Evaluation of Reading First in the Montgomery County Public Schools 2003–2006

Department of Shared Accountability

February 2007

Scot W. McNary Ph.D.

Montgomery County Public Schools

DEPARTMENT OF SHARED ACCOUNTABILITY

Dr. Faith Connolly, Director
850 Hungerford Drive
Rockville, Maryland 20850
301-279-3925

Dr. Jerry D. Weast
Superintendent of Schools

John Q. Porter
Deputy Superintendent
for *Information and*
Organizational Systems

Table of Contents

Executive Summary ix
Findings from Implementation Evaluation ix
Findings from Effectiveness Evaluation x

Introduction 1

Background 1
Scientifically Based Reading Research 2
MCPS Reading First 2

Evaluation Design 3

Methodology 4
Study Sample 4
Implementation Evaluation Data Collection Activities 4
Student Achievement Measures 5
Data Analysis 5
Data Availability 6

Results 7
Staff Interviews 7
Summary 11
Parent Interviews 11
Training 12
Classroom Observations: Consistency with MFRI Recommendations 12
Standardized Tests 13
DIBELS Subtests 16
Inferential Tests: Comparisons of Reading First and non-Reading First Schools 20
DIBELS Subtests 23
Discussion 25

Implementation Evaluation 25
Outcome Evaluation 25
Conclusion 27
Recommendation 27

References 28

Appendix A Target Population Demographics 30
Appendix B Staff and Parent Interview Protocols 32
Appendix C Training Data 42
Appendix D Assessment Data 45
Appendix E Results of Alternative Analyses of Student Achievement Data 62

List of Tables

Table 1	Achievement Test Data Availability……	7
Table 2	Staff Evaluations of Quality of Reading First Implementation and Whether to Recommend Reading First to a Colleague……	11
Table 3	Mean Minutes Spent in Reading First Recommended Instructional Activities, by Grade…….	13
Table 4	Differences Between Reading First and Comparison Schools for Grade 3 MSA Reading, Grade 3 MAP-R, and Grade 2 CTBS/TerraNova, by Cohort……..	22
Table 5	Differences Between Reading First and Comparison Schools for DIBELS Subtests 2005–2006, by Grade……	24
Table A1	Demographics of Target Population……	30
Table C1	Grades Taught by Reading First Teachers Attending any Training in 2004–2005……..	42
Table C2	Trainings Attended by Reading First Teachers in 2004–2005……...	42
Table C3	Grades Taught by Reading First Teachers Attending any Training in 2005–2006……..	43
Table C4	Trainings Attended by Reading First Teachers in 2005–2006……	43
Table C5	Inter-rater Agreement (Kappas) for Training Videotape and Classroom Observations……	44
Table D1	Grade 3 MSA Reading Means and Standard Deviations, by Year and Demographics and Service Characteristics…….	45
Table D2	Grade 3 MSA Reading Means and Standard Deviations, by Year and School……..	45
Table D3	Grade 3 MAP-R Means and Standard Deviations, by Year and Demographics and Service Characteristics…….	46
Table D4	Grade 3 MAP-R Means and Standard Deviations, by Year and School…..…….……...	46

Table D5 Grade 2 CTBS/TerraNova Reading Means and Standard Deviations, by Year and Demographics and Service Characteristics……. 47

Table D6 Grade 2 CTBS/TerraNova Reading Means and Standard Deviations, by Year and Demographics and Service Characteristics……. 47

Table D7 Kindergarten Student 2004–2005 DIBELS End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools…….. 48

Table D8 Kindergarten Student 2004–2005 DIBELS End-of-Year Test Performance, by School …… 49

Table D9 Kindergarten Student 2005–2006 DIBELS Letter Naming Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and Non-Reading First Schools…… 49

Table D10 Kindergarten Student 2005–2006 DIBELS Letter Naming Fluency End-of-Year Test Performance, by School…….. 50

Table D11 Kindergarten Student 2005–2006 DIBELS Phoneme Segmentation Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and Non-Reading First Schools……. 50

Table D12 Kindergarten Student 2005–2006 DIBELS Phoneme Segmentation Fluency End-of-Year Test, by School……. 51

Table D13 Kindergarten Student 2005–2006 DIBELS Nonsense Word Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and Non-Reading First Schools…… 51

Table D14 Kindergarten Student 2005–2006 DIBELS Nonsense Word Fluency End-of-Year Test Performance, by School…….. 52

Table D15 Grade 1 Student 2004–2005 DIBELS End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools……. 53

Table D16 Grade 1 Student 2004–2005 DIBELS End-of-Year Test Performance, by School…….. 54

Table D17 Grade 1 Student 2005–2006 DIBELS Phoneme Segmentation Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and Non-Reading First Schools……. 55

Table D18 Grade 1 Student 2005–2006 DIBELS Phoneme Segmentation Fluency End-of-Year Test Performance by School……... 56

Table D19 Grade 1 Student 2005–2006 DIBELS Nonsense Word Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and Non-Reading First Schools……. 56

Table D20 Grade 1 Student 2005–2006 DIBELS Nonsense Word Fluency End-of-Year Test Performance, by School…….. 57

Table D21 Grade 1 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and Non-Reading First Schools…… 57

Table D22 Grade 1 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by School…….. 58

Table D23 Grade 2 Student 2004–2005 DIBELS Oral Reading Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools……. 58

Table D24 Grade 2 Student 2004–2005 DIBELS Oral Reading Fluency End-of-Year Test Performance, by School…….. 59

Table D25 Grade 2 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and Non-Reading First Schools…… 59

Table D26 Grade 2 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by School…….. 60

Table D27 Grade 3 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools……. 61

Table D28 Grade 3 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools……. 61

Table E1 Differences Between Reading First and Comparison Schools for Grade 3 MSA Reading, 2003–2004, 2004–2005, and 2005–2006…… 63

Table E2 Differences Between Reading First and Comparison Schools for Grade 3 MSA Reading, 2003–2004, 2004–2005, and 2005–2006 with Grade 2 CTBS as Covariate …… 64

Table E3 Differences Between Reading First and Comparison Schools for Grade 3 MAP-R RIT Scores, 2004–2005 and 2005–2006 (Winter)……. 65

Table E4 Grade 2 Differences Between Reading First and Comparison Schools for CTBS Reading Scale Scores, 2003–2004, 2004–2005, and 2005–2006…. 66

Table E5 Differences Between Reading First and Comparison Schools for Kindergarten DIBELS Subtests, 2005–2006……... 67

Table E6 Differences Between Reading First and Comparison Schools for Grade 1 DIBELS Subtests, 2005–2006……. 68

Table E7 Differences Between Reading First and Comparison Schools for Grade 2 DIBELS Oral Reading Fluency, 2005–2006……...… 69

List of Figures

Figure 1 Grade 3 MSA Reading means by year and Reading First/Non-Reading First schools…..……... 14

Figure 2 Grade 3 MAP-R means by year and Reading First/Non-Reading First schools…… 15

Figure 3 Grade 2 CTBS/TerraNova reading means by year and Reading First/Non-Reading First schools…… 16

Figure 4 Kindergarten students 2005–2006 DIBELS Letter Naming Fluency end-of-year test performance by race/ethnicity in Reading First/ Non-Reading First schools' percentage of students at the Low-Risk (40+) level…… 17

Figure 5 Kindergarten students 2005–2006 DIBELS Letter Naming Fluency end-of-year test performance by race/ethnicity in Reading First/Non-Reading First schools' percentage of students at the Low-Risk (40+) level……. 17

Figure 6 Kindergarten students 2005–2006 DIBELS Nonsense Word Fluency end-of-year test performance by race/ethnicity in Reading First/Non-Reading First schools' percentage of students at the Low-Risk (25+) level…….. 18

Figure 7 Grade 1 students 2005–2006 DIBELS Phoneme Segmentation Fluency end-of-year test performance by race/ethnicity in Reading First/Non-Reading First schools' percentage of students at the Established (35+) level……..... 19

Figure 8 Grade 1 students 2005–2006 DIBELS Nonsense Word Fluency end-of-year test performance by race/ethnicity in Reading First/Non-Reading First schools' percentage of students at the Established (50+) level……..... 19

Figure 9 Grade 1 students 2005–2006 DIBELS Oral Reading Fluency end of year test performance by race/ethnicity in Reading First/Non-Reading First schools' percentage of students at the Low-Risk (40+) level……..... 19

Figure 10 Grade 2 students 2005–2006 DIBELS Oral Reading Fluency end of year test performance by race/ethnicity in Reading First/Non-Reading First schools' percentage of students at the Low-Risk (90+) level……..... 20

Figure 11 Effect sizes for cohorts 2003-2004 to 2005–2006 DIBELS on standardized tests..... 23

Executive Summary

The following report describes the implementation and conducts an evaluation of Reading First in the Montgomery County Public Schools (MCPS) from 2003 to 2006. Reading First is a grant-funded program designed to ensure all children read by the end of Grade 3. Reading First requires changes in curriculum, staffing, assessment procedures, and professional development. The goal of the report is to describe how well these changes were made and whether they are associated with student achievement.

Four schools in MCPS were selected to implement Reading First in Grades K–3 and four were selected as comparison schools. All schools have high proportions of students 1) receiving Free and Reduced-price Meal System (FARMS) assistance and 2) performing below grade level in reading. Highland, Rosemont, Summit Hall, and Wheaton Woods elementary schools were selected for Reading First. Glen Haven, Kemp Mill, Twinbrook, and Weller Road elementary schools were selected as comparison schools.

This evaluation addresses the following questions:
1) To what extent was the Reading First program implemented in schools as designed?
2) To what extent is the Reading First program perceived to be effective by a) teachers, b) Reading First coaches, c) principals, and d) parents/guardians?
3) How are students performing in reading in participating schools and how does this performance vary by student subgroups such as race/ethnicity and receipt of English for Speakers of Other Languages (ESOL)/FARMS/special education services?
4) How effective is Reading First in improving reading achievement at participating schools?

Methodology. Interviews with teachers and administrators, classroom observations, and Reading First training records were used to answer questions related to implementation. Staff interviews were used to assess perceived effectiveness. Standardized achievement tests, including the Maryland School Assessment (MSA), Comprehensive Test of Basic Skills/TerraNova (CTBS/TerraNova), Measures of Academic Progress-Reading (MAP-R), and Dynamic Indicators of Basic Early Literacy Skills (DIBELS) subtests were used to describe performance of students in various subgroups, and also to assess the effectiveness of Reading First in the four schools. The effectiveness analyses were conducted using a model that created a matched sample based on propensity scores and adjusted for demographic and service characteristic variables.

Findings from Implementation Evaluation

Evidence from the interviews suggests that most staff believed the instructional, assessment, and professional development goals of Reading First are being met. Some believed the program moved too quickly for non-English-speaking students and too slowly for some students performing at grade level. Staff also reported their Reading First curriculum lacks comprehension and writing emphasis, is difficult to schedule, doesn't accommodate ESOL instruction well, and inhibits creativity in some teachers. Although the majority of staff interviewed thought highly of the program and would like others to be aware of and

acknowledge the success they are having, many staff acknowledged that the program was not right for all students and teachers.
The classroom observations show that teachers are spending instructional time on the five basic reading skills (phonemic awareness, phonics, fluency, vocabulary, and comprehension) that is generally consistent with recommendations by Reading First guidelines. However, many teachers are including writing (drawing/painting for younger students) during reading instruction because they feel opportunities for writing are limited within the Reading First curriculum.

Findings from Effectiveness Evaluation

For standardized tests such as the CTBS/TerraNova and MSA, no consistent differences were found. When controlling for prior CTBS/TerraNova performance, Grade 3 students showed a significant negative effect of Reading First in 2003–2004 and a significant positive effect for Reading First in 2005–2006, although the effects were very small. Small negative effects of Reading First were found for the most recent MAP-R data available for Grade 3 students.

Moderate effects of Reading First were found for kindergarten and Grade 1 with the Phoneme Segmentation Fluency (PSF) subtest. Small effects of Reading First were found for kindergarten and Grade 1 Nonsense Word Fluency (NWF) subtests, and for Grade 1 and Grade 2 Oral Reading Fluency (ORF) subtests.

In summary, Reading First appears to have been implemented as intended, with a few minor exceptions, but does not appear to produce marked differences in students' achievement, as measured by the MSA or CTBS/TerraNova. Differences in DIBELS outcomes much more clearly show that Reading First students' scores exceed those of non-Reading First students, but this may be due to Reading First teachers and students' experience with DIBELS.

It is recommended that the investigation of DIBELS subtests in Reading First and non-Reading First schools be conducted again in 2006–2007, when non-Reading First schools have had another year of experience administering it. Reexamining effects of Reading First on standardized achievement in 2006–2007 tests would also be useful since evidence exists of increasing effects from 2003 to 2006. This is especially important to consider given the lag in effect found in other state Reading First evaluations.

Evaluation of Reading First in the Montgomery County Public Schools

2003–2006

Scot W. McNary, Ph.D.[1]

Introduction

The following report describes the implementation and conducts an evaluation of Reading First in MCPS from 2003 to 2006. Reading First is a grant-funded program designed to ensure all children read by the end of Grade 3. Reading First requires changes in curriculum, staffing, assessment procedures, and professional development. The goal of the report is to describe how well these changes were made and whether they are associated with student achievement.

Background

The Maryland State Department of Education (MSDE) was awarded a Reading First State Grant to allocate funds to eligible school systems beginning in 2003. MCPS was one among nine school districts in the state of Maryland that were eligible to compete for sub-grants. After receiving funding, the Maryland Reading First Initiative (MRFI) began in 2003–2004. The goal of MRFI is to ensure that all Maryland children are proficient in reading by the end of Grade 3. By so doing, reading achievement gaps will be reduced between children who live in poverty and those who do not. To accomplish this, MRFI requires changes in the following three areas (MSDE press release, September 5, 2003):

1. Instruction
    a. using a single reading program that adheres to scientifically based reading research (SBRR) principles,
    b. providing at least 90 minutes of uninterrupted reading instruction time each day, and
    c. offering differentiated instruction for students learning at different rates.

2. Assessment
    a. frequent assessment of all students to monitor progress using DIBELS, and
    b. screening to identify students in need of extra help using DIBELS and other tests as needed.

3. Professional Development
    a. providing a reading coach to assist teachers in their professional development as Reading First teachers within the school, and
    b. providing training opportunities outside the school for teachers, coaches, and principals at Reading First schools to facilitate the use of SBRR in instructional practice.

[1] Based in part on earlier work prepared by Nyambura Maina and Joy Mordica in 2004.

Scientifically Based Reading Research

A central premise of Reading First is that instructional materials and practices should be empirically based, that is, informed by SBRR. Instructional techniques and teacher training methods that have evidence of effectiveness in increasing student reading skill are considered SBRR. The National Reading Panel (2001) summary report identifies the five components of instruction (phonemic awareness, phonics, fluency, vocabulary, and comprehension) that have considerable empirical support for their efficacy. Reading curricula that address these components and teacher professional development efforts that disseminate these findings are also considered to be scientifically based. Instructional strategies, assessments, and professional development opportunities that are based on SBRR are designed to enhance teachers' reading instruction and thereby increase student reading achievement (MFRI, 2003).

