading First Implementation Index (RFII) statistic for each school that responded. The RFII is intended to measure the degree to which the teachers, coaches, and principals are implementing the Reading First program in their school. RFII measures have been computed for 2004 and 2005, based on survey administrations in the spring of each year. Therefore, there are two years of RFII scores for Cohorts 1 and 2 Reading First schools and one for Cohort 3.

The RFII was used to divide Reading First schools into two groups labeled High Implementation schools and Low Implementation schools. The method for categorizing schools into these two groups was as follows: For every Reading First school in Cohorts 1 and 2 the 2004 and 2005 RFII scores were combined and averaged. For a Cohort 3 school, the RFII was simply its 2005 RFII score; no combining was used. The mean RFII for all schools (across all three cohorts) in 2005 was computed, equal to 36. If a school's RFII (combined or not) fell at or above 36 points, it was classified as a High Implementation school, otherwise as a Low Implementation school. Thus, this number was used as the cut-off value for differentiating high and low implementers. It is expected that this cut-off will remain the same in future

years and across cohorts to maintain longitudinal comparability and to ensure that "high implementation" has a uniform meaning in all cases.

Achievement and Implementation

The chapter reports achievement gains separately for High Implementation and Low Implementation Reading First schools. The objective, as stated above, is to determine the degree to which a school's implementation of Reading First predicts its achievement gains. If there is no relation, we cannot conclusively report that Reading First is effective in improving student reading achievement. By the same token, if there is such a relation and it cannot be explained otherwise we have support for the hypothesis that Reading First implementation causes achievement to rise, though such a hypothesis cannot be proven in a non-experimental research study such as this.

**Results**

Cohort 1 Results

Tables 4.1.1 through 4.1.3 provide achievement gains for Cohort 1 Reading First Schools. In general, Reading First schools moved more students into the Proficient and Above category than did the Comparison Group schools and All Elementary schools. From 2002 to 2005, Grade 2 Reading First schools moved higher percentages of students (10.8%) *into* the Proficient and Above category. In addition, they moved higher percentages of students *out of* the lower categories, Below Basic & Far Below Basic (-10.7%). In interpreting these tables, bear in mind that a high *negative* percentage is desirable in the bottom categories while a high *positive* percentage is desirable in the top categories.

In Grade 3, Reading First schools did not show positive gains, but also did not decrease in the Proficient and Above category. Comparison Group schools decreased 1% in the Proficient and Above category and schools statewide decreased 2.3% in the Proficient and Above category. Reading First schools were successful in moving students out of Below Basic and Far Below Basic into the Basic category (-3.8%). In interpreting Grade 3 results, it is important to note that students who took the STAR 2005 Grade 3 test were at most exposed to three years of program implementation. They were not exposed to Reading First materials in Kindergarten. For this reason, Cohort 1 Reading First schools may show a greater increase in the Proficient and Above categories for Grade 3 in 2006 than schools in later cohorts since their students will have received Reading First curricula from the earliest grade possible.

When compared on the Grade 3 CAT/6 MeanPR metric, the Reading First schools had strong MeanPR gains in contrast to the Grade 3 CSTs which showed no gains. Bear in mind that the two metrics are not strictly comparable, however, since the CAT/6 gains are in a percentile metric. Reading First schools also

had higher MeanPR scores than the Comparison Group schools and All Elementary schools on Reading, Language, and especially Spelling.

The most revealing comparison is between the High Implementation and Low Implementation Reading First schools and between High Implementation schools and the non-Reading First schools. In Cohort 1, 141 schools were categorized as High Implementation schools and 129 schools were categorized as Low Implementation. The gains made by the High Implementation schools are the highest of all the groups presented in this table. In Grade 2, High Implementation schools moved 12.1% of their students into the Proficient and Above category and moved 12.4% of their students out of the bottom categories. In Grade 3, they moved 0.5% students into Proficient and Above (other groups moved zero or less) and moved 4.3% students out of the bottom categories of the CST.

All Elementary schools had less dramatic gains than All Reading First schools, High Implementation schools *and* Low Implementation schools. When compared for statistical significance, the gains made by All Elementary schools are significantly lower than those for All Reading First schools and High Implementation schools on all metrics except for Grade 3 Basic. The Low Implementation Reading First schools have achievement gains on par with the Comparison Group schools.

**Table 4.1.1: 3-Year Gain, 2002 to 2005 for Cohort 1 Reading First and non-Reading First Schools**

| | Average School Gain Score (see *Measures of School Progress* above) | | | | |
|---|---|---|---|---|---|
| | Reading First Schools | | | Non-Reading First Schools | |
| | All Reading First Schools | High Implementation Schools [3] | Low Implementation Schools | Comparison Group Schools | All Elementary Schools [4] |
| **CST** | | | | | |
| **Grade 2** [1] | (N=276) [2] | (N=141) | (N=129) | (N=354) | (N=4143) |
| Proficient & Above | 10.8 | 12.1 | 9.5 [b] | 9.8 [b] | 9.4 [a,b] |
| Basic | -0.2 | 0.3 | -0.6 | 0.4 | -4.6 [a,b] |
| Below Basic & Far Below Basic | -10.7 | -12.4 | -9.0 [b] | -10.2 | -4.8 [a,b] |
| **Grade 3** | (N=275) | (N=140) | (N=129) | (N=356) | (N=4154) |
| Proficient & Above | 0.0 | 0.5 | -0.5 | -1.0 | -2.8 [a,b] |
| Basic | 3.8 | 3.8 | 3.7 | 3.3 | 2.8 |
| Below Basic & Far Below Basic | -3.8 | -4.3 | -3.2 | -2.2 | -0.0 [a,b] |
| **CAT/6 MeanPR** | | | | | |
| **Grade 3** | (N=275) | (N=140) | (N=129) | (N=357) | (N=4152) |
| Reading | 2.9 | 3.2 | 2.4 | 2.8 | 0.2 [a,b] |
| Language | 3.8 | 4.0 | 3.6 | 2.9 | 2.1 [a,b] |
| Spelling | 9.6 | 10.6 | 8.3 | 6.0 [a,b] | 3.6 [a,b] |

[a] Statistically significant difference at $p < 0.05$ as compared to All Reading First schools.
[b] Statistically significant difference at $p < 0.05$ as compared to High Implementation Reading First schools.

[1] The N of schools may be different between Grades 2 and 3 within the same cohort because not all schools have both grades. However, the N of schools within the same cohort for the Grade 3 CSTs and CAT/6 MeanPR match.
[2] N is always in reference to Schools and not students.
[3] The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools in 2004 or 2005. Schools may be missing an RFII because they did not submit surveys.
[4] All Elementary Schools excludes 821 (Cohort1, Cohort 2 and Cohort 3) Reading First schools.

Table 4.1.2 presents the 1-year EOY gain for Cohort 1 from 2004 to 2005. As mentioned, even though two years of EOY gain data are available for Cohort 1, we ignore the 2003 results as being less reliable due to incomplete data sets. To explain further, the 2002-2003 academic year was the first year of Reading First implementation and the first that the EOY test was administered. As a new test, its implementation was not mandatory. Some schools did not administer it, others did only partially. There were also problems with data submission and data collection, and a number of schools scored lower on the EOY than they would have under ordinary circumstances. School participation in the EOY assessment and data collection efforts improved in 2004, providing more reliable EOY numbers. For purposes of this report, only the 1-year EOY gain from 2004 to 2005 is reported for Cohort 1 schools. (For informational purposes, Appendix F provides 2-year EOY gains for Cohort 1 schools.)

Table 4.1.2 shows that Reading First schools produced gains ranging from 3.5% to 6.2% across grades on the EOY test between 2004 and 2005. The pattern of higher growth for High Implementation Reading First schools that was seen for the CSTs and CAT/6 MeanPR is not reproduced here; only Kindergarten and Grade 2 show an implementation effect. In terms of the percent of students at benchmark on the EOY, English-instruction students show higher success rates than waiver (Spanish-instruction) classrooms, except for Grade 1 in High Implementation schools. The Grade 2 Spanish-instruction statistics seem anomalous.

It is important to note the limitations of the EOY tests in interpreting these results. The EOY test does not have established reliability and validity as occurs with standardized tests. The administration and data collection specifications for EOY tests are evolving and adjusting to changes in the Reading First program (such as the addition of waiver classrooms). Another important limitation is that the base line of the gains is not the score from the year preceding Reading First funding but the score at the end of the first year. Plus, the span is only one year.

If the EOY tests are to be used for evaluative purposes in the future, it would be useful to subject them to validation procedures similar to those used for standardized tests. It is expected that in future years the EOY results will be more informative.

**Table 4.1.2: End-of-Year Fluency Test: Gain (2004-2005) and Percent Meeting Benchmark 2005, Cohort 1 Reading First Schools**

| | All Reading First Schools | | | | High Implementation Schools [1] | | | | Low Implementation Schools [1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2004-2005 Gain | | 2005 % at Benchmark | | 2004-2005 Gain | | 2005 % at Benchmark | | 2004-2005 Gain | | 2005 % at Benchmark | |
| | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean |
| Kindergarten | | | | | | | | | | | | |
| English | 264 | 3.9 | 275 | 80.2 | 133 | 4.2 | 140 | 81.4 | 125 | 3.4 | 129 | 78.7 |
| Spanish | - | - | 69 | 71.7 | - | - | 22 | 73.0 | - | - | 45 | 70.7 |
| Grade 1 | | | | | | | | | | | | |
| English | 274 | 3.5 | 276 | 43.3 | 140 | 2.9 | 141 | 43.2 | 128 | 3.9 | 129 | 43.1 |
| Spanish | - | - | 71 | 42.2 | - | - | 23 | 46.0 | - | - | 46 | 42.0 |
| Grade 2 | | | | | | | | | | | | |
| English | 273 | 6.2 | 276 | 44.7 | 139 | 7.7 | 141 | 46.0 | 128 | 4.9 | 129 | 43.2 |
| Spanish | - | - | 71 | 29.4 | - | - | 23 | 23.6 | - | - | 46 | 32.8 |
| Grade 3 | | | | | | | | | | | | |
| English | 272 | 4.1 | 275 | 44.3 | 139 | 3.7 | 140 | 43.7 | 127 | 4.3 | 129 | 44.8 |

[1]The N of schools under High Implementation and Low Implementation does not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools in 2004 or 2005. Schools may be missing an RFII because they did not submit surveys.

Table 4.1.3 compares the gains in the RFAI for High and Low Implementation schools. The pattern, showing higher gains for High Implementation schools, is consistent with the CST and CAT/6 Mean PR gains in Table 4.1.1. This is not surprising in light of the fact that the RFAI is a weighted combination of the CSTs and the CAT/6 MeanPR. However, it includes an EOY component which, in light of Table 4.1.2, is probably narrowing the difference between the High and Low Implementers. This highlights the need to further examine the EOY test, as it impacts the RFAI.

**Table 4.1.3: One Year RFAI Gain 2004 to 2005 Cohort 1 Reading First Schools**

| | All Reading First Schools | | High Implementation Reading First Schools [1] | | Low Implementation Reading First Schools | |
|---|---|---|---|---|---|---|
| | N of Schools | Mean % Gain | N of Schools | Mean % Gain | N of Schools | Mean % Gain |
| RFAI Gain | 277 | 4.1 | 141 | 4.3 | 130 | 4.0 |

[1]The N of schools under High Implementation and Low Implementation does not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools in 2004 or 2005. Schools may be missing an RFII because they did not complete the teacher survey.

One disadvantage of studying only total 3-year gains for the CSTs and CAT/6 MeanPR is that they say nothing about the shape of the trend lines across years. Neither do they reveal the absolute percentage of students in a given category at a given time. To provide a more complete picture of the historical performance of these schools, we present trend line graphs. Figure 4.1.1 displays the percentage of students in the CST Proficient and Above category in Grade 2 from 2002 to 2005. Trend lines are given for All Reading First schools, High Implementation schools, Low Implementation Reading First schools, and Comparison Group Schools. Figure 4.1.2 displays the same trend lines for Grade 3. Figures 4.1.3, 4.1.4, and 4.1.5 track the average school MeanPR for Reading, Language Arts, and Spelling for Grade 3 from 2002 through 2005 for the same school groups.

Figure 4.1.1, the Grade 2 trend lines, prompts the following observations:

- All groups follow an N-shaped curve – an initial jump, partial retreat, a larger jump. Though not fully understood, this pattern is probably an artifact of the CST, masking a simple upward trend. The fact that all groups follow it bears this out.
- The trend is positive for all groups in Grade 2.
- The All Reading First schools trend line appears to parallel the trend line of the Comparison Group schools, which is above it, but the gap narrows slightly. This indicates two things:

- The Comparison Group schools start in 2002 with more Proficient and Above students than the Reading First schools and have maintained their lead;
- The Reading First schools are slowly gaining on the Comparison Group, which is why their total gains are higher as reported in Table 4.1.1, though not significantly so.

- The trend line for the High Implementation Reading First schools starts in 2002 with almost the same Proficient and Above percentages as the other Reading First groups, yet it rises more rapidly than all the other trend lines, surpassing the Comparison Group in 2004 and 2005. Factors that depressed scores in 2004 also affected the High Implementation group, but much less so.
- Whatever the cause of the strong gains in the High Implementation school trend line, it is unrelated to the starting point since the High Implementation schools started in 2002 at around the same percentage of Proficient and Above students as the Low Implementation schools. Yet it rises much more steeply.
- The High and Low Implementation school trend lines are nearly parallel in the first year. It is in the second and subsequent years that the High Implementation differential effect becomes evident. This tells us that schools need at least one year for implementation effects to become evident.
- The Low Implementation trend line, while uniformly lower than the others, rises at about the same rate as the Control Group schools. This tells us that schools that are below average in the implementation of the Reading First program are not worse off by using Reading First. They merely improve at the same rate as non-Reading First schools.
- For Grades 2 and 3 and across cohorts (though it is especially evident in Grade 2), High Implementation schools post higher gains—trend upward more steeply—than Low Implementation schools. Absent another explanation, we conclude that Reading First is effective in improving achievement to the degree it is implemented.

The Grade 3 CST trends for Reading First schools in Figure 4.1.2 mimic other statewide CST trends (see http://star.cde.ca.gov/ for statewide STAR reports) and reflect no significant gain in percent Proficient and Above in Grade 3. None of the groups in Figure 4.1.2 show gains. As in the Grade 2 trends, we see that the Comparison Group schools began at a higher proficiency level in 2002 than the other groups, but in Grade 3, the Comparison Group schools maintained their lead in 2005, whereas the Grade 2 High Implementation scools surpass the Comparison Group schools. The differences narrowed steadily over

time between the Comparison Group schools trend line and both the All Reading First and the High Implementation Reading First trend lines. The Reading First schools remained at the same level while the Comparison Group declined. High Implementation schools were at a higher level than Low Implementation schools and the gap increased slightly over time. The flatness of these curves is probably an artifact of the Grade 3 CST test since it is not reproduced in the Grade 3 CAT/6.