MCPS Reading First

Core Reading Program. Under MFRI, candidate curricula are selected by a panel comprised by representatives from all 24 Maryland local education authorities (LEA; MFRI, 2003). Individual districts may select their own curricula from this list. MCPS has selected the Houghton-Mifflin curriculum. Curricula used for enrichment or intervention (for students performing below benchmark) are also approved by the panel, but may vary from school to school.

Assessment. The Reading First initiative requires that students' reading achievement be assessed for screening, diagnosis, progress monitoring, and outcome. Screening and progress monitoring are conducted using DIBELS. DIBELS is administered four times during the school year, with the initial assessment serving as a screening device. After screening, students may be provided activities within the curriculum that meet their current levels of reading achievement. Students who do not demonstrate satisfactory progress are selected for progress monitoring, which requires retesting over a short period of time, and, if found to be underperforming, referral for intervention instruction. Students' progress is monitored monthly through DIBELS until benchmark reading performance is achieved. If a student does not respond to supplemental instruction, diagnostic testing is administered and individualized instructional strategies are developed for the student. Yearly achievement assessments are conducted with standardized tests such as the Stanford Achievement Test Series, Tenth Edition (SAT 10) for Grades 1 and 2 (Reading First schools only), and the MSA for Grade 3 students.

Professional Development. All kindergarten through Grade 3 classroom teachers, special education, ESOL teachers participate in Reading First professional development activities, which consist of mandatory and elective trainings sponsored by MSDE, MCPS, Houghton-Mifflin (the curriculum publisher), and Montgomery County Council of the International Reading Association. Teachers learn about the pedagogical foundation for the five components of reading, instructional methods for the five components, differentiation within Reading First, and how to use DIBELS and other assessment data to aid instruction. Another part of teachers' professional development includes the resources of a specialist in Reading First at each school (a "coach") and a project specialist for each district. Coaches and specialists are considered experts available for consultation to teachers about lesson plans, instructional materials, or modeling of techniques. Principals are also required to obtain training in order to ensure implementation at their respective schools. Reading institutes are scheduled during the summer for training of new and continuing teachers. All aspects of training are intended to be informed by SBRR.

Evaluation Design

The Reading First program evaluation will be conducted using a multi-method design. Several data collection methods, including survey, classroom observation, interview, and school system archival data will be used to address the following evaluation questions.

1. To what extent was the Reading First program implemented in schools as designed?

2. To what extent is the Reading First program perceived to be effective by a) teachers, b) Reading First coaches, c) principals, and d) parents/guardians?

3. How are students performing in reading in participating schools and how does this performance vary by student subgroups, such as ethnicity, receipt of ESOL/FARMS/special education services?

4. How effective is Reading First in improving reading achievement at participating schools?

Evaluation questions 1 and 2, pertaining to the implementation and perceived effectiveness of the program, will be answered using data from staff interviews, classroom observations, and training attendance data.

Evaluation question 3, concerning performance of children in Reading First schools (overall and within subgroups) will be answered by an examination of the descriptive statistics on test performance (counts/percents for DIBELS subtests, means and standard deviations for standardized achievement tests).

Evaluation question 4, the effectiveness of Reading First, will be answered with inferential testing comparing Reading First students with non-Reading First students in three years of cohorts, from 2003–2004 to 2005–2006, using standardized achievement tests, MAP-R, and DIBELS subtests where available.

Methodology

Study Sample

Out of the 18 Title I schools in MCPS during 2003–2004 school year, eight were selected for study by an advisory committee producing the MRFI. Four were assigned to implement Reading First (Highland, Rosemont, Summit Hall, and Wheaton Woods) and four were assigned as comparison schools (Glen Haven, Kemp Mill, Twinbrook, and Weller Road).

All of the K–3 students in the four selected schools are included in the Reading First intervention since it is a schoolwide intervention. Students with valid DIBELS scores during the 2005–2006 school year and/or students who have test scores in springs of 2004, 2005, and 2006 constitute the sample for student achievement test analyses. A table containing the distribution of demographic and service provision characteristics among students included in the analysis can be found in Appendix A.

Implementation Evaluation Data Collection Activities

Staff interviews (N =12) were conducted at the teacher, coach, or principal's convenience in late spring 2006. All four principals, three of the Reading First coaches, and four teachers (one from each grade) were randomly selected for interview. One of the teachers interviewed was also randomly selected for observation (see below). An ESOL teacher was interviewed in place of the Reading First coach at one school (due to absence), resulting in 12 interviews. Interview protocols were developed specific to each role (i.e., principal, coach, teacher) and designed to elicit responses from the individual about his/her experiences with implementation and beliefs about effectiveness of the program (see Appendix B). Four of 12 staff interviewed had been exposed to only one year of the Reading First program (two principals, two teachers), whereas the other eight had been at their school for the entire duration of the program.

Parent interviews (N = 22) were conducted on four different mornings during late spring 2006 after parents dropped off children for school. Two parents self-administered the protocol questions, 4 responded as a group, and the remaining 16 consisted of one-on-one interviews. Thirteen of the interviews were conducted in Spanish.

Classroom observations (N = 16) were conducted using the Program Fidelity Checklist (Oregon Reading First Center, University of Oregon, 2006), a behavioral observation record used to provide feedback to teachers in Reading First schools. Observers were trained to use the checklist using videotaped classroom lessons (WGBH Educational Foundation, 2002) and post-rating discussion. Raters achieved satisfactory levels of agreement using the Program Fidelity Checklist following training and in the field. Additional data on the inter-rater agreement can be found in Appendix C. One teacher per grade per school was randomly selected for observation during May and June 2006. Observations were scheduled at the teacher's convenience to overlap with the reading instruction period. Training attendance data were collected by the Reading First project specialist for MCPS and cover the summer following the 2004–2005 and 2005–2006 school years.

Student Achievement Measures

Achievement outcomes were measured using standardized test scores. The TerraNova Comprehensive Test of Basic Skills (CTBS) is a nationally normed test, developed by CTB/McGraw-Hill. It assesses academic performance in reading, language, language mechanics, mathematics, and mathematics computation. It has been administered to Grade 2 students since at least 2001. In spring 2006, MCPS adopted the TerraNova Second Edition TN/2, a re-normed version of the original test for use with Grade 2 students. Spring 2003–2006 test results were used for this evaluation.

The MSA was first administered in the 2002–2003 school year to Grades 3 and 5. It has both reading and mathematics components and the tests are group administered in four 90-minute sessions over four days. Both selected response (multiple choice) and brief written response item types are included. Grade 3 test scores from 2003–2006 were used in the evaluation.

MAP-R is a computerized adaptive test developed by the Northwest Education Association to assess reading achievement. It was first administered to Grades 3–5 students in 2004–2005. The Rasch unIT (RIT) score is used for analysis. It identifies a student's current level of reading achievement. Spring 2005 and winter 2006 test results were used for this evaluation. The MAP-R was not administered in 2003, the first year of the Reading First implementation.

DIBELS is a test that was administered to Reading First schools only in 2005, and to both Reading First and non-Reading First schools in 2006. The test was combined with elements of the Montgomery County Public Schools Assessment Program Primary Reading (MCPSAP-PR) into a test suite called mClass-3d. It contains five tests from DIBELS and five from the MCPSAP-PR.

Data Analysis

Descriptive statistics were used to summarize the interview, classroom observation, and training attendance data to demonstrate a) the degree to which a scientifically based reading program has been designed and delivered, b) the types of intervention plans/strategies delivered at the Reading First schools, and c) the implementation status of the major activities and/or features of the program. Interviews underwent content analysis for evidence relevant to implementation in 2005–2006. Student test scores (MSA, TerraNova, DIBELS) were analyzed as a between groups analysis, comparing Reading First versus non-Reading First schools. Student scores served as the unit of analysis, and students were considered nested within schools.

Although Reading First and non-Reading First schools were matched on overall reading achievement and FARMS status, other differences may exist between these schools that confound comparisons between schools. One way of controlling for these differences is by matching students more closely using propensity scores (Dehejia & Wahba, 2002; Rosenbaum & Rubin, 1983). Propensity scores are composites of a set of specific predictors that predict group membership, in this case Reading First versus non-Reading First students. This should result in better estimates of the effect of the Reading First program.

Recent developments in the propensity score literature have suggested options for matching students explicitly on the propensity score as well as controlling for covariates (Abadie &

Imbens, 2006). Flexible matching algorithms that accommodate ties among propensity scores and/or other matching criteria are also recently available (Sekhon, 2006) and are used in this evaluation to help improve the comparison between essentially nonequivalent groups. The model used for this analysis creates matched samples between Reading First and non-Reading First groups, and also allows for control of additional covariates in the assessment of difference between groups. This model is referred to as the *propensity score matched and covariate-adjusted model*.

In addition to the propensity score matched model, three other models were used to assess the effect of the intervention. First, an *unadjusted model* was run, which is equivalent to an independent groups t-test using school type (Reading First versus non-Reading First) as the grouping variable and achievement as the outcome. This model was used to provide a reference point for comparing the other three models that adjust for demographic and service provision characteristics. Second, a *covariate-adjusted model* was used with the following covariates: gender, race, FARMS, ESOL, special education, and proportion of school days in attendance (except for 2005–2006, when attendance data was not available by the time this report was completed). This model was included because it contains a typical set of covariates found in many educational inferential analyses and the subgroups are regarded as important by NCLB. Third, a model containing all the covariates plus a propensity score adjustment was included to help determine the utility of propensity score adjustment relative to unadjusted and standard approaches to educational analyses. The propensity score was used to create a categorical variable with five levels based on quintiles of the propensity score. When used in this way, the propensity score is considered a stratification variable and can account for imbalance of the covariates between groups (Rosenbaum & Rubin, 1983). Covariates are included in the model along with the propensity score in order to control for any remaining imbalance between groups not accounted for by including the categorized propensity score. This relatively simple approach to including a propensity score can be used as a reference point from which to judge the utility of the more complicated and computationally intensive method proposed by Sekhon (2006). This model was referred to as the *covariate-adjusted and stratified on propensity score model*. All models were run using multiple regression analyses. Results for the *propensity score matched and covariate-adjusted model* are reported in the results section. Results for the other three types of analyses are reported in Appendix E for purposes of comparison.

Differences between Reading First and non-Reading First schools were expressed as effect sizes (Cohen's *d*; Cohen, 1988). For the propensity score matched analysis, this is an effect size estimate based on a one-sample t-test for matched pairs ($ES = t/\sqrt{df}$).

Statistical analyses were conducted with SPSS 11.5 (SPSS, 2002) and R 2.3.1 (R Development Core Team, 2006).

Data Availability

Data were not available for all students at all time points. The 2005–2006 school year is the first in which Reading First and non-Reading First schools could be compared on DIBELS subtests. Students in kindergarten and Grade 1 may be compared between schools. As part of the National Impact Study, Reading First students in Grades 1–2 took the Stanford Achievement Test 10 (SAT 10), but non-Reading First schools did not, and so they are not helpful in making

comparisons between schools. In 2005–2006 Grade 2 students took the TN/2. Table 2 summarizes the test score availability.

Table 1
Achievement Test Data Availability

Source	2003–2004	2004–2005	2005–2006
MSA (Grade 3)	available	available	available
MAP-R (Grade 3)	n.a.	available[4]	available[3]
CTBS (Grade 2)	available	available[1]	available
DIBELS (Grades K–2)	n.a.	available[2]	available

Note: [1]Wheaton Woods ES was the only Reading First school with CTBS/TerraNova and MAP-R data in 2004–2005. [2]No comparison schools took DIBELS and no Reading First schools took MCPSAP-PR tests. [3]Highland ES did not administer the MAP-R in 2005–2006. [4]Winter 2006 scores.

Results

Evaluation questions 1 and 2 concerned the effectiveness of the implementation and perceived effectiveness of Reading First from the point of view of staff and parents. Interview data comprise most of the data for these two questions, although training attendance data and classroom observation data are also informative and are discussed below.

Staff Interviews

Professional Development Opportunities. Staff described four different types of training and support. These include mandatory training for teachers such as the Maryland Institute for Better Reading (MIBR), training on how to use DIBELS, and training sponsored by the curriculum publisher (Houghton-Mifflin) for new teachers. Optional training on intervention methods, specific reading skills (e.g., phonemic awareness, phonics, vocabulary), and assessment methods were also mentioned. Training for Reading First coaches was described as more frequent and consistent with their job as local trainers for teachers at their school. Principals and coaches attended national conferences and several noted the conferences were inspiring and helpful in understanding the national context for Reading First. Finally, teachers met regularly with coaches (some weekly, others less frequently) during the 2005–2006 school year, and principals met with other Reading First principals at bimonthly meetings.

Opinions varied on the training. A new teacher believed she didn't have enough training and hoped to obtain "deeper training on the program." One staff member suggested some of the publisher-sponsored training was "one cut above advertising." Two of the three coaches interviewed expressed a desire to attend Language Essentials for Teachers of Reading and Spelling (LETRS) training. Two of the five teachers explicitly mentioned that their coaches were helpful with problems. Although 2 of the 12 staff said they didn't believe any additional training was necessary, two wanted more ESOL training, two others wanted more differentiation training, and two others mentioned learning more about basic research on the program and/or how children learn to read. Other suggestions included training in interventions, use of DIBELS,

comprehension skills, and interactions/observations at other Reading First schools. One staff suggested a return to the MCPS reading curriculum.

When asked how their professional work had changed as a result of Reading First, two principals said they had not changed at all but one said he was more aware of Reading First components when observing teachers. Coaches mentioned being better able to recognize teachers' strengths and weaknesses, the importance of modeling for teachers, showing teachers how to use data, and how to provide observation feedback with both positive and negative elements. Teachers mentioned increasing use of fluency and phonemic awareness instruction.

Three of the five teachers explicitly mentioned receiving feedback about their teaching from coaches. One mentioned being observed once per month, another mentioned being observed three times by the coach (once by a state staff member and once by the assistant principal). A third specified no frequency of observations, but described meeting about student data with the coach once per month.

Reading First Curriculum. All staff teachers said their schools used the Houghton-Mifflin (HM) curriculum. One teacher mentioned using Read Well for intervention, another mentioned the Voyager curriculum. This same teacher also said that half of the second graders at his school were receiving some sort of intervention. A second teacher mentioned using Junior Great Books during intervention for students reading above the DIBELS benchmark (i.e., who were doing well). Principals and coaches had broader knowledge of the intervention curricula. For students who were not performing well, curricula used included Soar to Success, Quick Read, Read Naturally, Early Reading Intervention, Voyager, and Read Well. For students who were performing well, curricula used included William and Mary, Junior Great Books, HM anthology, and the next grade level of HM series. At some schools, students who were performing adequately had an additional period of the HM curriculum. Coaches provided similar information but one coach added that both the HM curriculum and DIBELS were also being used by Grades 4 and 5 at her school.

Assessment/Progress Monitoring. All 12 staff reported that DIBELS/mClass was used for progress monitoring (used to determine whether students meet benchmarks) and all but two mentioned the HM curriculum end-of-theme tests. Two staff members mentioned standardized tests, MSA, CTBS/TN/2, and the SAT 10. One principal also uses the MCPSAP-PR. Another principal mentioned providing data in a public place for all to see how the students are performing in the aggregate. A few staff members mentioned that it was very helpful to have regular assessments and reporting with DIBELS data.

Differentiation. Students with special needs (ESOL/special education/struggling/Gifted and Talented) are accommodated by the HM curriculum, according to the six staff members who responded to this question. Three mentioned that students performing above grade level are allowed to work in the HM curriculum at the next grade level up.

The 90-Minute Reading Block. When asked about whether the 90-minute reading block was protected in Reading First schools, the 10 of 12 staff members who responded to this question agreed that it was. Coaches and staff noted some exceptions, including 1) accommodating ESOL testing, 2) breaking the 90-minute period into halves and scheduling the reading intervention

period between them for kindergarten and Grade 1 classes to accommodate non-Reading curricula, and 3) behavior problems. One principal reported that the reading block prevented adequate instructional time for science, writing, social studies, and mathematics.

Positive Aspects of Reading First. Staff were asked which aspects of the program were working well for students and teachers. All 12 interviewees provided responses to these questions. Principals and coaches were also asked which aspects were working well for coaches, and five staff responded to this question. Elements of the program that were working well included some of the five Reading First reading components, specifically, phonemic awareness/phonics/fluency (4/12 staff). Two believed the reading material was interesting to students, and two believed the structure and predictability of the program was beneficial. More global benefits were reported including that "kids are learning to read," are more "confident," and "reading, writing…speaking and thinking has improved." Two coaches believed that the "explicit and direct instruction" was helpful, as was DIBELS, in matching students with instructional needs.

Benefits for teachers included decreased time and need for planning due to the structure offered by the program (5/12 staff). One principal said that the common curriculum and assessment methods has allowed her staff to communicate "on the same page" and that this has "united (her) staff" and "improved staff climate." Still another principal noted that the structure associated with the scheduling of the program allowed teachers to have the same free time available for meetings. However, one teacher said that although the structure made it easy, it was boring. Teachers are now more skilled using data and are able to make use of data for planning (2/12 staff). Still other benefits included the availability of small groups for improved informal assessments, the availability of materials to send home, and teachers learn new teaching strategies from intervention training (e.g., Read Well, Read Naturally).

Aspects of Reading First in Need of Improvement. When asked what was not working well, all 12 staff had a response. Responses can be grouped into four general areas: limited scope of the program, problems with assessment, restrictions on the types of texts used, and organizational concerns. Four staff had specific suggestions for the curriculum related to reading components: two suggested developing comprehension instruction, one for kindergarten, another suggested a reduction in attention devoted to phonics, and a fourth suggested vocabulary instruction for students at or below grade benchmark. Two of 12 staff members suggested that the curriculum needed a writing component. Some staff members reported problems with assessments: 1) DIBELS lacks a comprehension assessment, 2) there are too many HM assessments and they take up too much time, 3) teachers question the validity and appropriateness of some of the assessments, 4) the assessments are not aligned with MSA, and 5) DIBELS lacks a leveled reading assessment.

Two staff members thought the choices of text were problematic: they were either too few, "goofy," or inappropriate for the student's age. Three staff members felt that the integration with ESOL needed more development, and one staff recommended improving the curriculum to better suit advanced students. A principal mentioned that all four Reading First schools manage ESOL with Reading First differently.

Three staff members wished for more communication with the county and central office or with other schools, and one staff member was concerned that too many resources were being allocated to staff not directly involved with students or assessment. Another staff member suggested that

the Reading First curriculum was not well planned and that training teachers to teach Reading First drove good teachers away from the school and did not help the ones who remained in their future attempts to find jobs at non-Reading First schools. Contrary to this perspective, another staff member believed that Reading First instruction helped teachers at her school become stronger teachers who could work wherever they might choose to work. The one concern mentioned by this staff member was that Reading First was a good basis, but students needed to have a curriculum to move up to: "approximately 65 percent of students are now ready to make the transition to the MCPS curriculum. They could definitely make that move, as a result of receiving Reading First prerequisite…a lot of our children are reading and it's 'where do you go from here?'"

One staff member's assessment was particularly bleak. He felt that little about Reading First was working, although he noted that "there are students who are lost and this type of instruction gives them something firm to hold on to." Still another staff member believed that the program was a hindrance to students who moved out of the area or to a middle school that did not have a Reading First program, because they would be confused by the standard curriculum and likewise for students new to a Reading First school. He also believed that teachers who only knew how to teach Reading First would have a difficult time finding jobs at non-Reading First schools.

Staff responses suggested three types of additional support that would be useful in implementing the program. These included more personnel, more training (how to do guided reading groups, what makes the program research-based), and more support from the county.

Other Comments. Before concluding, all interviewees were asked if there was anything else they wanted to mention about the program. Responses seemed to fall into two categories: very positive ones mentioning other outcomes of interest, or negative ones pertaining to doubts the staff member or others might have about the program. For example, one coach offered these comments: "Teachers really like it and see how it is working for the kids; I have grown professionally with this program." Another said, after changing to the Reading First program:

> "…we have everything we need, we don't have to share books, we don't have to look for books that could be used in place of books we don't have. You can focus on instruction, which makes it so much easier not to have to run around and look for things. The stories are very interesting. The kids love it. I really enjoy using the program. For the most part, it meets the need of our students. It could use some fine-tuning, but it has each of the components we need."

Other staff acknowledged that Reading First was not for everyone and pointed out that taking creativity away from teachers was sometimes good when consistency in instruction was desired. One teacher expressed a wish for more flexibility. When dealing with the lack of flexibility of the program, a coach suggested that it made a difference to teachers how she approached them. She stated that "You can approach it as 'here, you're going to be a robot and teach this' or it's 'here's a guide, some tools that are going to help you, we know you are the professional.'"

Two final questions were asked about the quality of the implementation and whether the staff member would recommend Reading First. The responses to these two questions are tabulated below (Table 3). More than half of staff members believed that the program was implemented with excellent quality and would recommend Reading First to colleagues. Two conditions mentioned on recommendations were that strong leadership was important and that Reading First is especially recommended for new teachers.

Table 2
Staff Evaluations of Quality of Reading First Implementation and Whether to Recommend Reading First to a Colleague

Quality Of Implementation		
Poor	0	
Fair	1	(8%)
Very Good	7	(58%)
Excellent	4	(33%)
Total	12	
Would You Recommend Reading First?		
Do Not Recommend	2	(17%)
Recommend With Reservations	1	(8%)
Recommend	2	(17%)
Strongly Recommend	7	(58%)
Total	12	

Summary

Evidence from the interviews suggests that most staff believe the instructional goals of Reading First are being met. All staff agreed that a single reading program approved by MSDE is being used as the core reading curriculum, and that at least 90 minutes of protected instruction time is devoted to reading. Staff agreed less on how well students of different reading abilities were being accommodated by Reading First. Some believed that the program moved too quickly for non-English-speaking students and too slowly for some students performing at grade level. DIBELS is used to classify which students are in need of more help and for progress monitoring in all schools. Coaches are available at all schools and several teachers noted that they were helpful for guidance. Principals, coaches, and teachers all acknowledged the availability and receipt of training, although some would like more.

Several additional conclusions can be drawn from a consideration of the interview responses. First, although there are many benefits to having a consistent and structured curriculum across schools and grades, it lacks important instructional elements (e.g., comprehension and writing emphasis), is difficult to schedule, doesn't accommodate ESOL instruction well, and inhibits creativity in some teachers. Even one principal who thinks highly of the Reading First/Houghton-Mifflin curriculum believes that it works well as a prerequisite to a more advanced curriculum like the MCPS curriculum. Second, the majority of staff interviewed thought highly of the program and would like others to be aware of and acknowledge their successes. On the other hand, many staff acknowledged that the program was not right for all students and teachers. Finally, it seems clear that staff members believe each teacher and administrator in the school must understand the value of Reading First before it can succeed.

Parent Interviews

Of the 22 parents who responded to the interview protocols, fewer than half (41%, 9 individuals) knew about Reading First. Of those who responded, five were aware of an emphasis on assessment, many believed their children were reading well, and three expressed some concern

that the program was too rigid (didn't accommodate above- or below-grade-level children and too restrictive in choices of books). A few parents expressed a desire for more communication from their schools about Reading First.

Training

The attendance data for teachers in Reading First schools for mandatory and elective training conducted in 2004–2005 and 2005–2006 can be found in Appendix C, Tables C1–C4. The proportion of K–3 teachers attending (11%–15%) was higher than for other grades or specialties, with the exception of ESOL teachers (16%). This is consistent with the overall goal of Reading First to have all children reading by Grade 3.

The majority of teachers had obtained the mandatory training required of them in 2004–2005 or 2005–2006 (MIBR, DIBELS) and many had attended trainings for topics specific to their teaching needs (Read Well, Read Naturally, Phonemic Awareness, Phonics, Vocabulary, Fluency) in 2004–2005. When trainings were changed to optional in 2005–2006, attendance decreased. Attendance data for trainings prior to fall 2004 were not available.

Classroom Observations: Consistency with MFRI Recommendations

The Maryland Reading First Initiative reprints a table the authors attribute to Simmons and Kame'enui (2003) that suggests that each of the five instructional goals for Reading First (Phonological Awareness, Phonics, Fluency, Vocabulary, Comprehension) should occur at different phases and with different emphases, depending on grade. The Institute for Development of Educational Achievement (IDEA, 2006) has a similar suggested sequence of instruction, although considerably more detailed. Observed class activities can be compared with these recommendations as a measure of implementation.

A total of 207 activities over 16 classrooms (four per school) were observed. Observation sessions took place during scheduled reading instruction times and lasted an average of 95.8 minutes (*SD* = 23.9; Median = Minimum = 60 Maximum = 157), and contained an average of 16.2 students (*SD* = 2.5). At least 56 percent of observation sessions exceeded 90 minutes, the length of the protected reading block. Observation sessions of shorter duration may not have completely overlapped the entire reading block due to scheduling arrangements. Activities were noted if they were the main focus or an additional focus during a particular instructional period. For example, if a teacher defined words for students during a fluency reading exercise, the main activity would be considered fluency, but vocabulary would also be counted as an additional activity. The number of activities and number of minutes spent in each type of Reading First activity by grade is shown in Table 4. The bottom rows labeled "As Main or Additional Activity" represent the sum of minutes spent in any activity (e.g., kindergarten students spent 18 minutes in fluency activities as a main activity, 9 as an additional activity, and 18+9=27 as main or additional activities).

Table 3
Mean Minutes Spent in Reading First Recommended Instructional Activities, by Grade

Grade	Total Number of Activities	Phonemic Awareness	Phonics	Fluency	Vocabulary	Comprehension	Other
Mean Minutes Spent in Activity							
As Main Activity							
K	82	9	15	18	5	14	23
1	43	6	12	14	9	10	8
2	47	0	7	11	27	17	49
3	53	0	22	11	15	15	16
As Additional Activity							
K	82	12	10	9	11	9	18
1	43	0	9	11	12	19	9
2	47	8	12	13	10	15	15
3	53	0	11	8	12	9	13

Students in kindergarten classes experienced more activities in general, and spent more time on average in Phonemic Awareness and Fluency activities in particular, than any other grade. Grade 3 students spent more time on average in Phonics activities than other grades. Grade 2 students spent more time on average in Vocabulary and Comprehension activities. Grade 1 students spent an average of 21–30 minutes on all Reading First activities (except Phonemic Awareness). Both kindergarten and Grade 2 students spent an average of more than 40 minutes in "other" activities, not associated with the five recommended reading activities during the reading instruction session. However, raters noted that almost all of the activities classified as "other" involved writing letters, sentences, or paragraphs, depending on grade level, or drawing and painting.

According to Simmons and Kame'enui (2003) Phonemic Awareness instruction should continue through the third quarter of Grade 1. This suggests that by the end of the school year (the fourth quarter) only kindergarten students should be receiving Phonemic Awareness instruction. This prediction is consistent with the classroom observation data since kindergarten students received the largest average number of minutes of Phonemic Awareness instruction and Grade 3 students received no Phonemic Awareness instruction, Grade 2 students received an average of 8 minutes (11% of mean instruction time) of Phonemic Awareness instruction as an additional activity (not a main activity), and Grade 1 students received an average of 6 minutes (10% of mean instruction time) of Phonemic Awareness instruction as a main activity. Comparisons of the other reading instruction activities with recommended emphasis are not possible since the MFRI do not specify a recommended number of minutes of instruction for specific kinds of activities.

Standardized tests

Evaluation question 3 is concerned with disaggregated performance of students on standardized tests by student subgroups, such as race/ethnicity and receipt of ESOL/FARMS/special education services. A descriptive comparison of test scores across year-cohorts and within cohorts provides a general idea of patterns within and across cohorts. For the purposes of this evaluation, cohorts refer to different years of Reading First implementation (i.e., 2003–2004, 2004–2005, 2005–2006).

Grade 3 MSA Reading. Comparing across years, cohorts of Grade 3 students' test scores increased sharply from 2003–2004 to 2004–2005, and more gradually from 2004–2005 to 2005–2006. This pattern varied somewhat within subgroups. For example, differences between cohorts of students appeared larger between 2004–2005 and 2005–2006. This included Asian American students and Highland students. Twinbrook, Rosemont, and Weller Road students decreased in average test scores by 2005–2006. Between groups (but within cohorts) several patterns were observed: female students tended to have higher average scores than male students, White and Asian American students had higher scores than African American and Hispanic students, students not receiving ESOL services had higher scores than students receiving ESOL services, and students not receiving FARMS services had higher scores than students currently receiving FARMS services. Reading First school students had lower scores than non-Reading First students, although this difference had decreased by 2005–2006 (Appendix D, Tables D1–D2). Figure 1 displays MSA Reading means by school type and cohort.