Figures 4.1.3 and 4.1.4 show trend lines for the CAT/6 MeanPR Reading, Language Arts and Spelling tests from 2002 to 2005. While not as dramatic as the Grade 2 CST trend lines, they tell the same story. Instead of a decline, the trend lines for all groups are positive. High Implementation leads to higher gains than Low Implementation. In Reading and Language Arts, the Comparison Group starts with a substantial lead and the Reading First schools narrow the gap over time. In Spelling, the Reading First schools not only start out ahead but increase their lead dramatically. Spelling is a notable strength of the approved Reading First curricula.

**Figure 4.1.1: Cohort 1 CSTs, 2002 to 2005, Percent Proficient and Above Trend Lines, Grade 2**

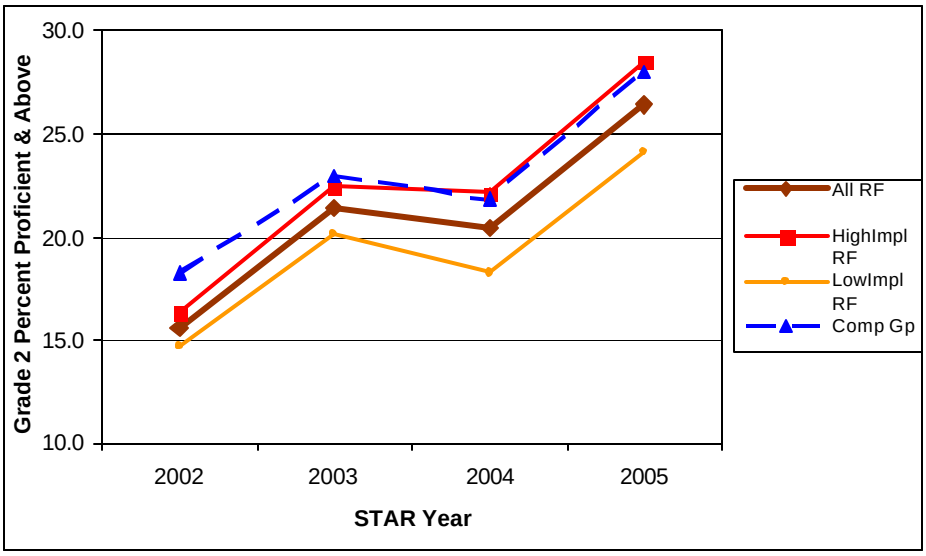



**Figure 4.1.2: Cohort 1 CSTs, 2002 to 2005, Percent Proficient and Above Trend Lines, Grade 3**

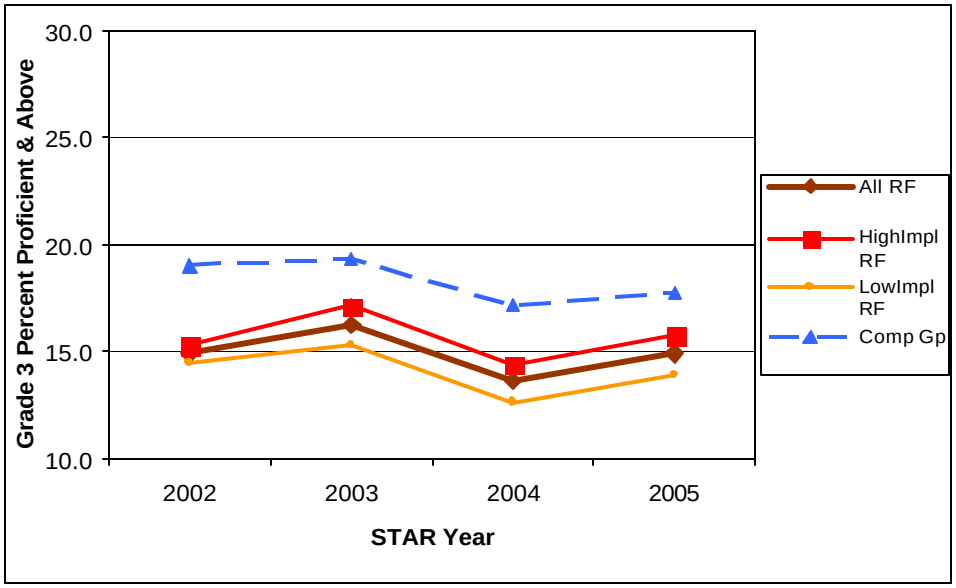


**Figure 4.1.3: Cohort 1 Mean Percentile Rank, CAT/6 Reading, 2002 to 2005, Grade 3**

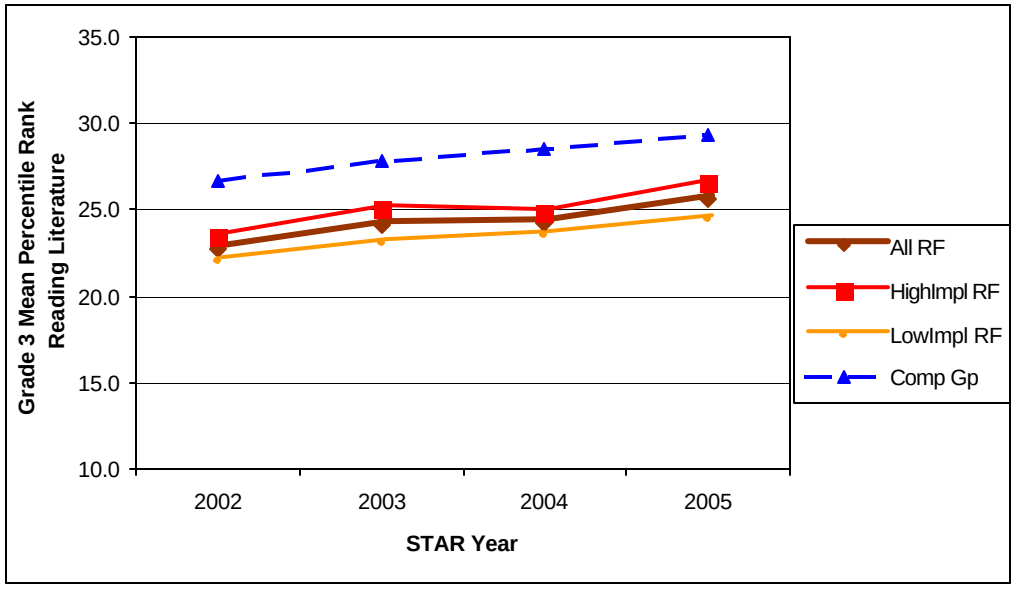



**Figure 4.1.4: Cohort 1 Mean Percentile Rank, CAT/6 Language Arts, 2002 to 2005, Grade 3**

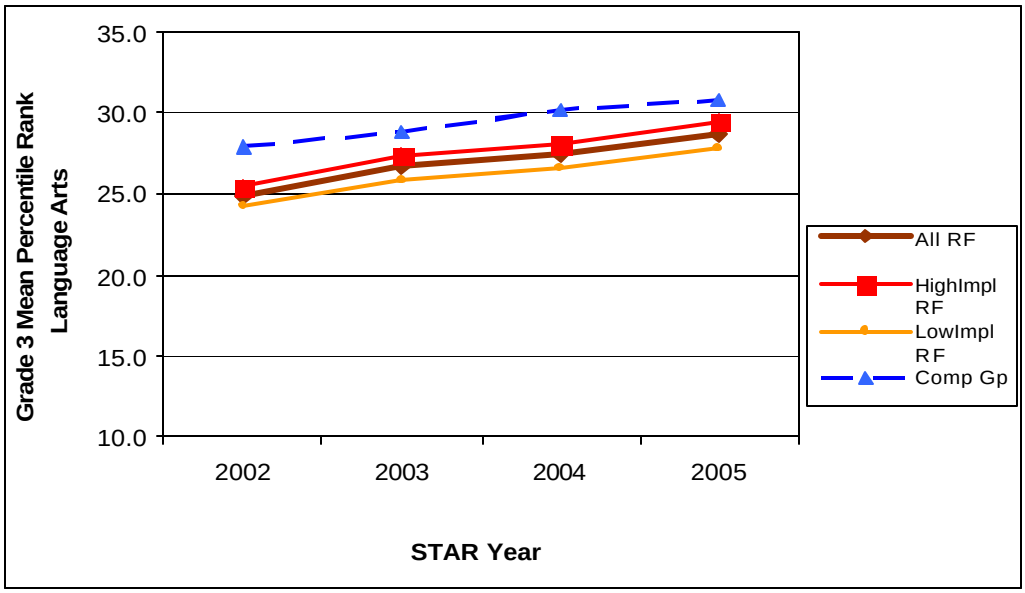
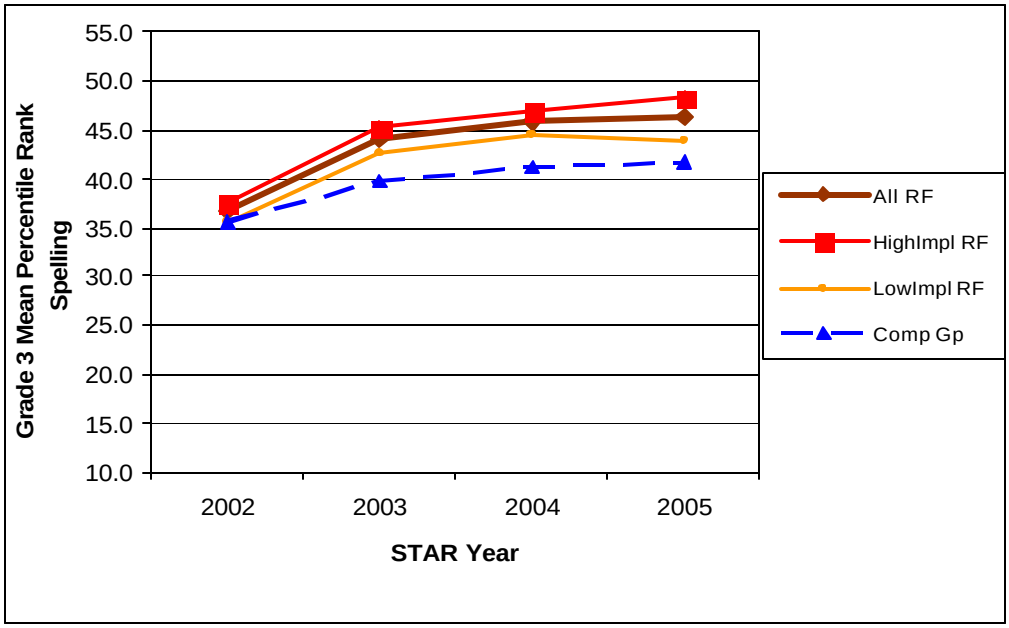


**Figure 4.1.5: Cohort 1 Mean Percentile Rank, CAT/6 Spelling, 2002 to 2005, Grade 3**



Cohort 2 Results

Table 4.2.1 shows the 2-year gains for Cohort 2 Reading First schools. STAR 2005 scores are compared to STAR 2003 scores to measure growth. On average, Reading First schools moved 5.7% of students into the Grade 2 Proficient and Above category, compared to 5.2% for the Comparison Group schools and 4.6% for All Elementary schools. As with Cohort 1, these gains are not observed in Grade 3, where Reading First schools and Comparison Group schools show declines in the percentage of students Proficient and Above between 2003 and 2005. However, High Implementation Reading First schools in Cohort 3 show a slight (0.3%) gain in the Grade 3 Proficient or Above category. The gains for the High Implementation schools are significantly higher than those made by Comparison Group schools and All Elementary schools on multiple metrics, as signified by the "b" superscripts in the Comparison Group and All Elementary schools columns. High Implementation Reading First schools were significantly better at moving students out of the bottom categories of the CST. They also show significantly larger gains for the CAT/6 MeanPR. Low Implementation schools post Proficient and Above gains on the CST similar to the Comparison Group schools.

The EOY results for Cohort 2 are similar to results for Cohort 1. Table 4.2.2 provides the 1-year (2004 to 2005) gain on the EOY test for English-instruction students in Cohort 2 schools. High Implementation schools outperform than Low Implementation schools in Grade 2 but not in Grades K, 1, or 3. When compared on percent at benchmark, English-instruction students do better than Spanish-instruction students except for Grade 1 Low Implementation schools. High Implementation English-instruction students show higher percent at benchmark than Low Implementation students. This pattern is reversed for Spanish-instruction students where students in Low Implementation schools do better than students in High Implementation schools. Again, implementation has no obvious relation to achievement in the EOY data.

Table 4.2.3 presents the one-year gain on the RFAI for Cohort 2 Reading First schools. The average Cohort 2 Reading First school gained 4.3 points on the RFAI. High Implementation schools show higher gains on the RFAI (4.6 points) than Low Implementation schools (4.1 points). This cohort has only one year of RFAI gains to examine. The RFAI is an important indicator of academic performance and, with only one set of gains to examine for this cohort, it will be monitored closely in future reports.

**Table 4.2.1: Cohort 2 Schools 2-Year Gains, 2003 to 2005**

| | Average School Gain Score (see *Measures of School Progress* above) | | | | |
|---|---|---|---|---|---|
| | Reading First Schools | | | Non-Reading First Schools | |
| | All Reading First Schools | High Implementation Schools[3] | Low Implementation Schools | Comparison Group Schools | All Elementary Schools[4] |
| **CST** | | | | | |
| **Grade 2[1]** | (N=378)[2] | (N=177) | (N=163) | (N=361) | (N=4250) |
| Proficient & Above | 5.7 | 6.8 | 5.1 | 5.2 | 4.5[b] |
| Basic | -2.6 | -2.7 | -2.8 | -4.5[a,b] | -4.6[a,b] |
| Below Basic & Far Below Basic | -3.1 | -4.2 | -2.2 | -0.6[a,b] | 0.1[a,b] |
| **Grade 3** | (N=376) | (N=177) | (N=163) | (N=357) | (N=4269) |
| Proficient & Above | -0.5 | 0.3 | -1.5 | -1.6[b] | -2.3[a,b] |
| Basic | 1.7 | 1.7 | 1.7 | 1.2 | 1.1 |
| Below Basic & Far Below Basic | -1.3 | -2.1 | -0.2 | 0.4[b] | 1.1[a,b] |
| **CAT/6 MeanPR** | | | | | |
| **Grade 3** | (N=376) | (N=177) | (N=163) | (N=358) | (N=4262) |
| Reading | 2.6 | 3.1 | 1.8 | 1.3[a,b] | 0.4[a,b] |
| Language | 3.6 | 4.1 | 2.8 | 1.7[a,b] | 1.1[a,b] |
| Spelling | 5.1 | 6.0 | 4.1 | 1.5[a,b] | 0.5[a,b] |

[a]Statistically significant difference at $p < 0.05$ as compared to All Reading First schools
[b]Statistically significant difference at $p < 0.05$ as compared to High Implementation Reading First schools.