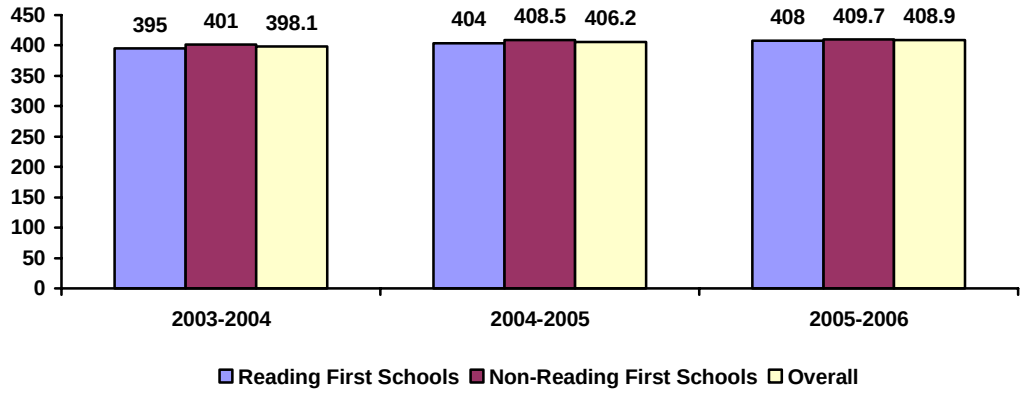

Figure 1. Grade 3 MSA reading means by year and Reading First/non-Reading First schools.

Grade 3 MAP-R. Comparing across years, cohorts of Grade 3 students' test scores dropped from 2004–2005 to 2005–2006. MAP-R scores were not available prior to fall 2004. This pattern varied somewhat within subgroups. For example, some differences between cohorts of students were very small between 2004–2005 and 2005–2006. This included Asian American students and students from Glenhaven and Weller Road elementary schools. Wheaton Woods was the only Reading First school to administer the MAP-R to Grade 3 students in both years. Several patterns were observed between groups (within cohorts): female students tended to have higher average scores than male students, White and Asian American students had higher scores than African American and Hispanic students, students not receiving ESOL services had higher scores than students receiving ESOL services, and students not receiving FARMS services had higher scores than students currently receiving FARMS services. Reading First school students had lower scores than non-Reading First students, although this difference is based on only one Reading First school in 2004–2005 and three Reading First schools in 2005–2006 (Appendix D, Tables D3–D4). Figure 2 displays MAP-R means by school type and cohort.

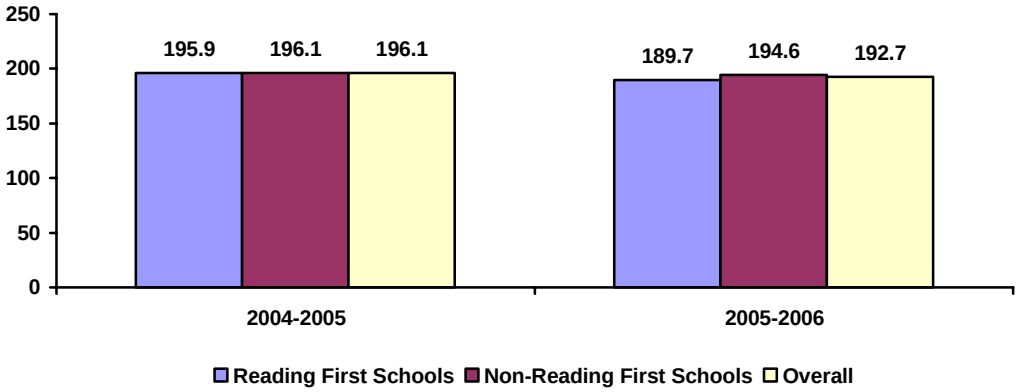

Figure 2. Grade 3 MAP-R means by year and Reading First/non-Reading First schools.

Grade 2 CTBS/TerraNova Reading. Comparing across years, cohorts of Grade 2 students test scores increased sharply from 2003–2004 to 2004–2005, and then decreased somewhat from 2004–2005 to 2005–2006. This pattern varied somewhat within subgroups. For example, some test scores remained the same or increased slightly over year-cohorts. White students' means showed little difference. Students currently receiving ESOL services and students from Glenhaven and Wheaton Woods elementary schools subgroups showed evidence of increasing means over cohorts. Reading First schools overall demonstrated increases in mean reading scores over cohorts, but these means are based on only one school in 2004–2005 (Wheaton Woods) because three of the schools did not administer the CTBS/TerraNova to Grade 2 students in 2004–2005. In 2005–2006 MCPS changed tests from the CTBS/TerraNova to the TN/2, so differences in performance may be attributable to differences in tests, making comparisons across cohorts difficult. Between groups (within cohorts) several patterns were observed: female students tended to have higher average scores than male students, White students had higher scores than Asian American, African American, and Hispanic students, students not receiving ESOL services had higher scores than students receiving ESOL services, and students not receiving FARMS services had higher scores than students currently receiving FARMS services. Reading First school students had lower scores than non-Reading First students, although this difference was at its narrowest in 2004–2005, when only Wheaton Woods provided data for the Reading First schools (Appendix D, Tables D5–D6). Figure 3 displays CTBS/TerraNova and TN/2 means by school type and year.

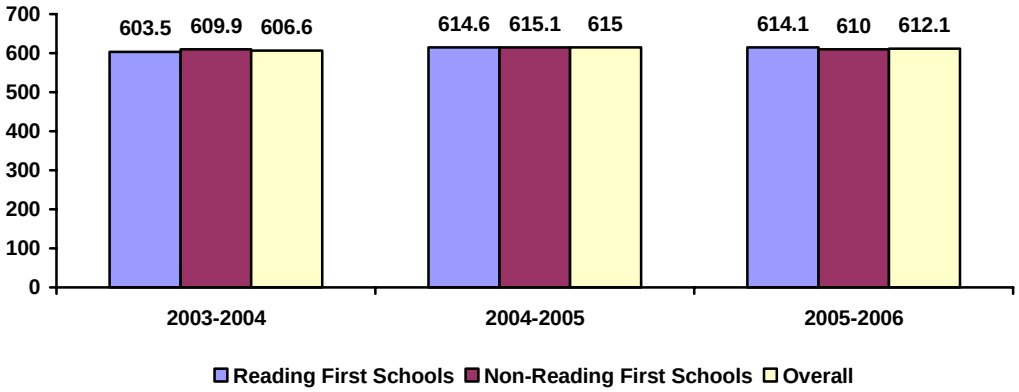

Figure 3. Grade 2 CTBS/TerraNova reading means by year and Reading First/non-Reading First schools.

DIBELS subtests

Non-Reading First schools did not administer DIBELS until 2005–2006 (as part of mClass3D), so 2004–2005 DIBELS results are available for students within Reading First schools only. Performance comparisons on DIBELS subtests are made by comparing the proportion of students in each subgroup whose performance at the end of the year is classified as either "Low Risk" or "Established," the highest level of performance for each test. The ORF subtest is the only subtest administered to Grade 2–3 students.

Kindergarten DIBELS: Letter Naming Fluency, Phoneme Segmentation Fluency, and Nonsense Word Fluency. Several patterns were observed: more female students were classified as "Low Risk"/"Established" than male students, more Asian American students were classified as "Low Risk"/"Established" than White, African American, and Hispanic students, and more students not receiving FARMS services were classified "Low Risk"/"Established" than students previously or currently receiving FARMS services (Appendix D, Tables D7–D14). Comparing Reading First and non-Reading First schools revealed somewhat more non-Reading First students classified as "Low Risk" on the LNF subtest compared with Reading First students, and this was apparent across subgroups. The proportion of students at "Low Risk" by school type and race/ethnicity is shown in Figure 4.

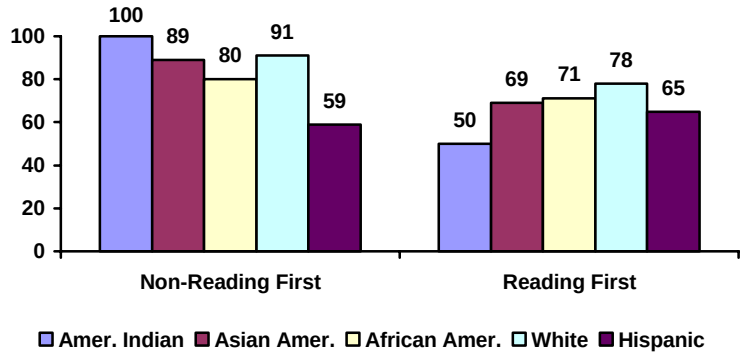

Figure 4. Kindergarten students 2005–2006 DIBELS LNF end-of-year test performance by race/ethnicity in Reading First/non-Reading First schools percentage of students at the Low Risk (40+) level.

On the PSF subtest more Reading First students were classified as "Established" than non-Reading First students, and these differences overwhelmed any differences in demographics or service provision (Appendix D, Tables D11–D12). The proportions of students classified as "low risk" by race/ethnicity and school type is shown in Figure 5.

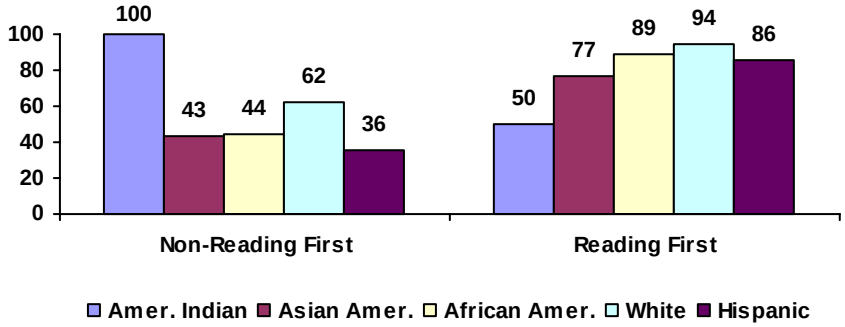

Figure 5. Kindergarten students 2005–2006 DIBELS LNF end-of-year test performance by race/ethnicity in Reading First/non-Reading First schools percentage of students at the Low Risk (40+) level.

More Reading First students were classified as "Low Risk" on the NWF subtest than non-Reading First students, although no differences in proportions of students classified as "Low Risk" were found between Asian American students or students who had previously received ESOL or FARMS services in the two school types (Appendix D, Tables D13–D14). Figure 6 shows the proportion of students at "low risk" by race/ethnicity and school type.

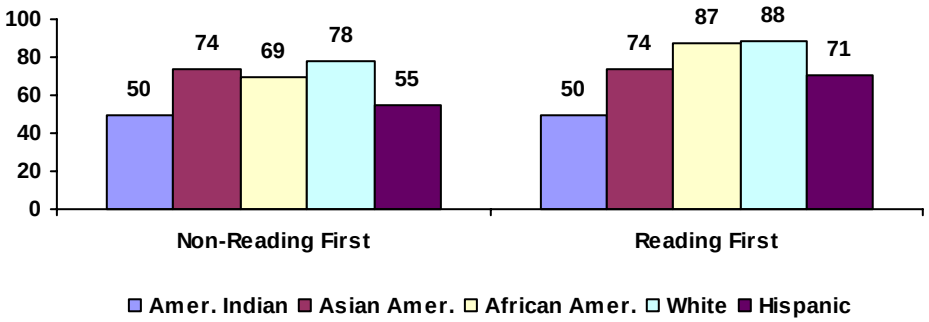

Figure 6. Kindergarten students 2005–2006 DIBELS NWF end-of-year test performance by race/ethnicity in Reading First/non-Reading First schools percentage of students at the Low Risk (25+) level

Grade 1 DIBELS: PSF, NWF, and ORF. In 2004–2005 few patterns were observed since the majority of students were classified as "Low Risk"/"Established," with four exceptions: more than half of 1) Hispanic students, 2) students currently receiving ESOL services, 3) students currently receiving special education services, and 4) students from Highland were classified as at least "Some Risk" ("At Risk" or "Some Risk") on the ORF subtest (Appendix D, Tables D15–D16).

In 2005–2006 more Reading First students were classified as "Established" on the PSF subtest compared with non-Reading First students, and this was evident across all subgroups. On the NWF subtest more Reading First students were classified as "Established" than non-Reading First students, and these differences overwhelmed any differences in demographics or service provision, just as with kindergarten students. Fewer Reading First students were classified as "At Risk" or "Some Risk" on the ORF subtest than non-Reading First students, although equivalent performance was found between Reading First and non-Reading First students who had never received ESOL services (Appendix D, Tables D17–D22). Figures 7–9 show the proportion of students who performed at the Established level or who were at "Low Risk" for PSF, NWF, and ORF by race/ethnicity and school type.

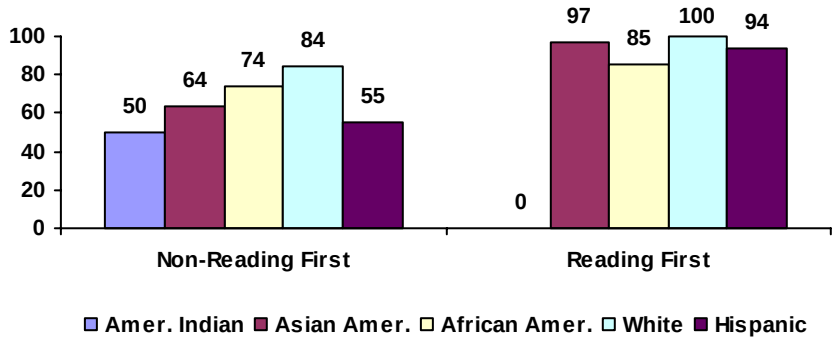

Figure 7. Grade 1 students 2005–2006 DIBELS PSF end-of-year test performance by race/ethnicity in Reading First/non-Reading First schools percentage of students at the Established (35+) level.

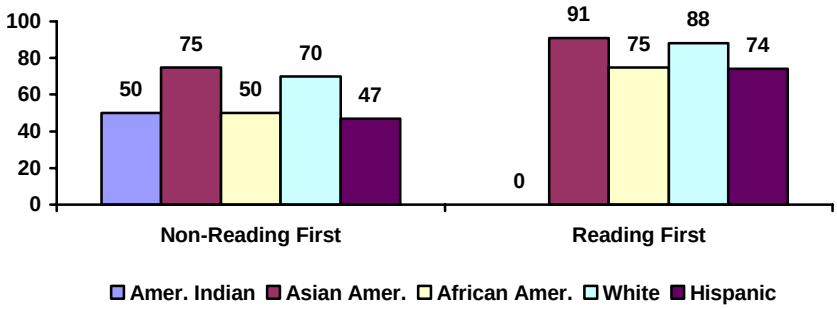

Figure 8. Grade 1 Students 2005–2006 DIBELS NWF end-of-year test performance by race/ethnicity in Reading First/non-Reading First schools percentage of students at the Established (50+) level.

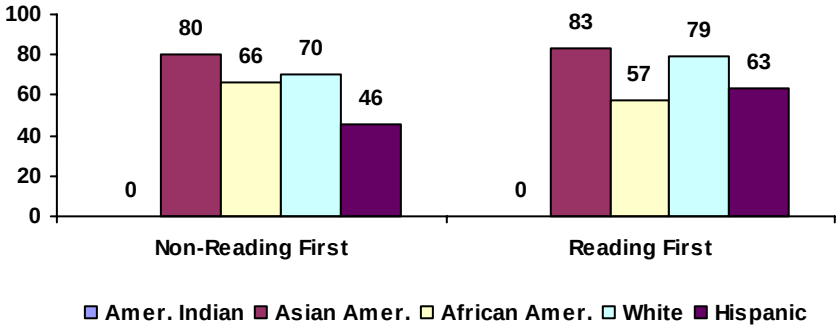

Figure 9. Grade 1 students 2005–2006 DIBELS ORF end-of-year test performance by race/ethnicity in Reading First/non-Reading First schools percentage of students at the Low Risk (40+) level.

Grade 2 DIBELS: ORF. In 2004–2005 more Asian American students were classified as "Low Risk" than White, African American, or Hispanic students. Fewer students currently receiving ESOL or FARMS services were classified as "Low Risk" than students not receiving them. Relatively low proportions of students at Summit Hall and Highland were classified as "Low Risk" compared with Wheaton Woods and Rosemont students (Appendix D, Tables D23–D24).

In 2005–2006 more Reading First students were classified as "Low Risk" on the ORF subtest compared with non-Reading First students, and this was evident across all subgroups, except for female students, students currently receiving ESOL services, and students currently receiving special education services, where the proportions appear equivalent between school types. More Highland students were classified as at least "Some Risk" than any other schools, including non-Reading First schools (Appendix D, Tables D25–D26). Figure 10 displays the proportion of students at "Low Risk" on the ORF subtest by race/ethnicity and school type.

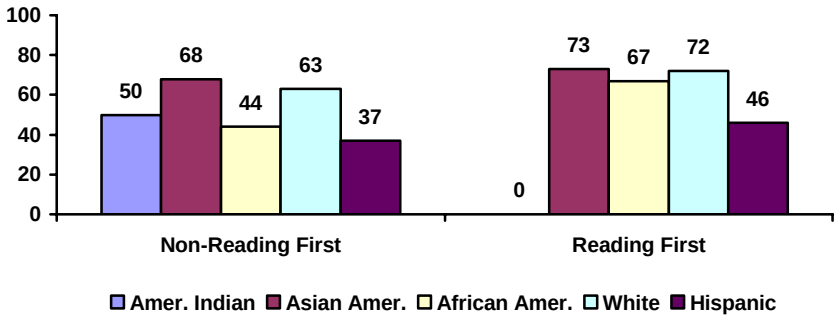

Figure 10. Grade 2 students 2005–2006 DIBELS ORF end-of-year test performance by race/ethnicity in Reading First/non-Reading First schools percentage of students at the Low Risk (90+) level.

Grade 3 DIBELS: ORF. In 2005–2006 more Asian American students were classified as "Low Risk" than White, African American, or Hispanic students. Fewer students currently receiving ESOL or FARMS services were classified as "Low Risk" than students not receiving them. Relatively low proportions of students at Highland were classified as "Low Risk" compared with other Reading First schools (Appendix D, Tables D27–D28). No non-Reading First school Grade 3 students took the DIBELS ORF subtest.

Inferential Tests: Comparisons of Reading First and non-Reading First Schools

Evaluation question 4 pertains to the effectiveness of Reading First. Effectiveness is evaluated as relative benefit of the program over a comparison group that was not exposed to the program. All analyses used a propensity score matched and covariate adjusted analysis by first creating matched samples based on a propensity score and then calculated a mean difference between Reading First and non-Reading First schools, adjusted for the covariates. Effect size estimates (Cohen's *d*) are calculated for each analysis.

Grade 3 MSA Reading. The matching algorithm produced well-matched samples for all three cohorts. After matching, the Kolmogorov-Smirnov (*KS*) test statistics for 2003–2004,

2004–2005, and 2005–2006 were 0.04, 0.02, and 0.004, with bootstrapped *p*-values of 0.99, 0.78, and 0.80, respectively. The low values of the *KS* test statistics obtained and high bootstrapped *p*-values obtained after matching suggest that the differences in covariates between matched Reading First and non-Reading First samples was considerably reduced by the matching algorithm and resulted in well-balanced groups for all three cohorts, although groups were initially reasonably well matched for 2005–2006. No differences in MSA reading scores were observed between Reading First and non-Reading First schools in any cohort (Table 5).

Grade 3 MSA Reading controlling for Grade2 CTBS/TerraNova performance. Using the previous year's CTBS scores allows comparisons to be made controlling for prior ability and also provides another source of information on which to match students in the propensity score matching algorithm. However, the reduction in sample size by including only those students who have data in two different years is substantial. For 2003–2004, the sample size of 691 is reduced to 623, a reduction to 90 percent of the original analysis sample. For 2004–2005, the reduction is from 718 to 584, or 81 percent of the original 2004–2005 sample. For 2005–2006, the reduction is the most dramatic, from 661 to 356, or 54 percent of the original sample. Some of the loss is attributable to students who have no prior ability testing for reasons of transfer or other individual causes. However, many losses are due to the policy of some Reading First schools to not administer the CTBS in 2004–2005. Consequently, inferences based on these analyses must be made with caution since they represent a subset of the original analysis samples.

Similar to the unadjusted analyses of the Grade 3 MSA, the matching algorithm produced well-matched samples for all three cohorts. *KS* statistics for 2003–2004, 2004–2005, and 2005–2006 were 0.07, 0.09, and 0.10, with bootstrapped *p*-values of 0.26, 0.996, and 0.77, respectively. This suggests that the matching algorithm resulted in well-balanced groups for all three years, although groups were well matched with respect to covariates prior to matching for 2005–2006.

In 2003–2004 a negative effect of the intervention was found with Reading First schools having significantly smaller average MSA reading scores, controlling for prior performance (Table 5). Controlling for prior performance allows the interpretation of the dependent variable to become one of relative gain. From this perspective, gain in reading achievement among non-Reading First schools exceeded the gain in reading achievement among Reading First students. However, the effect for the difference is small (0.13). In 2004–2005 no difference was found. In 2005–2006 a positive effect of Reading First was found with a small effect size (0.12). Although these results appear to favor Reading First in 2005–2006, it should be restated that these analyses are based on a single Reading First school—Wheaton Woods. Wheaton Woods was the only school that administered the CTBS/TerraNova in 2004–2005, which was the estimate of prior ability in these analyses. It is unknown what the effect of intervention would have been had the other three Reading First schools been included in this analysis.

Grade 3 MAP-R RIT Scores. The MAP-R was not administered until fall 2004, so only 2004–2005 and 2005–2006 RIT scores are available for analysis. In 2004–2005 only Wheaton Woods students provided MAP-R data, so comparisons are made with only one Reading First school. No differences between Reading First and non-Reading First schools were found that year. In 2005–2006 Highland students provided no MAP-R data, so comparisons are based on three Reading First schools. The matching algorithm produced well-matched samples for MAP-R RIT score analyses. *KS* statistics for 2004–2005 and 2005–2006 were 0.08 and 0.11, with bootstrapped *p*-values of 0.72 and 0.97 respectively, suggesting matching reduced covariate bias

between samples for both years. The 2005–2006 samples were initially well-balanced before matching was undertaken. No difference was found between Reading First and non-Reading First students in 2004–2005. A significant difference favoring non-Reading First schools was found in 2005–2006, but the effect size is small (-0.12; Table 5).

Grade 2 CTBS/TerraNova. The matching algorithm produced well-matched samples for CTBS/TerraNova analyses. *KS* statistics for 2003–2004, 2004–2005, and 2005–2006 were 0.03, 0.11, and 0.001, with bootstrapped *p*-values of 0.42, 0.38, and 1.00, respectively. This suggested the matching algorithm effectively reduced covariate bias between samples for all three years. No differences in CTBS/TerraNova scores were found in any year (Table 5). In 2004–2005 only Wheaton Woods students provided CTBS/TerraNova data, so comparisons are made with only one Reading First school for that year.

Table 4
Differences Between Reading First and Comparison Schools for Grade 3 MSA Reading and MAP-R; and Grade 2 CTBS/TerraNova, by Cohort

	Mean Difference	*SE*	*t*	*Df*	*ES*
Grade 3 Reading MSA					
2003–2004 (*N*=691 339/352[1])	-3.62	2.90	-1.25	675	-0.05
2004–2005 (*N*=718 344/374)	-1.71	2.82	-0.60	702	-0.02
2005–2006 (*N*=661 345/316)	1.32	2.51	0.53	647	0.02
Grade 3 Reading MSA controlling for Grade 2 CTBS					
2003–2004 (*N*=623 319/304)	-7.56	2.36	-3.21	606	-0.13
2004–2005 (*N*=584 282/302)	-2.16	2.45	-0.88	567	-0.04
2005–2006 (*N*=356 270/86)	9.28	4.19	2.22	341	0.12
Grade 3 MAP-R RIT Score					
2004–2005 (*N*=448 341/107[2])	-0.52	1.60	-0.32	432	-0.01
2005–2006 (winter *N*=549 336/213[3])	-3.51	1.27	-2.76	535	-0.12
Grade 2 CTBS/TerraNova					
2003–2004 (*N*=738 357/381)	-5.58	3.01	-1.89	722	-0.07
2004–2005 (*N*=438 338/100[4])	3.68	5.27	0.70	422	0.03
2005–2006 (*N*=648 315/333)	4.08	2.57	1.58	635	0.06

Note: [1]Overall sample size and sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *Df*=Degrees of Freedom. *ES*=Effect size (Cohen's *d)*. [2]Wheaton Woods was the only Reading First school that supplied 2004–2005 MAP-R test scores. [3]Highland was the only Reading First school that did not supply 2005–2006 MAP-R test scores. MAP-R data from 2003–2004 are not available. [4]Only Wheaton Woods had CTBS/TerraNova data available for analysis in Grade 2. RIT Score = Rasch unIT Score.

Effect sizes by cohort. All effect sizes for tests on which there were at least two years of comparisons made were collected and are shown graphically in Figure 11. The effect sizes are all small (≤ |.10|) and so are not interpretable as findings in favor of Reading First, but the pattern of effect sizes for each cohort appear to increase with time, except for MAP-R.

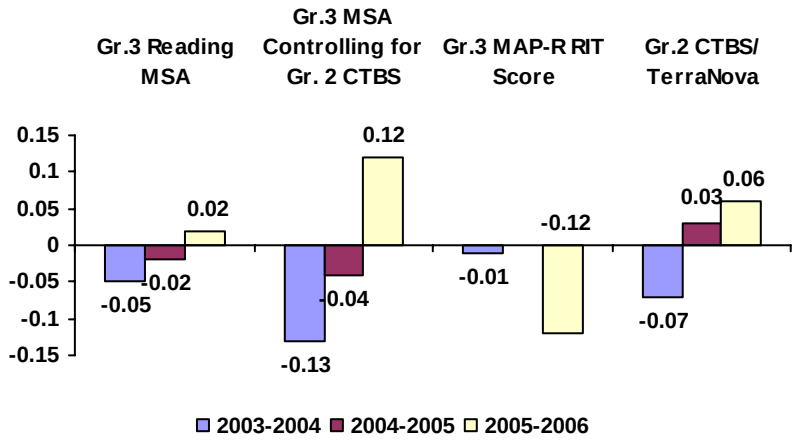

Figure 11. Effect sizes for cohorts 2003–2004 to 2005–2006 on standardized tests.

DIBELS Subtests

DIBELS subtests were administered to both Reading First and non-Reading First schools in 2005–2006, so this is the only time point at which the two schools can be compared. Although there are benchmark cutoff points which were used to treat scores as categorical outcomes, they are used as continuous variables in the analyses below.

Kindergarten DIBELS subtests. The matching algorithm produced well-matched samples for the kindergarten analyses. For 2005–2006, after matching *KS* = 0.009 and the bootstrapped *p*-value = 0.43, suggesting matching reduced covariate bias between samples.

No difference was found on the LNF test, but a significant difference in favor of Reading First was found for the PSF test (moderate effect size) and a significant but small effect favoring Reading First for the NWF subtest (Table 6).

Grade 1 DIBELS subtests. As with the kindergarten analyses, the matching algorithm produced well-matched samples for the Grade 1 analyses. For 2005–2006, after matching *KS* = 0.006, bootstrapped *p*-value = 0.46, suggesting matching reduced covariate bias between samples.

Significant differences were found favoring Reading First for all three subtests (Table 6). Significant effects were found favoring Reading First: a moderate effect on the PSF subtest and a small effect on the NWF subtest. The effect for the ORF subtest was significant but very small.

Grade 2 DIBELS ORF subtest. As with both kindergarten and Grade 1 analyses, the matching algorithm produced well-matched samples for the Grade 2 analysis. For 2005–2006, after matching *KS* = 0.01, bootstrapped *p*-value = 0.99, suggesting matching reduced covariate bias between samples.

The ORF subtest is the only DIBELS subtest administered to Grade 2 students at the end of the year. A significant difference was found favoring Reading First (Table 6), however, the effect size is very small (*ES* = 0.08).

Table 5
Differences Between Reading First and Comparison Schools for DIBELS Subtests 2005–2006, by Grade

Effect	Mean Difference	*SE*	*t*	*Df*	*ES*
Kindergarten DIBELS subtest					
Letter Naming Fluency (*N*=699 353/346[1])	0.53	1.26	0.42	684	0.02
Phoneme Segmentation Fluency (N=699 353/346)	16.90	1.17	14.46	684	0.55
Nonsense Word Fluency (*N*=699 353/346)	6.92	1.55	4.47	684	0.17
Grade 1 DIBELS subtests					
Phoneme Segmentation Fluency (*N*=650 296/354)	12.41	1.08	11.52	635	0.45
Nonsense Word Fluency (*N*=650 296/354)	14.54	2.45	5.93	635	0.23
Oral Reading Fluency *(N*=650 296/354)	5.16	2.54	2.03	635	0.08
Grade 2 DIBELS Oral Reading					
Fluency (*N*=675 320/355)	5.87	2.73	2.15	660	0.08

Note: [1]Overall sample size and sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *DF*=Degrees of Freedom. *ES*=Effect size (Cohen's *d*).

Discussion

Implementation Evaluation

Responses from teachers, coaches, and principals in the staff interviews suggest that instructional goals (SBRR curriculum, 90 minutes of protected reading instruction time) are being met, although differentiation needs more work. Assessment goals are being met by using DIBELS, which is used consistently across all four schools, as are professional development goals (coaches are available, evidence of availability, and attendance at training). Taken together, these observations suggest that schools generally adhere to the Reading First goals. Particular strengths appear to be the structure and availability of materials, although some teachers and administrators believe the structure inhibits creativity. Two weaknesses of the program are 1) difficulty accommodating below- and above-grade level students, and 2) scheduling reading time so that it is not interrupted and students receiving other services, such as ESOL, still receive them and so that other topics are not eliminated from the student's school day. Two staff members had strong negative reactions to the program, believing that the standard curriculum is superior.