[1]The N of schools may be different between Grades 2 and 3 within the same cohort because not all schools have both grades. However, the N of schools within the same cohort for the Grade 3 CSTs and CAT/6 MeanPR match.
[2] N is always in reference to schools and not students.
[3]The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools in 2004 or 2005. Schools may be missing an RFII because they did not submit surveys.
[4]All Elementary Schools excludes 821 (Cohort1, Cohort 2 and Cohort 3) Reading First schools.

**Table 4.2.2: End-of-Year Fluency Test: Gain (2004-2005) and Percent Meeting Benchmark for 2005, Cohort 2 Reading First Schools**

| | All Reading First Schools | | | | High Implementation Schools [1] | | | | Low Implementation Schools [1] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2004-2004 Gain | | 2005 % at Benchmark | | 2004-2004 Gain | | 2005 % at Benchmark | | 2004-2004 Gain | | 2005 % at Benchmark | |
| | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean |
| Kindergarten | | | | | | | | | | | | |
| English | 364 | 3.8 | 378 | 76.4 | 171 | 3.6 | 175 | 78.7 | 155 | 4.0 | 162 | 73.6 |
| Spanish | - | - | 123 | 68.3 | - | - | 40 | 60.5 | - | - | 57 | 72.2 |
| Grade 1 | | | | | | | | | | | | |
| English | 371 | 4.8 | 382 | 37.0 | 178 | 4.2 | 178 | 38.7 | 155 | 4.6 | 163 | 35.9 |
| Spanish | - | - | 124 | 36.0 | - | - | 41 | 34.2 | - | - | 57 | 37.5 |
| Grade 2 | | | | | | | | | | | | |
| English | 368 | 5.7 | 381 | 38.8 | 174 | 7.1 | 178 | 41.6 | 157 | 5.2 | 162 | 37.2 |
| Spanish | - | - | 124 | 25.8 | - | - | 41 | 23.3 | - | - | 57 | 24.7 |
| Grade 3 | | | | | | | | | | | | |
| English | 371 | 4.2 | 379 | 36.8 | 175 | 4.2 | 177 | 38.2 | 160 | 4.5 | 162 | 35.7 |

[1]The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools in 2004 or 2005. Schools may be missing an RFII because they did not submit surveys.

**Table 4.2.3: 1-Year RFAI Gain 2004 to 2005, Cohort 2 Reading First Schools**

| | All Reading First Schools | | High Implementation Reading First Schools [1] | | Low Implementation Reading First Schools | |
|---|---|---|---|---|---|---|
| | N of Schools | Mean % Gain | N of Schools | Mean % Gain | N of Schools | Mean % Gain |
| RFAI Gain | 384 | 4.3 | 180 | 4.6 | 164 | 4.1 |

[1]The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools in 2004 or 2005. Schools may be missing an RFII because they did not submit surveys.

Figures 4.2.1 and 4.2.2 show the CST trend lines for Proficient and Above for Cohort 2 Grades 2 and 3 Reading First schools. Since STAR 2003 was the pre-Reading First year for Cohort 2 schools, the trend lines are shown from 2003 to 2005. Both in Grades 2 and 3, it is evident that the All Reading First schools, the Comparison Group schools and the Low Implementation Reading First schools follow similar trends of gains and losses. The percent Proficient and Above fell in 2004 and increased in 2005, sharply for Grade 2, moderately for Grade 3.

What is striking in these trend lines is the pattern of growth for the High Implementation schools. Although the 1-year trend for Grade 2 parallels the other trend lines, in the second year the slope sharpens

dramatically. In Grade 3, while the other trend lines drop, the High Implementation trend line moves upward. Like High Implementation schools in Cohort 1, Cohort 2 High Implementation schools break away from the average and show startling gains in percent Proficient and Above in the second year of implementation.

Figures 4.2.3, 4.2.4, and 4.2.5 display trend lines for CAT/6 MeanPR in Reading, Language Arts and Spelling. Like the CST trend lines, the CAT/6 MeanPR lines parallel each other except for the High Implementation schools trend line. The High Implementation schools set themselves apart with a brisk upward trend.

**Figure 4.2.1: Cohort 2 CSTs, 2002 to 2005, Percent Proficient and Above Trend Lines, Grade 2**

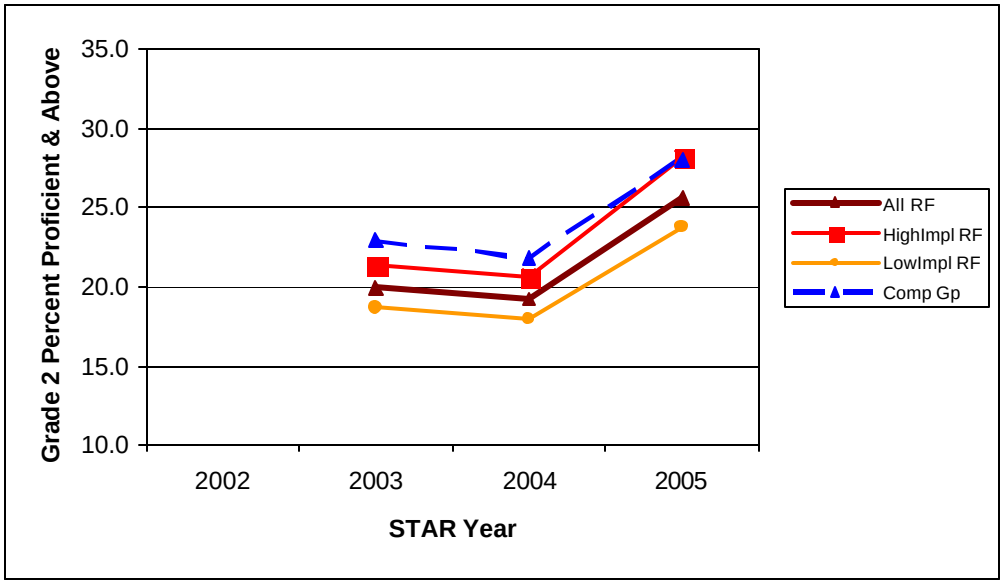



**Figure 4.2.2: Cohort 2 CSTs, 2003 to 2005, Percent Proficient and Above Trend Lines, Grade 3**

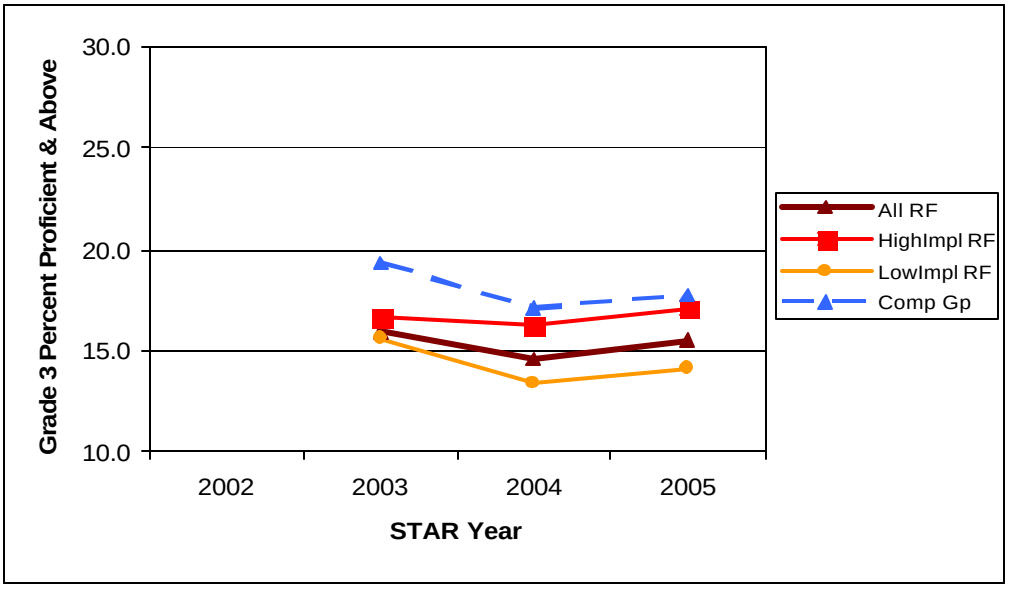


**Figure 4.2.3: Cohort 2 Mean Percentile Rank, CAT/6 Reading, 2003 to 2005, Grade 3**

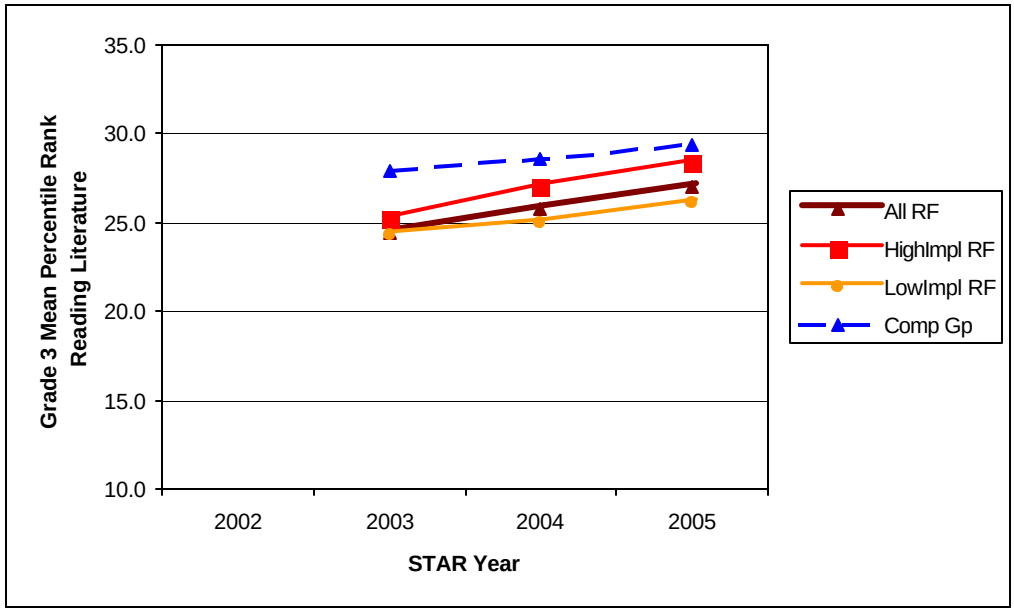



**Figure 4.2.4: Cohort 2 Mean Percentile Rank, CAT/6 Language Arts, 2003 to 2005, Grade 3**

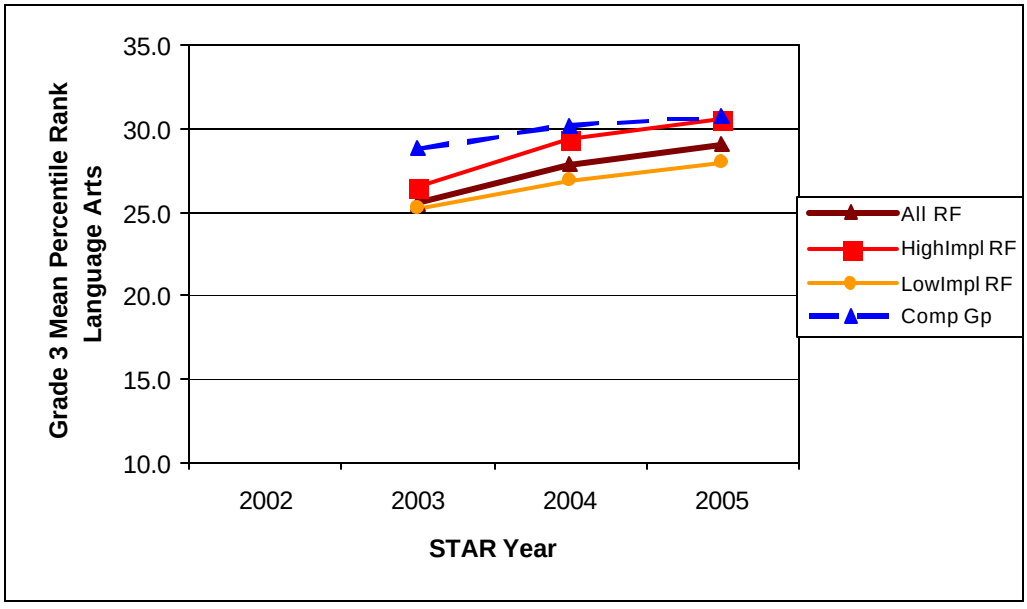


**Figure 4.2.5: Cohort 2 Mean Percentile Rank, CAT/6 Spelling, 2003 to 2005, Grade 3**


Cohort 3 Results

Table 4.3.1 presents the 1-year gains for Cohort 3. Cohort 3 is the most recent addition to the Reading First program in California. We saw in Cohorts 1 and 2 that Reading First schools tend to parallel non-Reading First schools in the first year of implementation. Table 4.3.1 and the trend lines in Figures 4.3.1 and 4.3.2 replicate that pattern. Between 2004 and 2005 Reading First schools show marginally better growth than other groups of comparison schools. Since there is only one year of achievement gains for Cohort 3 schools, it is too early to draw reliable conclusions about the impact of the Reading First program for this cohort.

It is interesting to note from the trend line graphs that the Cohort 3 Reading First schools, when they started in the program, were 3 to 5 percentage points lower than the Comparison Group schools in 2004. That is, the starting point for Cohort 3 schools is quite a bit lower for the Reading First schools than it was for Cohorts 1 and 2. This suggests that the Comparison Group schools that were drawn in 2004 to match the Cohorts 1 and 2 Reading First schools may not be comparable to the Cohort 3 schools.

Another observation is that the High Implementation schools also started at a higher proficiency level than the Low Implementation schools in 2004. In the previous two cohorts, the difference in starting

point between High Implementation, Low Implementation, and All Reading First schools is not as wide as it is for the Cohort 3 schools.

Note that Cohort 3 schools do not have gains data for the EOY, the first Cohort 3 EOY test having been administered at the end of the 2004-2005 academic year. The percent at benchmark numbers are presented in Table 4.3.2. Aside from Grade 1, English-instruction students show higher proficiencies than Spanish-instruction students. Again, no consistent implementation effect is apparent.

**Table 4.3.1: Cohort 3 Schools 1-Year Gains, 2004 to 2005**

| | Average School Gain Score (see *Measures of School Progress* above) | | | | |
|---|---|---|---|---|---|
| | Reading First Schools | | | Non-Reading First Schools | |
| | All Reading First Schools | High Implementation Schools[3] | Low Implementation Schools | Comparison Group Schools | All Elementary Schools[4] |
| **CST** | | | | | |
| **Grade 2[1]** | (N=144)[2] | (N=37) | (N=102) | (N=371) | N=(4331) |
| Proficient & Above | 6.0 | 7.0 | 5.8 | 6.0 | 6.0 |
| Basic | 0.8 | -0.3 | 1.2 | -1.4[a] | -3.3[a,b] |
| Below Basic & Far Below Basic | -6.9 | -6.6 | -7.0 | -4.6[a] | -2.7[a,b] |
| **Grade 3** | (N=146) | (N=37) | (N=104) | (N=370) | (N=4355) |
| Proficient & Above | 1.2 | 1.5 | 0.9 | 0.5 | 1.3 |
| Basic | 1.2 | 0.8 | 1.6 | 0.2 | -0.4[a] |
| Below Basic & Far Below Basic | -2.3 | -2.1 | -2.3 | -0.7 | -0.9 |
| **CAT/6 MeanPR** | | | | | |
| **Grade 3** | (N=146) | (N=37) | (N=104) | (N=370) | (N=4345) |
| Reading | 1.2 | 1.0 | 1.3 | 0.7 | 0.7 |
| Language | 1.2 | 1.4 | 1.2 | 0.5 | 0.5 |
| Spelling | 1.6 | 2.5 | 1.3 | 0.3 | 0.4 |

[a]Statistically significant difference at $p < 0.05$ as compared to All Reading First schools.
[b]Statistically significant difference at $p < 0.05$ as compared to High Implementation Reading First schools.

[1]The N of schools may be different between Grades 2 and 3 within the same cohort because not all schools have both grades. However, the N of schools within the same cohort for the Grade 3 CSTs and CAT/6 MeanPR match.
[2] N is always in reference to schools and not students.
[3]The N of schools under High Implementation and Low Implementation do not always add up to the N of total Reading First schools in that row because of missing RFII statistics for schools in 2004 or 2005. Schools may be missing an RFII because they did not submit surveys.
[4]All Elementary Schools excludes 821 (Cohort1, Cohort 2 and Cohort 3) Reading First schools.

As was pointed out in Chapter 3, we see that the number of High Implementation schools in Cohort 3 (N=37) is much less than the number of Low Implementation schools (N=102). If implementation remains at these low levels, we expect the Cohort 3 schools to show slower rates of growth than those in the first two cohorts. The most pressing need in these schools, as mentioned, appears to be better access

to professional development. However, if Cohort 3 follows the pattern set by Cohort 2, we may well see a jump both in implementation and in achievement by the end of the second academic year, i.e., 2005-2006.

**Table 4.3.2: End-of-Year Fluency Test: Percent Proficiency for 2005, Cohort 3 Reading First Schools**

| | All Reading First Schools | | High Implementation[1] Reading First Schools | | Low Implementation[1] Reading First Schools | |
|---|---|---|---|---|---|---|
| | N of Schools | % Proficient | N of Schools | % Proficient | N of Schools | % Proficient |
| Kindergarten | | | | | | |
| English | 149 | 71.2 | 44 | 66.2 | 100 | 73.6 |
| Spanish | 60 | 66.4 | 19 | 67.9 | 39 | 64.8 |
| Grade 1 | | | | | | |
| English | 145 | 29.6 | 41 | 28.3 | 99 | 29.4 |
| Spanish | 57 | 31.6 | 17 | 31.7 | 39 | 31.2 |
| Grade 2 | | | | | | |
| English | 145 | 33.2 | 40 | 31.3 | 101 | 33.7 |
| Spanish | 57 | 19.1 | 17 | 19.8 | 39 | 18.7 |
| Grade 3 | | | | | | |
| English | 145 | 29.7 | 40 | 31.6 | 101 | 29.2 |

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

**Figure 4.3.1: Cohort 3 CSTs, 2004 to 2005, Percent Proficient and Above Trend Lines, Grade 2**

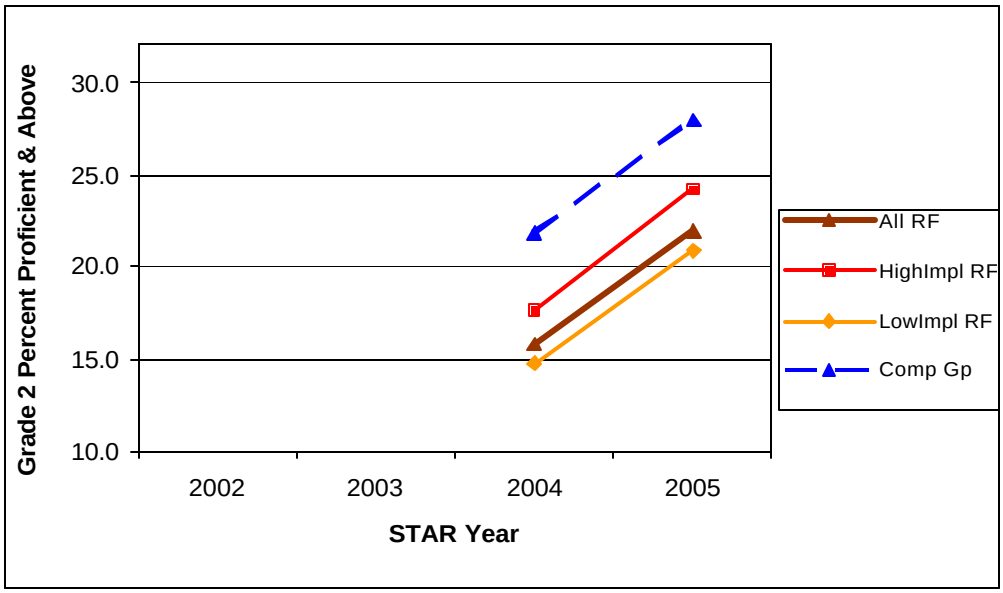



**Figure 4.3.2: Cohort 3 CSTs, 2004 to 2005, Percent Proficient and Above Trend Lines, Grade 3**

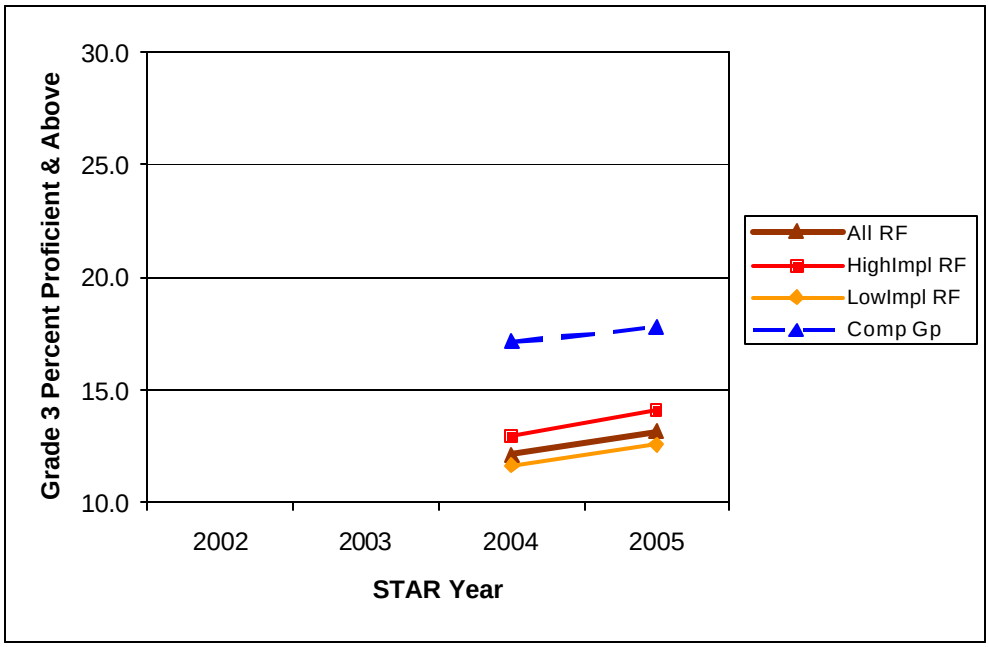


**Figure 4.3.3: Cohort 3 Mean Percentile Rank, CAT/6 Reading 2004 to 2005, Grade 3**

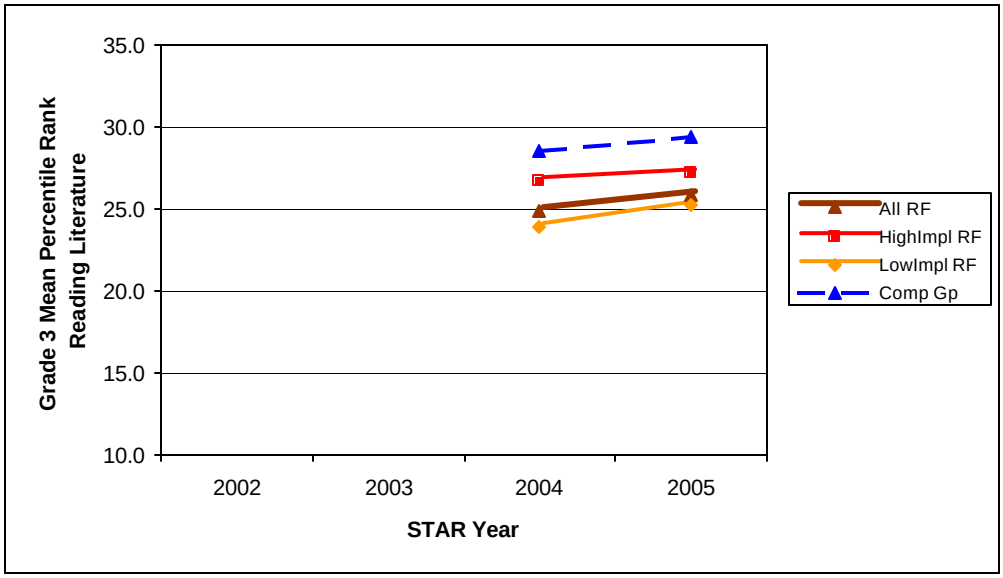



**Figure 4.3.4: Cohort 3 Mean Percentile Rank, CAT/6 Language Arts 2004 to 2005, Grade 3**

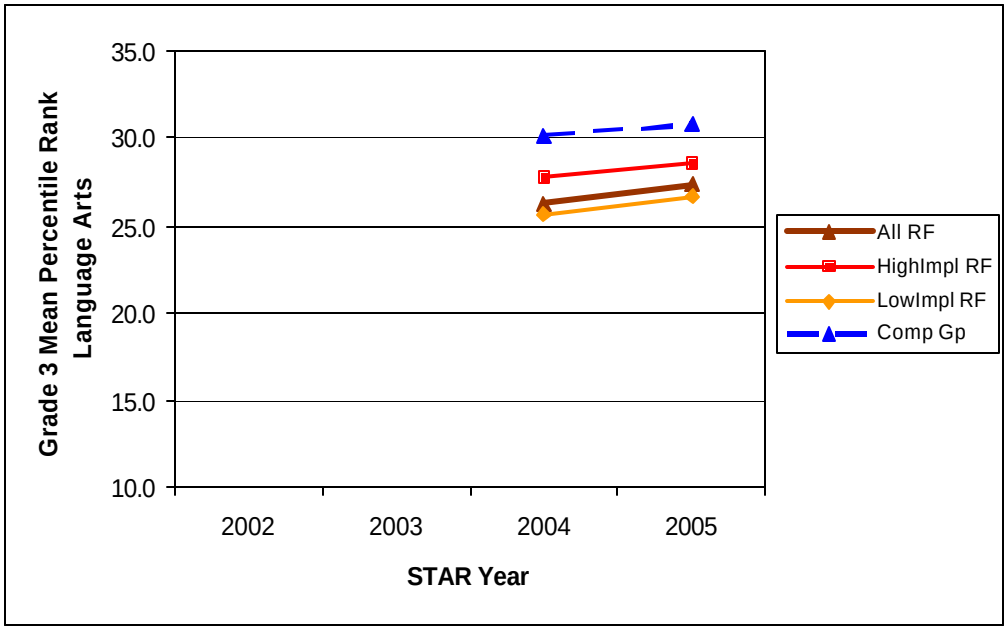
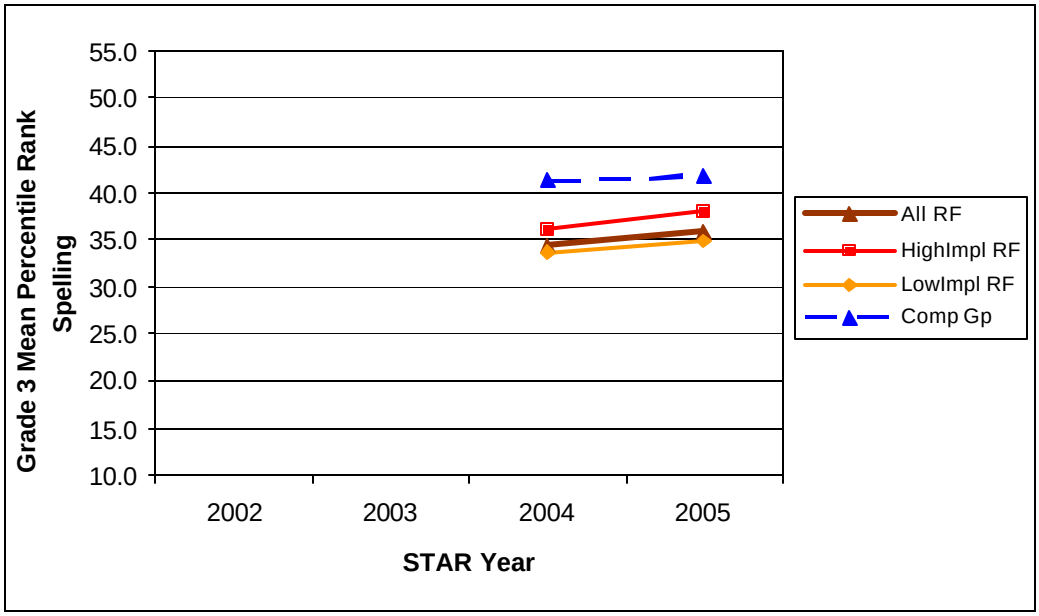


**Figure 4.3.5: Cohort 3 Mean Percentile Rank, CAT/6 Spelling 2004 to 2005, Grade 3**



How does Reading First implementation relate to achievement?

Figure 4.4.1 displays the Grade 2 CST achievement gains for Cohort 1 for five levels of Reading First implementation. In accord with other graphs in this chapter, it provides visual evidence that Reading First implementation has a substantial and positive effect on Grade 2 achievement gains. Simply put, it demonstrates that *implementation matters*, and that for most schools the program is effective when implemented, at least for Grade 2.