The classroom observations show that teachers are spending instructional time on the five basic reading skills (phonemic awareness, phonics, fluency, vocabulary, and comprehension) which is generally consistent with recommendations by Reading First guidelines. However, many teachers are including writing (drawing/painting for younger students) during the reading instructional period. Teachers feel the need to find ways to fit in related subjects such as writing during reading time so that students receive some exposure to writing related to reading instruction.

Training attendance records suggest that the majority of teachers have obtained mandatory trainings related to Reading First, although attendance at trainings in 2005–2006 had decreased. In part, this was due to teachers having already obtained training at a prior time and not needing to repeat it, but it may also have been due to the change in status of certain classes from mandatory to elective.

Outcome Evaluation

Descriptive analyses of standardized tests show that cohorts of Grade 3 students have generally increased in MSA reading achievement scores (2003–2004 to 2005–2006) but decreased in MAP-R scores (2004–2005 to 2005–2006). Grade 2 cohorts appear to increase and then decrease from 2003–2004 to 2005–2006. Within cohorts, female students have higher scores than male students across grades and tests, White and Asian American students have higher scores than African American and Hispanic students, and students receiving ESOL or FARMS services have lower scores than students not receiving them. These between group differences are also generally observed for DIBELS benchmark attainment. With 2005–2006 data when Reading First and non-Reading First comparisons are possible, more Reading First students appear to be performing in the "Low Risk" or "Established" categories than students in non-Reading First schools (except for LNF in kindergarten). These differences have not been submitted for statistical testing, and so no statement about the significance of these descriptive differences is suggested.

The inferential analyses comparing Reading First to non-Reading First schools show that for standardized tests such as the CTBS/TerraNova and MSA, no robust differences were found. Small negative effects of Reading First were found for the most recent MAP-R data available for Grade 3 students. The analyses of Grade 3 MSA controlling for prior performance on the CTBS/TerraNova revealed small negative effects of Reading First in 2003–2004 but small positive effects in 2005–2006. The latter should be considered with caution since the only Reading First school involved in that comparison was Wheaton Woods. In general, standardized testing in 2004–2005 was conducted with incomplete data because some Reading First schools did not administer certain tests to their students. The lack of consistent administration of standardized tests limits the number of inferences that can be made.

Results for DIBELS subtests were much more favorable for Reading First. Moderate effects of Reading First were found for kindergarten and Grade 1 PSF subtests. Small effects of Reading First were found for kindergarten and Grade 1 NWF subtests, and for Grade 1 and Grade 2 ORF subtests. However, since non-Reading First schools did not administer DIBELS subtests before 2005–2006, it is unknown how the differences between Reading First and non-Reading First schools may have appeared over time. It is also important to consider that Reading First students, having been exposed to the test multiple times over a longer period of time than non-Reading First students, may perform better on the test because of student practice effects or because Reading First teachers had more experience with DIBELS administration procedures (Good, Kaminski, Smith, Simmons, Kame'enui, & Wallin, 2002). Differences attributable to exposure to the outcome measures used are considered problems in instrumentation (Cook & Campbell, 1979) and limit causal inferences about program effectiveness.

Another possible explanation is that teachers in Reading First schools are not blind to the intervention students are receiving. Since teachers administer the DIBELS subtests but not the standardized achievement tests, any administration bias that may exist would work in favor of DIBELS. Teachers with favorable attitudes toward the intervention may unknowingly influence the performance of students in ways consistent with their attitudes, although the magnitude of the effects of teacher expectancies of students has been found to be small and of short duration (Jussim & Harber, 2005).

Performance on DIBELS subtests are generally highly correlated with performance on standardized reading achievement examinations, accounting for as much as 55 percent of variance in reading achievement, especially the ORF subtest (Carlisle et al., 2004). In MCPS in 2005–2006, Grade 2 students TN/2 reading scale scores correlated $r = .54$ ($N = 640$, $r^2 = .30$) with the DIBELS ORF subtest scores. It is interesting to note that the smallest effects of Reading First in this evaluation were found with the ORF subtest, and no effects were observed with standardized testing. This suggests that a continuum of proximal to distal outcomes of Reading First may exist, consisting of the LNF/PSF subtests on the proximal end, in which large differences were found between Reading First and non-Reading First schools; and standardized achievement tests (CTBS/TerraNova, MSA) on the distal outcome end, where no differences were found between Reading First and non-Reading First schools. ORF falls between the two ends of this continuum. If Reading First has an effect, it may be that it produces effects in proximal outcome measures (LNF and PSF) because Reading First students and teachers in MCPS have had more exposure to DIBELS than their counterparts.

The effect of Reading First may be limited to DIBELS subtests since it does not appear to produce differences among distal outcomes like the MSA, MAP-R, or CTBS/TerraNova. From this point of view, it could be argued that the LNF and PSF subtests do not share much common variance with the ORF test or standardized reading tests, and that differences between Reading First schools and non-Reading First schools are found in the unique variance of the DIBELS subtests. If a Reading First effect exists, but can only be detected by DIBELS performance this may cast doubt on the efficacy of the intervention, since students' performance on standardized testing has considerable effects on their academic careers.

Conclusion

Reading First appears to be implemented as intended, with a few minor exceptions, but does not appear to produce marked differences in students' achievement, as measured by the MSA or CTBS/TerraNova. Differences in DIBELS outcomes much more clearly show that Reading First students' scores exceed those of non-Reading First students, but this may be due to the close relationship between the use of DIBELS and the instruction received at schools, experience with DIBELS, or both.

Recommendation

The investigation of DIBELS subtests in Reading First and non-Reading First schools should be revisited again in 2006-2007, when non-Reading First schools have had another year of experience administering it. Reexamining effects of Reading First on standardized achievement tests would also be useful, since evidence exists of increasing effects from 2003 to 2006. This is especially important to consider given the lag in effect found in other state Reading First evaluations.

References

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74, 235-267.

Carlisle, J. F., Cortina, K., Zeng, J., & Schilling, S. G. (2006). Gains in reading achievement over two years in Michigan's Reading First schools. *Technical Report #3.1, Evaluation of Reading First in Michigan.* Ann Arbor: University of Michigan.

Carlisle, J. F., Schilling, S.G., Scott, S .E., & Zeng, J. (2004). Do fluency measures predict reading achievement? Results from the 2002–2003 school year in Michigan's Reading First schools. *Technical Report #1, Evaluation of Reading First in Michigan*. Ann Arbor: University of Michigan.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Edition). Hillsdale, NJ: Erlbaum.

Cook T. D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Boston: Houghton Mifflin.

Good, R. H., Kaminski, R. A., Smith, S., Simmons, D., Kame'enui, E., & Wallin, J. (2002). Reviewing outcomes: Using DIBELS to evaluate a school's core curriculum and system of additional intervention in kindergarten. In S. R. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for observing teaching and learning*. Baltimore: Paul H. Brookes.

Haager, D., Dhar R., Moulton, M. & Varma, S. (2005). *The California Year 3 Reading First Evaluation Report*. Sacramento, CA: California Reading First.

Institute for the Development of Educational Achievement (IDEA). (2006). *Learn how to teach it*. University of Oregon.

Jussim L., & Harber K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. Personality and Social Psychology Review, 9 (2), 131–155.

Landis, J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, (33*), 159–174.

Levesque, J., Schnell, T., Richardson, L., Pickering, R., Quinn, K., Ding, C., & Hyken, T. (2005). *Missouri Reading First annual performance report*. St. Louis, MO: University of Missouri-St. Louis.

McDonald N., Trautman T., & Corey, D. (2005). *Influencing early literacy skills: An experimental study of a computer-based intervention.* Oklahoma City, OK: The American Education Corporation.

Noble, A., & Grusenmeyer L. (2005). *Evaluation of Delaware's Reading First initiative: Year II report*. Newark, DE: University of Delaware, Delaware Education Research and Development Center.

Oregon Reading First Center (2006). *Program fidelity checklist*. Eugene, OR: University of Oregon.

Pressley, M., Hilden, K. R., & Shankland, R. K. (2006). *An evaluation of end-grade-3 dynamic indicators of Basic Early Literacy Skills (DIBELS): Speed reading without comprehension, predicting little*. East Lansing, MI: Michigan State University, College of Education, Literacy Achievement Research Center (LARC).

R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Sekhon, J. S. (2006). *Matching: Algorithms and software for multivariate and propensity score matching with balance optimization via genetic search*.

SPSS. (2002). *SPSS for Windows, Rel. 11.5.1*. Chicago: SPSS Inc.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. (2005). *Analysis of state K–3 reading standards and assessments.* Washington, D.C.

WGBH Educational Foundation. (2002). *Teaching reading K–2: A library of classroom practices. Videotape Library.* Videotape. ISBN: 1–57680–520–4.

Appendix A Target Population Demographics

Table A1
Demographics of Target Population

		Year					
		2003–2004		2004–2005		2005–2006	
		N	%	*N*	%	*N*	%
Kindergarten	Gender						
	Female	387	46.9	386	51.0	388	49.5
	Male	438	53.1	371	49.0	396	50.5
	Race/Ethnicity						
	African American	163	19.8	166	21.9	163	20.8
	American Indian	2	0.2	7	0.9	7	0.9
	Asian American	112	13.6	84	11.1	87	11.1
	Hispanic	418	50.7	392	51.8	427	54.5
	White	130	15.8	108	14.3	100	12.8
	Special Services						
	ESOL (Never)	461	55.9	353	46.6	326	41.6
	ESOL (Prior)	3	0.4	1	0.1	23	2.9
	ESOL (Now)	361	43.8	403	53.2	435	55.5
	FARMS (Never)	325	39.4	274	36.2	278	35.5
	FARMS (Prior only)	36	4.4	44	5.8	69	8.8
	FARMS (Current)	464	56.2	439	58.0	437	55.7
	Special Education = No	757	91.8	695	91.8	722	92.1
	Special Education = Yes	68	8.2	62	8.2	62	7.9
	Reading First School						
	No	411	49.8	347	45.8	401	51.1
	Yes	414	50.2	410	54.2	383	48.9
Grade 1	Gender						
	Female	408	50.5	363	46.5	381	50.3
	Male	400	49.5	418	53.5	376	49.7
	Race/Ethnicity						
	African American	182	22.5	159	20.4	180	23.8
	American Indian	5	0.6	2	0.3	3	0.4
	Asian American	95	11.8	77	9.9	82	10.8
	Hispanic	416	51.5	421	53.9	399	52.7
	White	110	13.6	122	15.6	93	12.3
	Special Services						
	ESOL (Never)	434	53.7	428	54.8	303	40.0
	ESOL (Prior)	134	16.6	66	8.5	89	11.8
	ESOL (Now)	240	29.7	287	36.7	365	48.2
	FARMS (Never)	238	29.5	250	32.0	255	33.7
	FARMS (Prior only)	79	9.8	82	10.5	81	10.7
	FARMS (Current)	491	60.8	449	57.5	421	55.6

	Reading First School						
	No	398	49.3	409	52.4	363	48.0
	Yes	410	50.7	372	47.6	394	52.0
Grade 2	Gender						
	Female	387	45.4	393	50.1	370	47.8
	Male	466	54.6	391	49.9	404	52.2
	Race/Ethnicity						
	African American	165	19.3	167	21.3	171	22.1
	American Indian	4	0.5	5	0.6	2	0.3
	Asian American	105	12.3	94	12.0	69	8.9
	Hispanic	453	53.1	428	54.6	405	52.3
	White	126	14.8	90	11.5	127	16.4
	Special Services						
	ESOL (Never)	442	51.8	403	51.4	392	50.6
	ESOL (Prior)	184	21.6	178	22.7	154	19.9
	ESOL (Now)	227	26.6	203	25.9	228	29.5
	FARMS (Never)	248	29.1	191	24.4	228	29.5
	FARMS (Prior only)	103	12.1	110	14.0	103	13.3
	FARMS (Current)	502	58.9	483	61.6	443	57.2
	Special Education = No	753	88.3	701	89.4	690	89.1
	Special Education = Yes	100	11.7	83	10.6	84	10.9
	Reading First School						
	No	403	47.2	390	49.7	378	48.8
	Yes	450	52.8	394	50.3	396	51.2
Grade 3	Gender						
	Female	382	48.4	381	45.6	386	49.1
	Male	407	51.6	454	54.4	400	50.9
	Race/Ethnicity						
	African American	166	21.0	172	20.6	170	21.6
	American Indian	5	0.6	4	0.5	6	0.8
	Asian American	87	11.0	78	9.3	101	12.8
	Hispanic	412	52.2	467	55.9	422	53.7
	White	119	15.1	114	13.7	87	11.1
	Special Services						
	ESOL (Never)	417	52.9	434	52.0	377	48.0
	ESOL (Prior)	169	21.4	223	26.7	214	27.2
	ESOL (Now)	203	25.7	178	21.3	195	24.8
	FARMS (Never)	219	27.8	215	25.7	191	24.3
	FARMS (Prior only)	77	9.8	144	17.2	138	17.6
	FARMS (Current)	493	62.5	476	57.0	457	58.1
	Special Education = No	690	87.5	723	86.6	681	86.6
	Special Education = Yes	99	12.5	112	13.4	105	13.4
	Reading First School						
	No	397	50.3	412	49.3	402	51.1
	Yes	392	49.7	423	50.7	384	48.9

Appendix B Staff and Parent Interview Protocols

Reading First Teacher Interview

Hello, my name is _____. The department of Shared Accountability (DSA) is in the process of conducting an evaluation of Reading First (RF) program. I would like to ask you about the RF program at your school. Our goal is to describe the training and activities that principals, RF coaches, and RF teachers engage in at different schools. The information that you provide will help us understand how the program operates in practice. This is a semi-structured interview—that is, I have a series of questions to which I would like for you to respond, but please feel free to make whatever extended comments you would like if you deem it to be necessary. I should also inform you that quotes may be taken from your interview as the evaluation report is prepared. No one will be identified by name as the source of the quotation. Do you have any questions before we begin?

<u>About the interviewee</u>

1. How long have you been a ***Reading First teacher***? How long at this school?
2. Please describe your role in RF implementation at your school.

<u>Professional Development</u>

1. Please describe any kinds of training, support, or other professional development provided to you during the 2005–2006 school year to implement RF.
    a. (probe if needed) To what extent were you able to attend RF training opportunities?
    b. (probe if needed) Describe the role of the RF Coach in your professional development.
    c. (probe if needed) What has been the most valuable resource for you as RF teacher?
2. As a result of your RF professional development experiences, what changes have you made in how you work with students?
3. Does your school provide teachers with feedback about their RF instruction? If so, how is it provided?
4. What type of training opportunities would you like to see offered in the future?