In Figure 4.4.1, we break the level of implementation into five groups according to level of implementation, or quintiles, to discern differences between the higher and lower levels of implementation. Each trend line represents a quintile. The quintiles are based on the average RFII score of 2004 and 2005 for the Cohort 1 schools. The progress of each implementation quintile is tracked over four years. Although all five quintiles begin at nearly the same level of achievement, after three years they "fan out" across the achievement spectrum. The highest implementation quintile ends with the highest percent of students Proficient and Above. The lowest implementation quintile ends with the lowest percent Proficient and Above. The intermediary quintiles spread out in between in an orderly fashion. Table 4.4.2 shows the same analysis by quintile for Grade 3 as for Grade 2.

The "fan" pattern for Grade 2 appears to demonstrate, strikingly, that implementation of Reading First improves achievement scores. But is the effect statistically significant, i.e., could it have occurred by chance? To answer this question, we constructed regression models for Grades 2 and 3 to predict achievement on the CSTs. Achievement was defined as the "school mean performance level" on the CSTs for the 2004-05 school year. As predictor variables, we used the school's starting point (its mean performance level on the CSTs for the year prior to its first year of Reading First implementation), its number of years in the program as of 2005, its mean RFII (the average of all its RFII statistics as of 2005), its percent of English Learners (ELs), and its percent of Socio-economically Disadvantaged (SED) students. To capture the total implementation effect, we multiplied a school's mean RFII by its number of years in the program to create a new predictor variable whose effect size was computed in a separate regression analysis. The same procedure was followed for Grade 3. The effect of each predictor variable on achievement, as well as its statistical significance, is provided in Tables 4.4.1 and 4.4.2. Further details are provided in Appendix F.

These tables demonstrate, as discussed below, that the impact of Reading First implementation on achievement is statistically significant for Grade 2, and to a lesser extent for Grade 3.

**Figure 4.4.1: Cohort 1 CSTs, 2002 to 2005, Percent Proficient Trend Lines, Grade 2, by Implementation Quintiles**

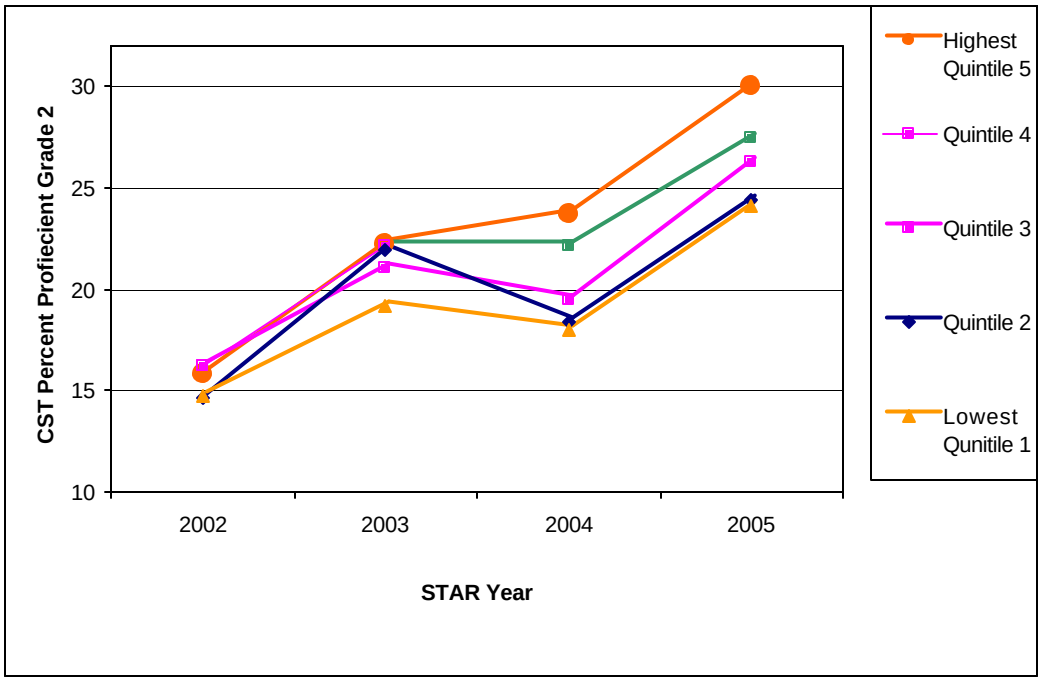



**Figure 4.4.2: Cohort 1 CSTs, 2002 to 2005, Percent Proficient Trend Lines, Grade 3, by Implementation Quintiles**

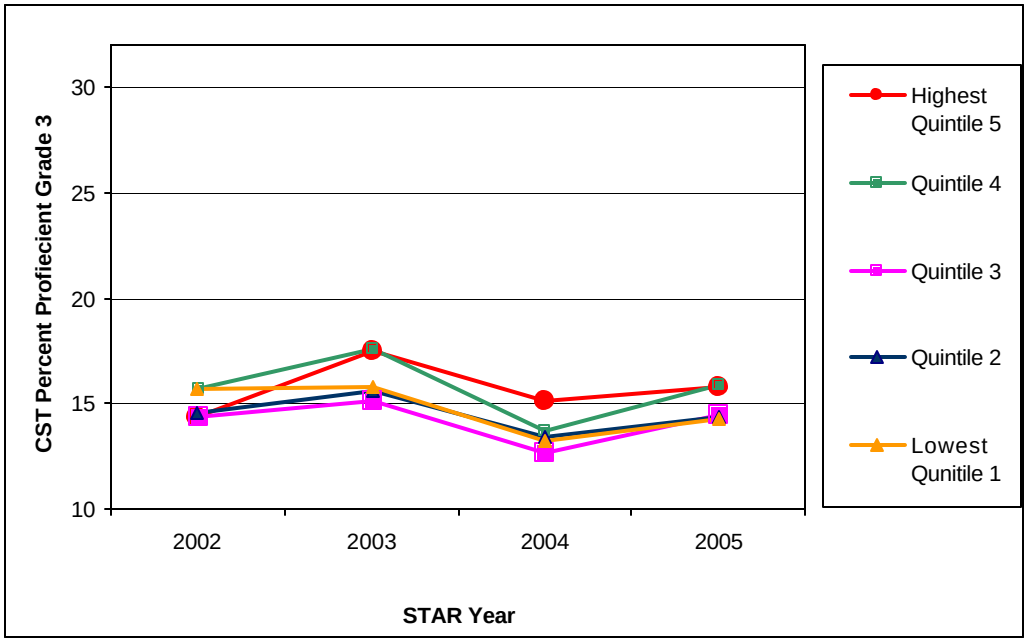


**Table 4.4.1: Prediction of Grade 2 CST School Mean Performance Level, 2004-05**

| Independent/Predictor Variables | Beta (Standardized Effect Size) | t-statistic (t>1.96 implies significance above 95% confidence level) | Significance (probability of occurring by chance) |
|---|---|---|---|
| Starting Point | 0.537 | 17.498 | 0.000 |
| Years in Program | 0.150 | 5.183 | 0.000 |
| RFII Mean | 0.141 | 4.856 | 0.000 |
| RFII Mean*Years in Program [1] | 0.213 | 7.555 | 0.000 |
| School Percent EL | -0.128 | -3.875 | 0.000 |
| School Percent SED | 0.011 | 0.370 | 0.712 |

[1] The "RFII Mean*Years in Program" Beta coefficient was computed in a separate regression analysis with "RFII Mean" and "Years in Program" removed.

**Table 4.4.2: Prediction of Grade 3 CST School Mean Performance Level, 2004-05**

| Independent/Predictor Variables | Beta (Standardized Effect Size) | t-statistic (t>1.96 implies significance above 95% confidence level) | Significance (probability of occurring by chance) |
|---|---|---|---|
| Starting Point | 0.478 | 15.208 | 0.000 |
| Years in Program | 0.041 | 1.366 | 0.172 |
| RFII Mean | 0.083 | 2.745 | 0.006 |
| RFII Mean*Years in Program [1] | 0.076 | 2.554 | 0.011 |
| School Percent EL | -0.183 | -5.390 | 0.000 |
| School Percent SED | -0.028 | -0.851 | 0.395 |

[1] The "RFII Mean*Years in Program" Beta coefficient was computed in a separate regression analysis with "RFII Mean" and "Years in Program" removed.

Tables 4.4.1 and 4.4.2 are interpreted as follows: The first column lists the predictor variables used to predict 2005 CST achievement. The second column provides the standardized effect size, the amount of change in achievement that may be expected given a unit change in the predictor variable, expressed as standard deviation units, or *z*-scores. Thus, for Grade 2, an increase of one standard deviation in the predictor variable called "RFII Mean*Years in Program" is associated with an increase of 0.213 standard deviations in the CST School Mean Performance Level for 2004-05. The third column is the t-statistic, a type of statistical "signal to noise" ratio. When *t* is 1.96 or greater the effect size is statistically significant at the 95% confidence level. The fourth column is computed from *t* and may be interpreted as the probability of achieving that effect size by chance.

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

For both tables, the strongest predictor of 2005 achievement is the school's starting point, its mean performance level in the year before implementing Reading First. This is to be expected and has no special significance other than to convert the 2005 mean performance level into the equivalent of a gain score. The remaining predictor variables are thus conceptually equivalent to predictors of gain scores. The next strongest Grade 2 predictor is "Mean RFII*Years in Program," which has an effect size of 0.213. Note that both are forms of implementation – the mean RFII per year and the number of years that the program has been implemented. Their product is a sensible way of measuring the total amount of Reading First implementation that has taken place in a school. We see that for Grade 2 this product yields an effect size that is both larger and more significant than other predictors of school 2005 achievement, even the percent of English Learners in the school. The effect remains statistically significant in Grade 3 but is dampened by the relative absence of cross-year growth.

Mean RFII on its own is a significant predictor of 2005 achievement for both grades, but Years in Program is not a significant predictor of 2005 achievement for Grade 3 whereas it is very strong for Grade 2. The percent of EL students is a significant predictor for both grades. Percent of SED students is not a significant predictor of 2005 achievement.

It is evident throughout this chapter that Grade 2 and Grade 3 CST scores do not respond to Reading First implementation in the same way. The fact that Grade 2 shows strong gains for both Reading First and non-Reading First schools, whereas Grade 3 does not, suggests two things:

1. The difference between the two grades has little to do with Reading First;
2. For whatever reason, the Grade 3 CSTs are measuring a different language proficiency construct than the Grade 2 CSTs, a construct for which there does not appear to have been much change in teaching practice over the past four years. While Reading First implementation does have a statistically significant effect on this Grade 3 construct, it is much smaller than its effect on the Grade 2 construct.

It is also worth bearing in mind that Grade 3 students have at most only had Reading First-sponsored instruction since Grade 1. None have had Reading First since Kindergarten (though many have been in Open Court or Houghton Mifflin since that time). It is possible that larger gains will appear in Year 4 as students taught under Reading First since Kindergarten move into Grade 3. Nonetheless, it is clear that

either schools are not growing with respect to the Grade 3 CST construct and need to re-gear themselves accordingly, or the Grade 3 CST construct is not targeted on the skills that schools and the State actually think are important. The latter is made more likely by the fact that the Grade 3 CST construct does not appear to match that for the Grade 3 CAT/6.

The figures and tables above yield the following observations:

- Reading First implementation predicts to a statistically significant extent, and almost certainly has a causal effect on, achievement gains.
- The variation in achievement for different implementation quintiles is not an effect of starting point. Schools at all implementation quintiles start at approximately the same achievement level.
- The achievement growth appears to be linear and continuous, setting aside the uniform drop in 2004. No "plateau" effect is yet evident.
- All of the Reading First implementation quintiles show growth in Grade 2.
- Cohort 1 in Grade 2 showed a downward trend on CST scores following the second year of the program for lower levels of implementation.
- SED and EL are not proxies for implementation. That is, the implementation effect is present regardless of the SED and EL level of the school.
- The Grade 2 and Grade 3 CSTs have very different trend lines and do not seem to be measuring the same construct. The Grade 2 construct is much more sensitive to the types of effects caused by implementation of Reading First.

Can the achievement gains of High Implementation schools be explained by demographics?

Figure 4.4.1, taken on its own, makes the case that Reading First implementation *causes* achievement gains. There is room for skepticism, however. Suppose what we *think* is implementation is actually a proxy for some other school characteristic that is associated with the capacity to make large achievement gains. One can imagine, for instance, that schools with lower percentages of Socio-economically Disadvantaged (SED) students might find it easier to generate achievement gains than schools with more SED students. The same might be said for schools with lower percentages of English Learners (EL). Is it possible that the survey questions used to construct the RFII are somehow picking up SED or EL status instead of Reading First implementation? If so, then it is not implementation that is raising achievement but SED or some other hidden demographic variable which is a structural characteristic of the school and

more difficult to change. This would fatally undermine a policy argument that achievement scores could be increased merely by increasing the degree of Reading First implementation.

While the regression equations above show mathematically that implementation is not a proxy for school SED and EL status, the question is important enough to explore more closely. The percentage of SED and EL students in a school is often found to have significant correlations with achievement scores. To examine this relationship, Reading First schools were separated into three demographic groups. These demographic groups are the same referred to as "clusters" in the Year 2 Report. In that Report, clustering methodology was used to select a demographically matched Comparison Group (known then as Comparison Group B). The three clusters consisted of:

1. Schools with high percentages of High-SED and EL students;
2. Schools with high percentages of High-SED and moderate percentages of EL students; and
3. Schools with moderate to high percentages of High-SED and low percentages of EL students.

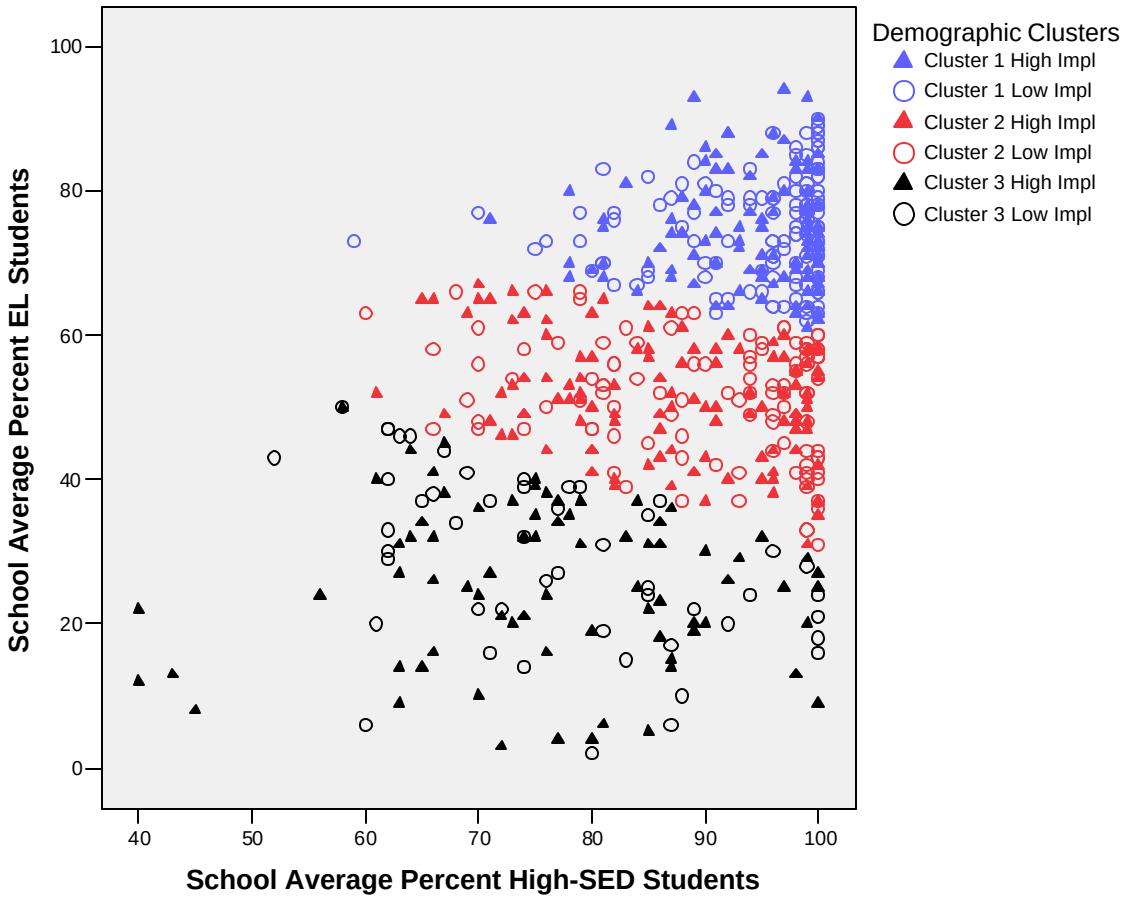
Cohort 1 and 2 Reading First schools were divided into High and Low Implementation groups based on whether their average 2004/2005 RFII statistics were above or below the mean of all the Cohort 1, 2, and 3 schools. Schools with RFIIs equal to or greater than the mean were classified as High Implementation schools. The rest were classified as Low Implementation schools. (This is the same methodology used to create the High and Low Implementation groups discussed elsewhere in the chapter.) All schools were then graphed onto a demographic scatterplot (Figure 4.5) to determine whether the High Implementation schools tend to fall into one of the demographic clusters. That would raise the possibility that what we have been calling "implementation" is actually a proxy for that demographic cluster.

Figure 4.5 shows that High and Low Implementation schools are well-dispersed across the clusters. The only skewness in the dispersion is found among the High-SED/High-EL schools (top-right, labeled 1 and 2), in which the Low Implementation schools tend to be more densely concentrated along the right edge of the graph in the region associated with schools that have 100% High-SED students. Otherwise, "implementation" does not appear to be related to cluster, which supports our case that the RFII is not simply a proxy for a more intractable school demographic.

This finding is reinforced by Table 4.5 in which achievement gains on each metric for High and Low Implementation schools are computed for each cluster. If "implementation" is independent of demographic cluster, the large achievement gain differences between High and Low Implementation schools should be reproduced within each cluster. We find that they are.

The fact that "implementation" as embodied by the RFII is independent of the EL and SED demographic characteristics does not rule out the possibility that it might be dependent on some other characteristic which is not Reading First implementation. There is always room for doubt in this regard in a non-experimental design. It is also possible that the RFII might contain information from survey questions that have some artificially high relationship with achievement gains. We have conducted statistical audits of the survey to test this hypothesis, but so far no evidence of artificial relationships with achievement gains as emerged. As far as can be determined, the RFII does in fact measure Reading First implementation and little else.

**Figure 4.5: Scatterplot of Cohorts 1 and 2 Reading First Schools by Demographic Cluster and High and Low Implementation Schools**




Cluster 1: High percent High-SED and High percent EL students

Cluster 2: High percent High-SED and Moderate percent EL students

Cluster 3: High to Moderate percent High-SED and Low percent EL students

Table 4.5 provides 2-year gains (2003 to 2005) for Cohorts 1 and 2. The reason Cohorts 1 and 2 are merged in this analysis is to yield larger Ns and therefore increase the reliability of these comparisons. For each demographic cluster, gains are provided for High Implementation and Low Implementation schools.

**Table 4.5: 2-year Gains (2003 to 2005) Cohort 1 and 2 Reading First Schools by Cluster and by Program Implementation**

| | Mean Percent Gain | | | | | |
|---|---|---|---|---|---|---|
| | Cluster 1 Schools (High Percent High-SED, High Percent EL) | | Cluster 2 Schools (High Percent High-SED, Moderate Percent EL) | | Cluster 3 Schools (Moderate Percent High-SED, Low Percent EL) | |
| | High Implementation | Low Implementation | High Implementation | Low Implementation | High Implementation | Low Implementation |
| **CST** | | | | | | |
| **Grade 2** | (N=120) | (N=125) | (N=122) | (N=111) | (N=76) | (N=51) |
| Proficient & Above | 6.4 | 5.8 | 6.7 | 3.5[a] | 6.1 | 4.1 |
| Basic | -1.7 | -2.7 | -1.3 | -4.6 [a] | -4.9 | -3.9 |
| Below Basic & Far Below Basic | -4.6 | -3.2 | -5.5 | 1.2 [a] | -1.4 | -0.2 |
| **Grade 3** | (N=120) | (N=125) | (N=122) | (N=111) | (N=75) | (N=51) |
| Proficient & Above | -0.8 | -1.6 | -1.1 | -0.8 | 1.1 | -1.7 |
| Basic | 1.5 | 1.5 | 1.7 | 1.6 | 1.7 | 0.3 |
| Below Basic & Far Below Basic | -0.6 | 0.0 | -0.6 | -0.7 | -3.0 | 1.5 |
| **CAT/6 MeanPR** | | | | | | |
| **Grade 3** | (N=120) | (N=125) | (N=122) | (N=111) | (N=75) | (N=51) |
| Reading | 1.9 | 1.6 | 2.6 | 2.1 | 2.8 | 1.2 |
| Language | 2.5 | 2.3 | 3.5 | 2.8 | 3.8 | 2.6 |
| Spelling | 3.7 | 2.5 | 4.9 | 3.5 | 6.0 | 3.0 |

[a] Statistically significant at $p < 0.05$ as compared to High Implementation schools within the same cluster.
Note: N is always in reference to schools and not students.

Are school-level achievement gains sustainable?

When a program is associated with positive gains, it is sometimes assumed that it is natural for such achievement gains to continue indefinitely. This expectation misconstrues the nature of institutional change. Schools are given the task of moving a population of students from one proficiency level to another. Each year they start with a new cohort that is at an educational level similar to the previous one. In this sense, school progress is not cumulative in the way that student progress is. As a rule an educational program is effective, and has been effectively implemented, if it moves a specified percentage

of students into the proficient and above category each year. As this percentage is approached, the institutional growth curve is likely to flatten. We expect a similar pattern with Reading First. So far, however, Reading First growth trends appear to be fairly linear and do not show signs of flattening.

Does low implementation of the Reading First program hurt schools?

Finally, it is important to ask if low implementation of the Reading First program hurts schools in terms of student achievement? Based on the data presented in this chapter is, we would say no. Tables 4.1.1, 4.1.2 and 4.1.3 compare the gain scores of the average Reading First school with the gain scores of High Implementation and Low Implementation Reading First schools, Comparison group schools, and All Elementary schools. In all three cohorts we see that for most metrics of analysis, Low Implementation schools show gains that are similar to, or higher than, Comparison Group schools and All Elementary schools. If low implementation of the Reading First program were having a negative impact on the schools, we would expect to see Low Implementation schools perform worse than the Comparison Group schools. It is notable that except for Cohort 1 Grade 2, Low Implementation schools move *more* students out of the bottom categories of the CSTs than do Comparison Group schools. We therefore conclude that low implementation of the Reading First program, though not desirable, is still at least as effective in helping the students in the bottom CST categories as Comparison Group schools, and that it certainly is not causing a decline in students achieving standards.

**Conclusions**

As stated in the beginning of the chapter, Reading First will be said to show evidence of being effective if:

1. Achievement gains in Reading First schools are positive;
2. Reading First schools show higher achievement gains than comparable non-Reading First schools;
3. High Implementing Reading First schools show higher achievement gains than Low Implementing Reading First schools.

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

For Cohorts 1, 2, and 3, the answer is "yes" on all metrics except the Grade 3 CSTs. Even on the Grade 3 CSTs, High Implementation Reading First schools show positive growth.

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

Based on comparisons between All Reading First schools and the Comparison Group schools, for Cohorts 1 and 2 the answer is "yes" on almost all the STAR metrics, and the difference is often statistically significant. The pattern is less clear for Cohort 3, but still present. When the comparison is between High Implementation Reading First schools and the Comparison Group schools, the differences are much larger and more often statistically significant. Again, the pattern is less clear for Cohort 3.

*Criterion 3: Do High Implementing schools show higher achievement gains than Low Implementing Reading First schools?*

For Cohorts 1 and 2, the answer is a "yes" based on the STAR data. The effect is clearly pronounced in the Grade 2 CST data and Grade 3 CAT/6 MeanPR data, less clearly in the Grade 3 CST data. The effect is not reproduced as clearly in Cohort 3, but our findings suggest that implementation does not impact achievement until the end of the second year. The effect is not reproduced in the EOY data for reasons that are unknown.

We conclude that the Reading First program is having a positive impact on student growth. Reading First schools show consistent multi-year gains. Against the Comparison Group schools, Reading First schools show higher gains.

One limitation of this study is our lack of understanding of the Comparison Group schools. Little is known about the reading programs they use or the funding they receive under other programs, possibly considerable. It would not be surprising to learn that many of these schools also use the state-adopted programs approved for Reading First and had professional development to support their use. Given this possibility, it is notable that Reading First schools post higher gains than the Comparison Group schools.

We have found it necessary and helpful to study gains by cohort and to chart trend lines across years. This approach makes it possible to notice that Reading First schools generally defer their achievement gains until the end of the second year of program implementation.

The most important finding of this chapter is the close association that exists between implementation and achievement. The data show that gains made by High Implementation schools are significantly higher than those made by Low Implementation schools and by Comparison Group schools. The relationship between implementation and achievement persists across demographic clusters. While there is not sufficient history to claim that there will always be a strong relationship between Reading First implementation and achievement, especially in light of differences between cohorts and achievement

metrics, our findings do justify the expectation that such a relationship will continue to be found in future studies.

## References

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.

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

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

## Chapter 5: Focus Group Interviews

The purpose of this chapter is to examine the nature and depth of California's implementation of Reading First in ways that could not be captured in the quantitative measures documented in the preceding chapters. In this chapter, we report qualitative findings obtained from focus group interviews regarding school-level implementation of Reading First. The focus groups yielded descriptive information about the attitudes of school-based personnel, the extent to which they implemented aspects of the program, and anecdotes about how implementation occurred in their schools. Two experienced reading specialists conducted focus groups with school principals and reading coaches. Questions were designed to accomplish two goals:

1. To verify and support findings from the implementation survey with descriptive information
2. To provide additional information regarding implementation that could not be obtained from survey data

Interviews were conducted in northern and southern regions of the state during September 2005. In all cases, meetings were held at district offices. Interviews ranged from 90 to 115 minutes. Potential interviewee schools were randomly selected from a list of Reading First schools. A letter was sent to the school district Reading First coordinator of each selected school, asking for the principal and coach participation in the focus group. In a few instances, school districts declined to participate due to schedule conflicts or difficulties releasing personnel. A total of 31 districts were represented in focus groups. The number of school principals or assistant principals and coaches participating is provided in Table 5.1.

**Table 5.1 Focus Group Participants**

| | Number of Principals or Assistant Principals | Number of Coaches or Content Experts | Number of Districts |
|---|---|---|---|
| Northern Region | 15 | 18 | 10 |
| Southern Region | 21 | 23 | 11 |
| Total | 36 | 40 | 21 |

**Focus Group Procedures**

Focus group interviews are generally used to gather descriptive information regarding a predetermined topic from a group with similar experiences or backgrounds. Focus groups offer an effective way to examine the perceptions of key stakeholders regarding a specific topic. In a small group setting, participants have multiple opportunities to provide their opinions in response to open-ended questions presented by a facilitator. Participants also have opportunities to respond to comments of other participants. In a focus group setting, one participant's response may trigger a comment from another individual that would not have occurred without the trigger, resulting in a richer understanding of the topic than an individual interview would provide. The interactive discourse regarding a predetermined topic provides information about the collective perceptions of a stakeholder group about that topic. Focus groups are often used in an exploratory way, that is, to discover issues or topics that are important to a particular group. In the case of this Reading First evaluation study, focus group interviews were used to explore points of consensus among principals and coaches at participating schools regarding implementation issues. The focus groups provide illustrations and descriptions of implementation at the school level that are impossible to tap solely with survey data.

Focus group interviews were conducted using procedures that promote interactive dialogue among participants while maintaining a focus on a key topic or focus question. Procedures were based on guidelines for focus group interviews in educational settings (Vaughn, Schumm and Sinagub, 1996). Participants were asked to reflect on their school and district implementation of Reading First during the recently completed 2004-2005 year and center their responses on that time frame. Though interviewers used a pre-determined set of questions, the focus group procedures allowed for some flexibility for the interviewer to follow a topic as it unfolded in the course of the conversation. The facilitator's role was to keep the discussion focused on the topic, ensure that all participants had an opportunity to respond, and probe with additional questions if the discussion lagged.

Each interview lasted 90 to 115 minutes. Participants in each group included personnel in leadership roles in Reading First implementation. Principals, assistant principals, and reading coaches comprised the mixed groups, though in one interview, a district-level content expert participated. Group size ranged from six to ten.

Participants were briefed on the topic of the interview prior to the scheduled date. At the beginning of each session, the facilitator explained the procedures and purpose of the interview. The facilitator explained that responses would be kept anonymous and that only the facilitator would have access to the tape recordings to assist in writing a summary report. The facilitator asked the interview questions in

order and facilitated the discussion, budgeting time to ensure sufficient time for all questions. Each facilitator submitted a detailed summary report to EDS. Table 5.2 provides the introductory script used at the beginning of each session and the interview questions.