<u>Curriculum Implementation</u>

1. Describe how the RF curriculum is implemented at your school.
    a. (probe if needed) What curriculum is used for instruction? For supplementation? For intervention?
2. How is the curriculum coordinated with other programs (e.g., Title I, special education, language arts)?
3. How do you work with your colleagues to implement the RF curriculum?
    a. (probe if needed) Do teachers meet to collaborate on RF instruction? How often? In subject or grade level? Across subject or grade level?
4. Please describe what methods you have available for monitoring student progress in reading and how you use them.
5. What resources are available for students receiving ESOL, special education, or Gifted and Talented services in RF? For students who are struggling?

   a. (probe if needed) Who provides the additional help and how do they provide it?
6. Describe the role of the principal and RF coach in implementing RF at your school.
7. In your opinion, is reading instruction time sufficiently protected? why or why not?

Program Evaluation

1. What aspects of the program are working well for students? teachers?
2. What aspects of the program need improvement?

Other

1. Is there any additional support that you could use to provide more effective reading instruction?
2. Are there any other people I should speak with at your school to better understand how RF is implemented here?
3. Is there anything else you would like to tell me about the RF program at your school?

Two final questions:

Overall, how would you rate the overall quality of implementation of Reading First program in your school?

□ Poor □ Fair □ Good □ Very Good □ Excellent

Would you recommend the program to another school with similar demographic characteristics?

□ Strongly Recommend	□ Recommend	□ Recommend with reservations	□ Do not recommend

Reading First Coach Interview

Hello, my name is _____. The department of Shared Accountability (DSA) is in the process of conducting an evaluation of Reading First (RF) program. I would like to ask you about the RF program at your school. Our goal is to describe the training and activities that principals, RF coaches, and RF teachers engage in at different schools. The information that you provide will help us understand how the program operates in practice. This is a semi-structured interview—that is, I have a series of questions to which I would like for you to respond, but please feel free to make whatever extended comments you would like if you deem it to be necessary. I should also inform you that quotes may be taken from your interview as the evaluation report is prepared. No one will be identified by name as the source of the quotation. Do you have any questions before we begin?

About the interviewee

1. How long have you been a ***Reading First Coach***? How long at this school?
2. Please describe your role in RF implementation at your school.

Professional Development

1. Please describe any kinds of training, support, or other professional development provided to you during the 2005–2006 school year to implement RF.
    a. (probe if needed) To what extent were you able to attend RF training opportunities?
    b. (probe if needed) Describe the role of the RF Principal in your professional development.
    c. (probe if needed) What has been the most valuable resource for you as RF coach?
2. As a result of your RF professional development experiences, what changes have you made in how you work with teachers?
3. Does your school provide teachers with feedback about their RF instruction? If so, how is it provided?
4. What type of training opportunities would you like to see offered in the future?

Curriculum Implementation

1. Describe how the RF curriculum is implemented at your school.
    a. (probe if needed) What curriculum is used for instruction? For supplementation? For intervention?
2. How is the curriculum coordinated with other programs (e.g., Title I, special education, language arts)?
3. How do you work with teachers to implement the RF curriculum?
    a. (probe if needed) How often do you meet with teachers?
4. Please describe what methods are available for monitoring student progress in reading and how they are used.
5. What resources are available for students receiving ESOL, special education, or Gifted and Talented services in RF? For students who are struggling?
    a. (probe if needed) Who provides the additional help and how do they provide it?
6. Describe the role of the principal in implementing RF at your school.
7. In your opinion, is reading instruction time sufficiently protected? Why or why not?

Program Evaluation

1. What aspects of RF are working well for students? Teachers? Coaches?
2. What aspects of RF need improvement?

Other

1. Is there any additional support you could use to provide more effective reading instruction?
2. Are there any other people I should speak with at your school to better understand how RF is implemented here?
3. Is there anything else you would like to tell me about the RF program at your school?

Two final questions:

Overall, how would you rate the overall quality of implementation of RF program in your school?

□ Poor □ Fair □ Good □ Very Good □ Excellent

Would you recommend the program to another school with similar demographic characteristics?

□ Strongly Recommend	□ Recommend	□ Recommend with reservations	□ Do not recommend

Reading First in Montgomery County Public Schools (MCPS)
Parent Survey (Grades K–3)

The MCPS' Department of Shared Accountability (DSA) is evaluating the Reading First (RF) program during the 2005–2006 school year. We have several questions for you about how the RF program is being implemented at your child's school. Your responses to the questions will help us understand the program better. We will use the results of the overall evaluation to help decide how the school administrators can improve RF at your school. All of your responses will be confidential. This means that although we will use your responses in the evaluation, no one will be able to identify you with any particular statement you may make on this survey because your names will not be included on the form. This should take about 10–15 minutes.

The following nine questions ask you about the RF program at your school. Principals, RF coaches, and RF teachers and parents do different types of activities to make this program work. We want to know how well you think RF is being implemented and how it affects your student's reading achievement.

Reading First: Parent Questions

In what grades are your children?

_____

How long has your child or children been at this school?

_____

1. Describe what you know about the RF program at this school.

2. How do you know about the RF program? Please describe any communication you receive about the program.

3. How is your child's progress in reading monitored?

4. What aspects, if any, of the reading program at your school do you think are working well?

5. What aspects, if any, of the reading program at your school do you think need improvement?

6. Is there anything else you would like to tell me about the RF program at your student's school?

2nd and 3rd Grade Parents Only

2nd grade parents – Answer only if your child was at this school for Kindergarten
3rd grade parents – Answer only if your child was at this school for 1st grade

This is the second year of the RF at your school

7. Is the amount of time spent on reading instruction different since RF started two years ago? If so, how?

8. Describe any other academic changes you have noticed in your student since RF started two years ago.

Reading First Principal Interview

Hello, my name is _____. The department of Shared Accountability (DSA) is in the process of conducting an evaluation of Reading First (RF) program. I would like to ask you about the RF program at your school. Our goal is to describe the training and activities that principals, RF coaches, and RF teachers engage in at different schools. The information that you provide will help us understand how the program operates in practice. This is a semi-structured interview—that is, I have a series of questions to which I would like for you to respond, but please feel free to make whatever extended comments you would like if you deem it to be necessary. I should also inform you that quotes may be taken from your interview as the evaluation report is prepared. No one will be identified by name as the source of the quotation. Do you have any questions before we begin?

About the interviewee:

1. How long have you been a ***principal***? How long at this school?
2. How long have you worked in a RF school?

Professional Development

1. Please describe any kinds of support, training, or other professional development provided to you during the 2005–06 school year to implement RF.
    a. (probe if needed) To what extent were you able to attend RF training opportunities?
    b. (probe if needed) What has been the most valuable resource for you as RF principal?
2. As a result of your RF professional development experiences, what changes have you made in how you work with RF teachers and coaches?
3. What type of training opportunities would you like to see offered in the future?

Curriculum Implementation

1. Who are the members of the RF Leadership Team at this school?
    a. (probe if needed) Are all RF positions fully staffed?
2. Describe how the RF curriculum is implemented at your school.
    a. (probe if needed) What curriculum is used for instruction? For supplementation? For intervention?
3. How is the RF curriculum coordinated with other programs (e.g., Title I, special education, language arts)?
4. How do you work with your teachers and coaches to implement the RF curriculum?
    a. (probe if needed) How often do you meet with teachers and/or coaches?
5. Please describe what methods are available for monitoring student progress in reading and how they are used.
6. What resources are available for students receiving ESOL, special education, or Gifted and Talented services in RF? For students who are struggling?
    a. (probe if needed) Who provides the additional help and how do they provide it?
7. Describe the role of the principal in implementing RF at your school.
8. In your opinion, is reading instruction time sufficiently protected? Why or why not?

Program Evaluation

1. What aspects of RF are working well for students? Teachers? Coaches?
2. What aspects of RF need improvement?

Other

1. Is there any additional support that you, the coaches, or teachers could use to provide more effective reading instruction?
2. Are there any other people I should speak with at your school to better understand how RF is implemented here?
3. Is there anything else you would like to tell me about the RF program at your school?

Two final questions:

Overall, how would you rate the overall quality of implementation of RF program in your school?

□ Poor □ Fair □ Good □ Very Good □ Excellent

Would you recommend the program to another school with similar demographic characteristics?

□ Strongly Recommend	□ Recommend	□ Recommend with reservations	□ Do not recommend

Appendix C Training Data

The attendance data for teachers in RF schools for mandatory and elective training conducted in 2004–2005 and 2005–2006 is summarized in Tables C1-C4 below.

Table C1
Grades Taught by Reading First Teachers Attending Any Training in 2004 –2005

	Highland (*N*=58)	Rosemont (*N*=36)	Summit Hall (*N*=36)	Wheaton Woods (*N*=54)	Combined (*N*=184)
Grade/Position					
K	10% (5)	18% (6)	*	15% (8)	13% (22)
1	13% (7)	18% (6)	17% (6)	13% (7)	15% (26)
2	12% (6)	*	14% (5)	13% (7)	13% (23)
3	15% (8)	*	*	9% (5)	11% (19)
ESOL	17% (9)	*	17% (6)	15% (8)	16% (27)
Paraeducator	*	*	14% (5)	*	4% (7)
Reading Specialist	12% (6)	18% (6)	9% (3)	9% (5)	11% (20)
Special Education	10% (5)	*	*	*	7% (12)

Note: Shaded area represents RF focus on Grades K–3. * = sample size too small to report.

Table C2
Trainings Attended by Reading First Teachers in 2004–2005

	Highland (*N*=58)	Rosemont (*N*=36)	Summit Hall (*N*=36)	Wheaton Woods (*N*=54)	Combined (*N*=184)
Mandatory Training					
MIBR	78% (45)	75% (27)	83% (30)	65% (35)	74% (137)
DIBELS	76% (44)	75% (27)	36% (13)	56% (30)	62% (114)
Mandatory, Specific Interventions					
Read Well K	12% (7)	23% (8)	28% (10)	19% (10)	20% (35)
Read Well 1	40% (23)	39% (14)	25% (9)	24% (13)	33% (59)
Read Naturally	*	*	*	27% (5)	11% (20)
Early Reading Intervention	*	*	19% (7)	19% (10)	13% (24)
Mandatory, choose 2/4					
Phonemic Awareness	33% (19)	14% (5)	17% (6)	9% (5)	19% (35)
Phonics	44% (26)	39% (14)	11% (4)	60% (32)	43% (76)
Fluency	6% (4)	14% (5)	8% (3)	26% (14)	14% (26)
Vocabulary	59% (34)	67% (24)	31% (11)	43% (23)	49% (92)
Optional					
TOWK MSDE	*	*	*	*	3% (6)
GORT MSDE	*	*	*	*	3% (6)

Note: *=sample size too small to report

Table C3
Grades Taught by Reading First Teachers Attending any Training in 2005–2006

Grade/Position	Highland (*N*=56)	Rosemont (*N*=36)	Summit Hall (*N*=45)	Wheaton Woods (*N*=58)	Combined (*N*=195)
K	11% (6)	19% (7)	*	16% (9)	13% (26)
1	11% (6)	19% (7)	16% (7)	12% (7)	14% (27)
2	12% (7)	14% (5)	11% (5)	12% (7)	12% (24)
3	18% (10)	*	13% (6)	9% (5)	12% (24)
4	*	*	*	10% (6)	5% (10)
ESOL	20% (11)	*	13% (6)	14% (8)	15% (29)
Paraeducator	*	*	11% (5)	*	4% (7)
Reading Specialist	9% (5)	17% (6)	11% (5)	9% (5)	11% (21)
Special Education	*	*	*	*	7% (13)

Note: Shaded area represents Reading First focus on grades K–3. * = sample size too small to report.

Table C4
Trainings Attended by Reading First Teachers in 2005–2006

	Highland (*N*=56)	Rosemont (*N*=36)	Summit Hall (*N*=45)	Wheaton Woods (*N*=58)	Combined (*N*=195)
Mandatory					
Houghton Mifflin New Teachers	12% (7)	14% (5)	*	14% (8)	12% (23)
MIBR					
2005–2006	*	*	*	*	10% (14)
Previously completed	66% (37)	69% (25)	62 (27)	53% (34)	64% (124)
Optional (required in 2004–2005)					
Read Well K	*	*	*	*	1% (2)
Read Well 1	*	17% (6)	15% (7)	*	7% (13)
Read Naturally	*	17% (6)	*	*	4% (6)

Note: * = sample size too small to report.

Inter-rater Agreement

Rater agreement was assessed at two time points: once when training was completed and once in the field during a live classroom reading instruction period. To obtain data to assess inter-rater agreement following training, both observers independently rated the same previously unviewed videotape at the end of training. Agreement data were collected in the field when both observers independently rated the same classroom setting at the same time. Instances of agreement on the type of instructional activity (phonemic awareness, phonics, fluency, vocabulary, comprehension, or other) were counted along with disagreements for both occasions. Agreement was quantified using weighted kappa.

Kappa coefficients can range from 1.0, representing perfect agreement, to negative infinity, representing near perfect disagreement, with kappas of zero indicating that the level of agreement between raters is equivalent to chance. Landis and Koch (1977) suggested descriptive

labels for kappa coefficients greater than zero. Kappa coefficients between .41–.60 represent moderate inter-rater agreement, coefficients between .61–.80 represent substantial inter-rater agreement, and coefficients between .81–1.00 represent almost perfect agreement. Kappa coefficients obtained at the conclusion of training were in the moderate to substantial range (Table C5), except for the category of vocabulary instruction, which was poor. Discussion of instructional segments on the training video for which there were rater disagreements led to resolution for those segments and ultimately resulted in improved inter-rater agreement in the field. All inter-rater agreement kappa coefficients for the classroom observations were in the substantial or almost perfect range for classroom observations.