**Table 5.2 Focus Group Interview Questions**

**Introductory Script**

The purpose of our focus group today is to reflect on our state's implementation of *Reading First* during the last school year (2004-05) so that we can better understand what is working in the process and how we can improve our implementation. Your insight as a participant in *Reading First* will provide valuable information for our evaluation of California's *Reading First* program. The information you provide in this session will be kept in strict confidence. No names of individuals, schools or districts will be used in reporting the data. I will audiotape the focus group only to refresh my memory as I prepare a summary report. Again, I will not include any names of individuals, schools, regions or districts in the report. I will not start the tape until we have completed introductions.

1. You were all included in *Reading First* during the last school year (2004-05). How would you rate your level of fidelity to the program? Given a scale of 1-10, with 10 being high, what rating would you give your school's implementation, and why?
2. Time seems to be a major challenge for any educational initiative. How does your school manage the time commitment of *Reading First*? How do you protect the reading block?
3. How did the *Reading First* professional development and follow-up support impact the teachers and students in your school? What evidence do you see that the professional development is being used during daily classroom instruction?
4. What was your role in supporting *Reading First* at your school? How did *Reading First* change your daily routine and activities? How did you respond when you observed variation among classrooms?
5. If I were to walk into one of the best *Reading First* classrooms in your school, what would I see? What would make it different from other primary grade classrooms in your district or around the state?
6. All *Reading First* schools are using common assessments. To what extent did the assessment process impact your teachers? How do you know the assessments used made a difference in instructional delivery and student achievement?
7. If I were to come to your school to observe a grade-level collaborative meeting, what would it look like? Who would be there and what would be their roles? How do you know there is follow-up as a result of these meetings?
8. Did you supplement your core program with any supplemental materials or programs? Why or why not? If you did supplement, how did you determine the alignment with the core program?
9. Given our conversation today, how do you now feel about your school's fidelity of implementation? Did your rating change?
10. Today, you have had an opportunity to hear about implementation of *Reading First* at other schools. Did this conversation prompt you to think of any changes you would like to make in your school this year? What ideas will you take away from this meeting?

**Data Analysis Procedures**

The facilitators conducted the first level of data summary by compiling a summary report by question across interview sessions. One report was compiled for the northern region of the state and one for the southern region. The principal investigator (PI) then analyzed the summary reports using a systematic qualitative coding approach. The narrative reports were coded independent of interview questions to capture themes regarding implementation of Reading First from the whole data set rather than constricting themes to predetermined categories. The PI first read through the summary reports to identify significant themes that were evident across questions and region. Then, the narrative text was examined to identify conceptual units, groups of words that communicate a distinct message or bit of information. Each conceptual unit (i.e., phrase, sentence or paragraph) in the narrative report was independently evaluated using a constant-comparison approach to determine the extent to which it belonged to an existing theme or stood alone as a new theme (Miles & Huberman, 1994). Similar conceptual units were grouped together under the themes and thus characterized the theme with illustrative quotes and phrases. Unique but significant responses were also recorded throughout the coding process.

**Qualitative Findings**

In this section, each theme is described using summary statements, excerpts and quotes.

Theme 1: Length of time in the Reading First program makes a difference in implementation.

Respondents consistently reported that there were significant effects of having extended time in the Reading First program. There were several changes that took place over two or three years of participation. Most participants reported that the teachers in their school underwent a significant change in their beliefs and attitudes about the program and about reading instruction in general. Most interviewees cited initial resistance, or even fear, especially regarding the assessment routine, but resistance and fear diminished with the second and third year of implementation. Gains in reading scores were a significant impetus for these changes. Interviewees cited the frequent and continuous focus on assessment data as an impetus for change not only in attitudes, but also in practice.

One interesting comment reported differences in teachers' implementation over time depending on their years of experience: "The veteran teachers' level of implementation is high, but new teachers' implementation is lower even though they received the same professional development. It takes time in the program and experience to get high implementation."

Another important change that occurred over time was depth of understanding. Teachers gained a deeper understanding of the research with more time in the program. In addition, with more time in the program,

teachers gained a better understanding of how to teach students with special needs, including English-language learners and students with disabilities.

Theme 2: Reading First has contributed to the professionalization of the teaching force.

This theme is characterized by a variety of comments that had to do with how Reading First changed the ways teachers, coaches and principals think about their roles and how they operate on a daily basis. One interviewee observed a positive effect of a skilled professional workforce in place, "Highly qualified teachers are now at the lowest performing schools." There are several ways that Reading First has helped to define the professional roles of teachers, coaches and principals.

Reading First has provided a "common language" not only for professional communication but also classroom discourse. Principals, coaches and teachers now have a "common language" for communicating and learning about assessment and instruction. This consistent language is infused in professional development, regular meetings, impromptu discussions and coaching sessions. A common reading program in the school has given teachers a shared instructional focus. Numerous respondents reported a common language of instruction in the classroom. They cited instances of walking into different classrooms and hearing consistent language used in explaining concepts and strategies to students. There is coherence in instructional delivery, collegial planning and teamwork, fidelity to the program and pacing of instruction. There was consensus among participants that this was a positive change in the schools. The majority of principals reported that Reading First was a major support for them in establishing a school focus and one principal noted, "Reading First prevented us from starting over again every year."

Evidence of professionalism was also present in comments that had to do with gaining knowledge and skill in reading instruction. Teachers, coaches and principals have deepened their knowledge of reading curriculum, instructional practices and assessment. Though one might expect "fatigue" with so much professional development, there was no evidence of such. The interviewees enthusiastically reported positive changes as a result of gaining knowledge and skills with regard to reading instruction. As one participant commented, "Some teachers are meeting on their own time to improve their understanding of the program and to plan for instruction." Several comments indicated that teachers and coaches readily share ideas and experiences during meetings. In addition, several comments had to do with teachers' gaining insight through opportunities to be reflective about their craft: "Teachers are able to reflect on their own practice and have a deeper understanding of literacy development."

A notable finding has to do with the added benefit of extended professional development: "There is a difference between teachers who have reached mastery level and those who have not in the instructional language used in their classrooms."

Principals' roles have changed as they systematically visit classrooms and engage in discussion with teachers and coaches: "It is critical that both principals (assistant principals) and coaches are in classrooms daily so teachers become comfortable with this 'new way of doing business'."

Theme 3: assessments and data have been a driving force for change in Reading First schools.

Assessment is a central focus in Reading First schools. Respondents consistently recounted scenarios of the use of assessment data to guide discussions at meetings and instructional planning.

Seeing gains in student scores was the primary impetus for changes in teachers' enthusiasm. "At first, teacher buy-in to the program was at a one or a two (low), but that has moved up now because they see results." Assessment summaries are regularly included in grade-level meetings and some schools require teachers to organize data by grade and by class in a three-ring binder so teachers, coaches and principals can all see the data.

The frequency and consistency of assessment has led directly to changes in the classroom. First, assessment pushes teachers and students through the curriculum and thus to higher levels of learning. One comment stated, "Having access to the *Online Assessment and Reporting System (OARS)* has required teachers to complete instruction prior to the assessments," and another, "Teachers have moved away from complaining that the assessments are too difficult for the students now that they see their students moving up to higher levels of ability."

Assessment has also led to increased differentiation of instruction for varying levels of students or struggling learners. Said one respondent, "OARS is graphic; it is there; it's THE evidence of student learning… Those students who have instructional needs 'yell at you' through the colored data system." From the assessment data, teachers now see the purpose in "workshop" time or "universal access time" (i.e., time in small groups to differentiate instruction for varying student needs). One benefit of using a data reporting system is the ability to follow students over time. Multiple-year access to data reports allows teachers to view individual students' progress over time. There is more continuity within a school and teachers are using data reports in meetings with parents.

Theme 4: Coaching is a vital component of Reading First.

Interviewees unanimously supported the implementation of coaching in Reading First. Many comments noted the effectiveness and importance of the coaches' training. Coaches were reported as highly

competent due to depth of training, and that their knowledge has transferred to the classroom. One respondent noted the impact of coaching on the classroom with "that daily support that provides assistance every step of the way." Coaches universally affirmed the value of coach training.

The coaching process was noted in several comments. For the most part, principals and coaches felt that coaches played a vital role in helping teachers implement what they learned in their professional development. Coaches felt that there was a positive correlation between the level of professional development a teacher had received (i.e., entry level versus mastery level) and their willingness to work with the coach. Coaches reported that professional development opened the door for them because of common understandings and language. Coaches viewed their role as being clarified by Reading First. Reading First made the coaches seem less "on their own." They also reported more collaboration with the principal. However, some coaches felt they were not serving in their primary roles to the extent they felt they should and wanted to. They were not in classrooms to any appreciable extent because of other assigned duties—some related to Reading First and some not.

There was disagreement as to whether the coach should have a primary or secondary role in grade-level collaborative planning meetings. One respondent noted that coaches "made an effort to participate." Across the groups, there seemed to be a direct relationship between coach participation in grade-level meetings and the number of schools for which the coach was explicitly responsible for the meetings. There was agreement that the coach should participate, but it was not always logistically possible.

Theme 5: Reading First has led to increased collaboration in schools and collaboration has promoted implementation.

Interview participants reported increased collaboration in their schools as a result of Reading First, particularly with regard to analyzing data and planning instruction. Collaboration was important within and across grade levels. For instance, one comment was, "grade-level meetings provide opportunities for discussing and sharing new ideas as well as clarifying new learning," while another was, "vertical team meetings is where the continuity of the program is seen and a better understanding of the program as a whole is gained." One group used multi-grades for their meetings and had found this as a solution to the problem of combination classes—teachers became more willing to exchange students for reading. If there was a 1-2 combination, the first grade teachers took responsibility for reading instruction for the first graders. Some of the outcomes observed from the grade-level meetings included additional requests for coach support, uniformity in instructional approaches and strategies, a more trusting environment among and between teachers, more discussion and conversation related to program implementation, focus on assessment results, and studying the teacher manuals in pairs or groups of teachers. Some participants

reported on the benefits of half-day and full-day grade-level meetings so that an in-depth study of assessment results could take place.

School principals and coaches asserted that grade-level meetings were held in all schools. Generally, the meetings were attended by grade-level teacher teams and led by teachers. Meeting requirements included agenda, sign-ins, and plans for action between meetings. Student data was to drive the meeting. There was some disagreement as to what extent principals and coaches were and should be involved in these meetings.

Though increased collaboration was highly valued by interviewees, the process was not without difficulties. One comment stated, "Having to implement and keep on top of the three programs (one English and two Spanish) at the school is challenging because all teachers are not able to talk about the same reading program when they are brought together to collaborate." Some participants felt that the 80 hour follow-up (Passport) was difficult to accomplish. If the principal did not specifically require it, it was not likely to happen with all teachers. When there *was* follow-through on the 80 hours, it was thought to be very positive in driving collaboration, data analysis, and ongoing support for new teachers.

Theme 6: Reading First has increased administrators' participation in classroom-level implementation.

Principals discussed how Reading First had changed their daily routines and practices. Said one principal, "*Reading First* has provided structure and purpose to my day" and has led to opportunities for goal-setting at both staff meetings and teacher collaboration meetings. Some principals reported taking an active role in ensuring fidelity to the program. Some means for doing this included having teachers share aloud what they learned in professional development and this allowed principals to follow through with individual teachers in classroom visits and one-to-one meetings with teachers. Principals felt they need to convince teachers that feedback is good as it is a way for all to grow professionally.

Theme 7: Competing and conflicting initiatives are a constant threat to the school's focus on Reading First.

Principals expressed concern that competing or conflicting initiatives threaten to pull attention away from reading instruction and assessment data. Though no specific initiatives were identified and the interviewers did not probe for specific examples, except for a comment about training in other subject areas, some comments indicated that other initiatives served as a distraction. When there are multiple initiatives in place in a school, academic or otherwise, it is difficult to allot the time and attention needed for Reading First. Principals discussed the difficulty of maintaining focus both for themselves and the school due to district mandates, other grants and programs. In some cases, these initiatives may conflict

with the principles of Reading First, such as using reading materials or methodology that conflict with the adopted program. This was the exception rather than the rule. In some comments, the other initiatives had a similar focus on fidelity of implementation or compliance with requirements and this strengthened their ability to monitor and facilitate Reading First.

**Supplemental Programs and Materials**

Most principals and coaches reported that adherence to the guidelines of Reading First left little time for supplementary materials. Principals felt that Reading First gave them the permission to say, "You can't!" Many reported holding a tight line and limiting the use of supplements to ancillary material of the Open Court or Houghton Mifflin programs. Said one principal, "The 'little books' are packed away!"

When the use of supplemental materials was cited, it was in the areas of writing and intervention. There was frequent reference to the use of the writing program *Step Up to Writing*. Schools beginning the second year of the program indicated a need to supplement the writing strand of the core. Three supplemental programs that were identified were *Writer's Workshop, Writing for Excellence,* and *Six Traits for Writing.* In the area of intervention, many used intervention from the Special Education Referral Reduction Program list for assisting students over and above the core instruction. It was not uncommon for other programs to be used for extended day programs. *Accelerated Reader* and *Waterford* were named as supplemental programs being used in some schools. Most schools that have used the core program for more than two years have eliminated all programs in-class and are using only core program materials.

**Challenges and Issues To Address**

Though the general tone of the meetings was positive and reported many benefits of Reading First, there were various challenges cited. These did not fall into categories, so are listed here in no particular order.

1. Large schools that are on year-round, multiple-track calendars are difficult to manage because teachers from room to room are on differing pacing calendars, which makes it difficult for the principal to determine appropriateness and efficacy.
2. Principals have had insufficient training to fully support Reading First.
3. Coaches are pulled away for too many meetings.
4. Finding enough time is a constant challenge both in the classroom and for meetings and follow-up.
5. The drop in third grade CST scores was discouraging to teachers, coaches and principals.

6. Some coaches have too many responsibilities and often serve multiple schools.
7. Schools report receiving information and materials late.
8. Turnover in school leadership detracts from continuity.
9. Spanish bilingual classes are not as well supported in the program as are English classes.

## References

Vaughn, S., Schumm, J. S., & Sinagub, J. M. (1996). *Focus Group Interviews in Education and Psychology*. Thousand Oaks, CA: Sage Publications.

Miles, M. B., & Huberman, M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: Sage Publications.

# Chapter 6: Findings and Recommendations

California has completed three years of implementation of Reading First. Though the first year was a "start-up" year and implementation began toward the middle of the academic year, in this report we have an opportunity to examine three years of achievement and implementation data and examine trends in the impact of Reading First on California's elementary schools. There is substantial evidence that it takes extensive time and effort to accomplish significant and lasting change in school reform (e.g., Fullan, Good, Legg Burross & McCaslin, 2005; St. John, Manset-Williamson, Choon-Gung, & Michael, 2005). At this time, it is possible to begin to examine long-term benefits and challenges of implementing the ambitious reform effort of Reading First.

Chapter 6 summarizes the findings of the Year 3 Reading First evaluation study and presents an analysis of implications for the continuation of Reading First as well as more general policy issues. Findings are drawn from the previous chapters and are guided by the two overarching research questions stated in Chapter 1: How well has the Reading First program been implemented in participating schools and districts? And, What impact has the Reading First program had on participating schools? We first summarize the evaluation of implementation, then review the achievement findings, then examine the interaction of achievement and implementation.

## Findings

Findings are summarized by the following research questions.

### *How well did participating districts and schools implement their Reading First grants in accordance with California's Reading First plan?*

Chapter 3 outlines a rationale for building careful measurement of implementation into the evaluation of Reading First. In the absence of implementation measures, it would be difficult to ascertain the extent to which significant (or non-significant) achievement results can be explained by the program. Furthermore, an evaluation of implementation provides information about which aspects of the program are working well and which may need improvement. Chapters 3 and 5 provide an in-depth examination of the implementation of Reading First from the perspectives of teachers, principals and reading coaches. Appendix H reports an implementation statistic for each Reading First school where possible.

The 2003-2004 academic year (Year 2) was the first year the implementation survey was used. That year involved significant effort to develop and refine a valid and reliable measure. With the 2004-2005 survey data, we now have a better picture of implementation. Calculation of the Reading First Implementation Index (RFII) provides a global look at a school's implementation across different dimensions of implementation. Following are important findings from the implementation survey.

- The basic elements of Reading First have been adequately implemented in Reading First schools. Ninety-six percent of schools in 2005 were rated "adequate" or better by teachers in Reading First schools.
- In examining the dimensions of implementation derived from the survey, the Professional Development and the Coaching/Teacher Implementation dimensions receive the strongest ratings.
- The Reading First Implementation Survey, completed by teachers, principals and coaches, yields a measure of each individual school's implementation called the Reading First Implementation Index (RFII). The mean RFII for 2005 was 36, the same as that for 2004.
- Nonetheless, Cohorts 1 and 2 have raised their levels of implementation since the 2003-2004 school year, supporting the finding that implementation takes at least one or two years to take hold.
- Cohort 3 has begun at a lower level of implementation than Cohorts 1 and 2 did in their first years (though the RFII for Cohort 1 in its first year is not known) due mainly to lower levels of teacher and coach professional development. Because many Cohort 3 schools are in rural areas, it is possible that they have less access to high quality professional development, but the full reasons for Cohort 3's slow start are not yet known.

***What aspects of implementation have impacted Reading First schools?***

Findings from the focus group interviews reported in Chapter 5 "unpack" implementation issues by showing where there is consensus regarding implementation issues and providing anecdotal information about the benefits and challenges of implementing Reading First.

- Focus group interviews provide strong validation for the assessment, coaching, and collaboration elements of Reading First. There was consensus among coaches and principals that these aspects have had a significant and lasting impact on Reading First schools.
- Reading First has led to noteworthy improvement in the professionalism of the teaching force and work environment of schools. Teachers, coaches and school principals have deepened their

knowledge and professional skills in the area of reading instruction and, after sufficient exposure to the program, tend to demonstrate a strong commitment to the program. Being at the same point in the curriculum pacing schedule at the same time reinforces the ability of teachers to find a common professional vocabulary and to consult with their colleagues.

- A strength of the Reading First program has been the consistent message of full implementation and strong focus on assessment results. Competing initiatives and agendas in districts and schools are a constant threat to the success of a sweeping reform program like Reading First.
- Reading coaches play a vital role in professional development and program implementation in Reading First schools. The coaching force has become highly knowledgeable and skilled over time in Reading First and is now viewed as indispensable in ensuring implementation.

***What is the impact of the Reading First program on K-3 students in participating schools and districts?***
To answer this question, we first examine achievement data from Reading First schools over time. Only Cohort 1 and Cohort 2 schools have been in the program for multiple years. We then compare Reading First and non-Reading First Comparison Group schools to examine the differential impact of Reading First. Finally, we compare the achievement data of High Implementation and Low Implementation Reading First schools to determine the degree to which the program is effective when implemented.

<u>Reading First Schools Compared to Themselves Over Time</u>

- Reading First schools consistently show strong upward trends in the percentage of students in the Proficient and Advanced categories of the CST performance scale for Grade 2. These trends are apparent for all three Reading First cohorts. (Cohort 1 entered the program in the 2002-2003 school year, Cohort 2 in the 2003-2004 school year, and Cohort 3 in the 2004-2005 school year.) On the Grade 3 CST metric the trend lines are flat, reflecting a statewide trend.
- There is evidence that the academic advantage of participating in the Reading First program increases with time in the program. In school-level regression models of CST achievement, Years in Program is found to be a statistically significant predictor of 2005 CST scores for Grade 2 (but not Grade 3) after controlling for starting point and percent of English Learners.
- Reading First schools with the highest percentages of socio-economic disadvantage (SED) and English Learner (EL) students have shown significant gains in students scoring at Proficient and Above while also showing decreases in students scoring in the bottom CST categories.

- Reading First schools show substantial percentages of students moving out of the bottom CST performance categories (Below Basic and Far Below Basic). Similar trends are observed with regard to the California English Language Development Test (CELDT) Beginning and Early Intermediate categories of English language development. While this may be interpreted as evidence that Reading First leads to strong gains in English proficiency for both English-speaking and non-English speaking students, more systematic study of this issue is warranted. This study was not designed to fully address this question and results are preliminary at best.

Reading First schools as Compared to Non-Reading First schools

- When compared to a demographically matched sample of non-Reading First schools called the "Comparison Group," all three cohorts of Reading First schools show somewhat larger achievement gains than the Comparison Group over time, though the differences are often not significant. Why the differences between Reading First schools and Comparison Group schools are not more significant may, perhaps, be explained by a recent history of statewide and district reading initiatives that may have impacted Comparison Group schools. At present, however, such a history is not available, making Reading First / non-Reading First comparisons hard to interpret.

- High Implementation Reading First schools show higher gains and sharper growth curves than those of the non-Reading First Comparison Group.

- Reading First schools show larger percentages of students moving out of the bottom CST performance categories than Comparison Group schools do.

- These patterns are consistent across the Grade 2 and Grade 3 STAR test CSTs and the CAT/6.

High Implementation Reading First Schools as Compared to Low Implementation Reading First Schools

- When student achievement is disaggregated by school implementation level, the gains made by High Implementation schools are significantly higher than those for other Reading First schools, especially Low Implementation Reading First schools. In school-level regression models of CST achievement, Reading First implementation was found to be a statistically significant predictor of 2005 CST scores for both Grades 2 and 3 after controlling for starting point and percent of

English Learners. This *supports* (but does not *prove*, given the limitations of our research design[9]) the hypothesis that implementation of Reading First causes achievement to rise.

- High Implementation schools not only move more students into the Proficient and Above category, but the rate at which this movement occurs increases dramatically in the second and third years of program implementation. The focus group interviews support this finding. It can be concluded that Reading First takes at least 1 to 2 years to become integrated into a school's teaching and learning environment. Once that happens, it may be expected to have a strong effect on student learning and achievement. As mentioned, Years in Program is a significant predictor of CST achievement gains for Grade 2, though not for Grade 3.

- High Implementation schools are more successful than Low Implementation schools in moving students out of the bottom performance categories of the CSTs.

- Nonetheless, Low Implementation schools do not "lose ground" as a result of lower program implementation and do show achievement gains for many of the metrics measured. Their rate of growth parallels that of non-Reading First Comparison Group schools. Though low implementation is not "harmful," schools do not obtain the achievement gains associated with consistent and full implementation. This is a phenomenon that warrants further study.

- High Implementation schools are well-distributed along the socio-economic continuum of Reading First schools. High Implementation schools show higher gains in student achievement than Low Implementation schools regardless of where they are on that continuum.

Taken as a whole, these comparisons allow us to publish a reasonably confident finding that Reading First is working and that the effect is enhanced by higher degrees of implementation and length of time in the program.

***What evidence is there that the Reading First program has improved the effectiveness of participating schools and districts?***

As discussed above, the evidence that Reading First has improved the effectiveness of participating schools and districts is derived from school achievement and implementation data. When Reading First schools are studied by cohort over time, we see that as a group they show positive growth in student

[9] Because Reading First schools are not randomly assigned to Low Implementation and High Implementation groups as a true experimental design would require, no causal inferences can legitimately be made. We can only say that implementation significantly *predicts* achievement ($p<0.05$). Nonetheless, our findings are consistent with what would be expected if there were a causal relationship.

achievement. When compared to a demographically matched Comparison Group, Reading First schools show larger movements of students into the Proficient and Above categories and larger losses out of the bottom performance categories. In addition, implementation of the Reading First program is directly associated with gains in student achievement. High Implementation schools show significantly larger gains than the average Reading First school. Low Implementation schools tend to parallel the growth curves of the average non-Reading First elementary school and Comparison Group school. The fact that Low Implementation schools do not show significantly lower gain scores suggests that though strong implementation of the program certainly impacts student achievement, weak implementation does not necessarily harm student achievement. High Implementation schools continue to show larger gains than Low Implementation schools when controlling for socio-economic disadvantage and proportion of English learners. This suggests that the implementation level of a school is not an artifact of its demographic makeup.

***Anomalies, Unintended Consequences and Intriguing Results that Warrant Further Study***

One aim of this evaluation study was to identify unintended consequences that may be a result of Reading First. In the course of three years of implementation of the program, several issues arise that warrant further study. These include:

- The End-of-Year test results do not appear to be completely consistent with the STAR results. The reasons for this are not yet understood. Further studies should examine the reliability, validity, administration, and data management of the EOY tests.

- The RFAI appears to be a useful tool for measuring and monitoring a school's overall use and benefit from Reading First, in particular for assessing "significant progress." Since this Year 3 report is only the second year of its use, and since it includes an End-of-Year test component, the RFAI index should be further explored. Accurate, timely and consistent data reporting will facilitate this effort.

- There is a tendency for the percentage of students in Reading First schools who are at the level of proficient in English Language Development to increase while the percent of students in the lower categories of proficiency decreases. The CELDT results suggest that extended time in Reading First leads to better results in assisting EL students to acquire English. Further studies might explore *which* aspects of Reading First contribute to these findings (e.g., teacher professional development, curriculum, scientifically based instructional principles). It would be important to further explore differences in instruction, curriculum and procedures between Reading First schools and Comparison Group schools.

- The promise of Reading First to improve the quality of teachers at the state's most needy schools appears possibly to be a promise fulfilled. There is a trend for differential improvement in teacher qualifications in Reading First schools and Comparison Group schools. Though all California schools reported an increase in the number of teachers with full credentials, likely due to the *No Child Left Behind Act,* Reading First schools reported increases in teachers seeking advanced degrees and university course credit. Given the national interest in improving teacher qualifications, further exploration of this issue is important.

- In general, future reports should further explore disaggregated results. There is some evidence that Reading First may affect different groups in different ways. Some possible issues to explore would be urban versus rural locations, ethnicity and language issues, cohort differences, and grade level differences.

- This report includes clear evidence of differences in achievement for high and low implementation groups. Future reports should further define these groups and examine differential results over time and through the grades. There is evidence that in some cases non-implementers (i.e., the non-Reading First Comparison Group) outperformed low implementing groups. This may be in part due to inherent differences in these groups as well as differential starting points on academic achievement. The focus groups suggested that some schools may have difficulty receiving information and materials in a timely fashion, though no specific information was provided regarding the source of the problem. Because implementation relies on accurate and timely access to information and materials, this issue should be monitored closely.

- There is some evidence of differential impact of professional development on cohort groups. On the implementation survey, Cohort 3 reported less use of professional development, and anecdotal reports suggest that a substantial number of teachers are either reluctant to receive further professional development or are experiencing obstacles in trying to receive it. This may, in part, reflect the fact that there are more rural districts in Cohort 3 and they may have less access to high quality professional development.

**Policy Implications**

Observation 1

Reading First schools show increases in student achievement scores over time. High implementation schools show higher rates of growth than non-Reading First schools.

This finding is a strong indication for continuing to implement all aspects of Reading First. A strong focus on full implementation is warranted and some evidence suggests that clarifying the roles of principals and coaches in the day-to-day operation of Reading First in schools may support full implementation. Cohort 3 shows a particular need for stronger support in the area of Professional Development, especially at the LEA level.

Observation 2

Reading First cohorts started with the Reading First program at different points and they are demographically different. Therefore outcome results should always be studied at the cohort level. Combining cohorts can mask the real story. Time in the program is a critical factor in both reading achievement and implementation.

Observation 3

It is critical to study academic growth in relation to program implementation. Variation in program implementation has differential impact on academic growth, therefore it is important to separate schools on the implementation continuum. This is possible only when schools can be measured on implementation. The Reading First Survey administered to teachers, coaches and principals provides valuable information about implementation at the school level. The RFII is constructed from the response data to the surveys. Thus, the use of the Survey and computation of the implementation index should be continued in future years of the study.

Observation 4

The qualitative data collected from the focus group sessions corroborated an important inference made from the outcome data, which is that the length of time that schools have had Reading First funding has an impact on implementation. Usually teachers (especially less experienced teachers) need longer to embrace the program in its entirety. Data from focus group interviews also tells us that the Reading First program has helped to improve teaching practices and increase collaborative work at the school level.

**Policy Recommendations**

Recommendation 1

*Continue to focus on full implementation of Reading First.*

Review of school RFII measures suggests that Reading First is being adequately implemented across the state in essential respects, in particular professional development, program materials, coaching, and instructional practices. However, our findings show that high-implementing schools yield higher academic gains than moderate- or low-implementing schools. State and local Reading First personnel should focus extra efforts on schools with lower levels of implementation to ensure maximum benefit from participation in Reading First. Cohort 3 in particular may need extra monitoring and support from LEA's and the state and regional technical assistance services in this regard.

Recommendation 2

*Support participation in Reading First over multiple years. It takes at least two years of implementation to show significant achievement gains, even with extensive training and support. Continued support beyond the initial two years is essential to achieve significant and lasting results and to establish the long-term institutional changes needed for Reading First instructional practices to continue even after funding is discontinued.*

Quantitative and qualitative data support the notion that continued program participation leads to continued and lasting improvement in teaching practices and student outcomes. Extended time (e.g., 6 years) will allow school personnel to gain depth of knowledge, refine their skills, and integrate program principles into the fabric of their school operations. On the other hand, if individual schools show little evidence of benefiting from the program after three years of support, they should be dropped from the program as the Reading First NCLB legislation requires. Cohort 1 schools show steady academic gains over three years of participation, supporting the premise that extended support and participation leads to continued improvement.