Table C5
Inter-rater Agreement (Kappas) for Training Videotape and Classroom Observations

Activity	Training Videotape	Classroom
Phonemic awareness	.63	.72
Phonics	.58	.79
Fluency	.72	.69
Vocabulary	-.07	.91
Comprehension	.53	.71
Other	.63	.88

Appendix D Assessment Data

Table D1
Grade 3 MSA Reading Means and Standard Deviations,
by Year, Demographics, and Service Characteristics

	2003–2004			2004–2005			2005–2006		
	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*
Gender									
Female	403.6	35.3	331	410.6	35.6	326	411.9	33.2	315
Male	393.0	37.9	365	402.6	35.5	396	406.1	31.6	349
Ethnicity									
African American	397.3	35.7	141	405.7	34.1	144	406.8	36.2	141
American Indian	429.6	39.8	5	388.8	31.6	4	397.7	18.9	3
Asian American	416.8	37.8	79	417.5	39.6	71	424.6	32.9	87
Hispanic	390.5	34.1	367	399.8	32.9	404	403.3	27.3	357
White	410.0	39.8	104	425.4	38.1	99	421.0	38.7	76
Special Services									
Current ESOL (No)	404.5	37.3	524	411.1	34.9	568	413.2	32.9	513
Current ESOL (Yes)	378.3	28.4	172	388.0	32.8	154	394.0	26.1	151
Current FARMS (No)	410.2	38.3	245	415.7	38.7	270	418.6	34.2	260
Current FARMS (Yes)	391.4	34.7	451	400.5	33.1	452	402.6	29.7	404
Special Education (No)	401.2	36.4	600	409.1	35.1	621	411.7	31.9	573
Special Education (Yes)	378.1	35.1	96	388.5	34.6	101	391.0	30.8	91

Note: Includes both Reading First and non-Reading First Schools

Table D2
Grade 3 MSA Reading Means and Standard Deviations,
by Year and School

	2003–2004			2004–2005			2005–2006		
	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*
Schools									
Glenhaven	387.6	32.0	84	407.7	32.5	88	409.1	39.1	91
Highland[RF]	390.7	34.0	103	389.3	33.7	118	405.3	27.5	98
Kemp mill	395.8	34.6	98	406.8	33.4	92	410.2	30.8	99
Rosemont[RF]	399.1	39.9	61	417.0	31.5	67	400.3	27.9	51
Summit Hall[RF]	392.6	31.2	82	399.7	31.4	84	406.8	31.7	69
Twinbrook	407.4	34.4	88	413.8	41.5	71	411.6	33.0	71
Weller road	413.4	40.2	85	407.0	37.2	95	408.1	31.7	84
Wheaton Woods[RF]	399.1	43.3	95	415.5	36.2	107	415.1	34.4	101
Reading First School									
No	401.0	36.6	355	408.5	36.0	346	409.7	33.7	345
Yes	395.0	37.3	341	404.0	35.5	376	408.0	31.1	319
Overall	398.1	37.1	696	406.2	35.8	722	408.9	32.5	664

Note: [RF =] Reading First school

Table D3
Grade 3 MAP-R Means and Standard Deviations, by Year, Demographics, and Service Characteristics

	2004–2005			2005–2006		
	Mean	*SD*	*N*	Mean	*SD*	*N*
Gender						
Female	197.8	13.1	209	194.4	14.4	299
Male	194.6	16.0	243	191.0	15.3	333
Race/Ethnicity						
African American	195.8	14.8	98	192.7	16.0	132
American Indian	199.7	11.6	3	202.5	16.3	2
Asian American	199.7	13.3	51	199.5	11.7	83
Hispanic	193.7	14.3	236	190.0	14.3	340
White	202.1	15.9	64	195.7	16.3	74
Special Services						
Current ESOL (No)	197.7	14.4	363	195.3	14.2	496
Current ESOL (Yes)	189.3	14.3	89	183.0	13.5	136
Current FARMS (No)	200.5	14.0	186	197.8	14.2	250
Current FARMS (Yes)	193.0	14.5	266	189.3	14.5	382
Special Education (No)	198.0	13.1	402	194.5	14.0	556
Special Education (Yes)	180.4	17.7	50	178.9	14.7	76

Table D4
Grade 3 MAP-R Means and Standard Deviations, by Year and School

	2004–2005			2005–2006		
	Mean	*SD*	*N*	Mean	*SD*	*N*
Schools						
Glenhaven	193.4	15.1	87	193.4	15.5	86
Highland[RF]	N/A	N/A	N/A	N/A	N/A	N/A
Kemp mill	200.9	12.2	86	197.5	13.7	93
Rosemont[RF]	N/A	N/A	N/A	192.3	10.6	45
Summit Hall[RF]	N/A	N/A	N/A	185.4	16.4	70
Twinbrook	194.9	17.7	78	190.9	15.0	67
Weller road	195.2	13.9	93	195.6	14.3	80
Wheaton Woods[RF]	195.9	14.2	108	191.6	14.6	96
Reading First School						
No	196.1	15.0	344	194.6	14.7	326
Yes	195.9	14.2	108	189.7	14.8	306
Overall	196.1	14.8	452	192.7	14.9	537

Note: [RF]Reading First school. N/A= Not Available

Table D5
Grade 2 CTBS/TerraNova Reading Means and Standard Deviations,
by Year and Demographics and Service Characteristics

	2003–2004			2004–2005			2005–2006		
	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*
Gender									
Female	613.4	37.3	331	621.9	39.0	210	615.1	35.8	315
Male	601.2	41.1	410	608.6	39.1	228	609.2	32.3	335
Race/Ethnicity									
African American	604.3	40.1	144	617.2	37.1	89	608.9	32.5	140
American Indian	613.0	34.2	3	N/A	N/A	N/A	569.5	53.0	2
Asian American	614.1	40.4	84	626.3	42.0	63	624.2	36.0	57
Hispanic	600.0	37.2	399	607.7	38.4	230	607.6	30.9	352
White	627.5	41.3	111	628.6	39.5	56	626.3	40.0	99
Special Services									
Current ESOL = No	613.7	38.8	558	622.7	38.7	328	617.0	34.5	458
Current ESOL = Yes	585.2	35.3	183	592.0	32.8	110	600.3	30.2	192
Current FARMS = No	617.8	45.0	289	625.8	42.3	161	621.3	36.4	256
Current FARMS = Yes	599.5	34.4	452	608.7	36.6	277	606.1	31.2	394
Special Education=No	608.8	39.6	655	616.8	39.8	396	613.8	34.5	574
Special Education=Yes	590.0	38.4	86	597.4	32.7	42	598.7	27.7	76

Table D6
Grade 2 CTBS/TerraNova Reading Means and Standard Deviations,
by Year and Demographics and Service Characteristics

	2003–2004			2004–2005			2005–2006		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Schools									
Glenhaven	612.5	34.9	77	611.6	34.8	79	619.6	33.4	73
Highland[RF]	599.7	37.5	122	N/A	N/A	N/A	609.7	28.9	87
Kemp mill	618.3	42.1	100	620.2	44.8	95	605.9	44.7	80
Rosemont[RF]	612.3	40.4	72	N/A	N/A	N/A	618.0	35.7	80
Summit Hall[RF]	602.2	39.1	82	N/A	N/A	N/A	616.7	30.1	64
Twinbrook	609.4	44.8	87	610.4	32.1	76	604.9	37.5	82
Weller road	599.5	38.2	95	616.9	35.1	88	610.4	29.8	82
Wheaton Woods[RF]	602.9	38.4	106	614.6	46.3	100	613.0	29.2	102
Reading First School									
No	609.9	40.8	359	615.1	37.4	338	610.0	37.1	317
Yes	603.5	38.8	382	614.6	46.3	100	614.1	31.0	333
Overall	606.6	39.9	741	615.0	39.6	438	612.1	34.1	650

Note: [RF]Reading First school. Some Reading First schools did not administer the CTBS/TerraNova. The 2005–2006 test data are from the TN/2.

Table D7
Kindergarten Student 2004–2005 DIBELS End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools

	Letter Naming Fluency[1]			Phoneme Segmentation Fluency[2]			Nonsense Word Fluency[3]		
	0–28 % (*N*)	29–39 % (*N*)	40+ % (*N*)	0–9 % (*N*)	10–34 % (*N*)	35+ % (*N*)	0–14 % (*N*)	15–24 % (*N*)	25+ % (*N*)
Gender									
Female	10% (18)	18% (33)	72% (134)	2% (4)	13% (25)	84% (156)	6% (12)	12% (22)	82% (151)
Male	13% (24)	25% (44)	62% (110)	7% (12)	19% (34)	74% (132)	15% (20)	19% (34)	66% (118)
Race/Ethnicity									
African American	16% (10)	17% (11)	67% (43)	6% (4)	12% (8)	81% (52)	8% (5)	19% (12)	73% (47)
American Indian	0% (0)	0% (0)	100% (1)	0% (0)	0% (0)	100% (1)	0% (0)	0% (0)	100% (1)
Asian American	0% (0)	11% (4)	89% (31)	3% (1)	17% (6)	80% (28)	0% (0)	6% (2)	94% (33)
Hispanic	11% (24)	26% (56)	62% (133)	4% (8)	18% (39)	78% (166)	11% (24)	16% (34)	73% (155)
White	16% (8)	12% (6)	72% (36)	6% (3)	12% (6)	82% (41)	18% (9)	16% (8)	66% (33)
Special Services									
Current ESOL = No	12% (16)	21% (28)	67% (91)	4% (6)	16% (22)	79% (107)	10% (14)	18% (24)	72% (97)
Current ESOL = Yes	11% (26)	21% (49)	67% (153)	4% (10)	16% (37)	79% (181)	10% (24)	14% (32)	75% (172)
Current FARMS = No	8% (10)	13% (17)	79% (104)	3% (4)	12% (16)	85% (111)	8% (10)	11% (15)	81% (106)
Current FARMS = Yes	14% (32)	26% (60)	60% (140)	5% (12)	18% (43)	76% (177)	12% (28)	18% (41)	70% (163)
Special Education = No	10% (35)	20% (67)	69% (233)	3% (11)	14% (46)	83% (278)	8% (28)	14% (47)	78% (260)
Special Education =Yes	25% (7)	36% (10)	39% (11)	18% (5)	46% (13)	36% (10)	36% (10)	32% (9)	32% (9)
Overall	12% (42)	21% (77)	67% (244)	4% (16)	16% (59)	79% (288)	10% (38)	15% (56)	74% (269)

Note: [1]0–28 = At Risk, 29–39 = Some Risk, 40+ = Low Risk. [2]0–9 Deficit, 10–34 = Emerging, 35+ Established [3]0–14 = At Risk, 15–24 = Some Risk, 25+ Low Risk

Table D8
Kindergarten Student 2004–2005 DIBELS End-of-Year Test Performance, by School

	Letter Naming Fluency[1]			Phoneme Segmentation Fluency[2]			Nonsense Word Fluency[3]		
	0–28	29–39	40+	0–9	10–34	35+	0–14	15–24	25+
	% (*N*)	% (*N*)	% (*N*)	% (*N*)	% (*N*)	% (*N*)	% (*N*)	% (*N*)	% (*N*)
Rosemont	20% (19)	22% (21)	58% (56)	4% (4)	20% (19)	76% (73)	20% (19)	15% (14)	66% (63)
Summit Hall	6% (5)	27% (21)	67% (53)	4% (3)	18% (14)	78% (62)	4% (3)	16% (13)	80% (63)
Highland	17% (17)	31% (31)	52% (52)	8% (8)	18% (18)	74% (74)	14% (14)	24% (24)	62% (62)
Wheaton Woods	1% (1)	5% (4)	94% (83)	1% (1)	9% (8)	90% (79)	2% (2)	6% (5)	92% (81)
Overall	12% (42)	21% (77)	67% (244)	4% (16)	16% (59)	79% (288)	10% (38)	15% (56)	74% (269)

Note: [1]0–28 = At Risk, 29–39 = Some Risk, 40+ = Low Risk. [2]0–9 Deficit, 10–34 = Emerging, 35+ Established. [3]0–14 = At Risk, 15–24 = Some Risk, 25+ Low Risk

Table D9
Kindergarten Student 2005–2006 DIBELS Letter Naming Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and non-Reading First Schools

	Non-Reading First Schools						Reading First Schools					
	0–281		29–39		40+		0–28		29–39		40+	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Gender												
Female	9	6%	31	19%	125	76%	19	10%	44	24%	119	65%
Male	25	13%	29	15%	136	72%	16	10%	35	21%	115	69%
Race/Ethnicity												
African American	2	2%	15	17%	70	80%	3	5%	13	24%	39	71%
American Indian	0	0%	0	0%	2	100%	0	0%	1	50%	1	50%
Asian American	2	4%	3	7%	41	89%	1	3%	10	29%	24	69%
Hispanic	30	18%	37	22%	98	59%	30	13%	49	22%	145	65%
White	0	0%	5	9%	50	91%	1	3%	6	19%	25	78%
Special Services												
Current ESOL = No	4	2%	21	12%	146	85%	4	3%	24	20%	89	76%
Current ESOL = Yes	30	16%	39	21%	115	63%	31	13%	55	24%	145	63%
Current FARMS = No	4	3%	16	10%	134	87%	8	6%	26	20%	103	75%
Current FARMS = Yes	30	15%	44	22%	125	63%	27	13%	52	25%	130	62%
Special Education=No	29	9%	50	16%	238	75%	29	9%	72	22%	223	69%
Special Education=Yes	5	14%	10	28%	21	58%	6	27%	6	27%	10	45%

Note: [RF]Reading First school. [1]0–28 = At Risk, 29–39 = Some Risk, 40+ = Low Risk.

Table D10
Kindergarten Student 2005–2006 DIBELS Letter Naming Fluency
End-of-Year Test Performance, by School

	0–28[1]		29–39		40+	
	N	%	*N*	%	*N*	%
Schools						
Glenhaven	4	4%	13	13%	80	82%
Highland[RF]	7	8%	21	23%	62	69%
Kemp mill	19	18%	20	19%	66	63%
Rosemont[RF]	10	13%	21	27%	46	60%
Summit Hall[RF]	5	7%	16	22%	53	72%
Twinbrook	4	5%	12	16%	62	79%
Weller road	7	9%	15	20%	53	71%
Wheaton Woods[RF]	13	12%	21	20%	73	68%
Non-Reading First Schools	34	10%	60	17%	261	73%
Reading First Schools	35	10%	79	23%	234	67%
Overall	69	10%	139	20%	495	70%

Note: [RF]Reading First school. [1]0–28 = At Risk, 29–39 = Some Risk, 40+ = Low Risk.

Table D11
Kindergarten Student 2005–2006 DIBELS Phoneme Segmentation Fluency
End-of-Year Test Performance, by Demographics and Service Provision Characteristics
in Reading First and non-Reading First Schools

	Non-Reading First Schools						Reading First Schools					
	0–9[1]		10–34		35+		0–9		10–34		35+	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Gender												
Female	16	10%	71	43%	76	47%	5	3%	16	9%	161	88%
Male	29	15%	85	45%	76	40%	9	5%	18	11%	139	84%
Race/Ethnicity												
African American	9	10%	40	46%	38	44%	0	0%	6	11%	49	89%
American Indian	0	0%	0	0%	2	100%	0	0%	0	0%	2	100%
Asian American	5	11%	21	46%	20	43%	4	11%	4	11%	27	77%
Hispanic	31	19%	74	45%	60	36%	10	4%	22	10%	192	86%
White	0	0%	21	38%	34	62%	0	0%	2	6%	30	94%
Special Services												
Current ESOL = No	9	5%	62	36%	100	58%	0	0%	7	6%	110	94%
Current ESOL = Yes	36	20%	94	51%	54	29%	14	6%	27	12%	190	82%
Current FARMS = No	8	5%	56	36%	92	59%	5	4%	7	5%	126	91%
Current FARMS = Yes	37	19%	100	50%	62	31%	9	3%	27	13%	174	83%
Special Education=No	37	12%	138	43%	144	45%	11	3%	30	9%	285	87%
Special Education=Yes	8	22%	18	50%	10	28%	3	14%	4	18%	15	68%

Note: [RF]Reading First school. Note: [1]0–9 Deficit, 10–34 = Emerging, 35+ Established.

Table D12
Kindergarten Student 2005–2006 DIBELS Phoneme Segmentation Fluency End-of-Year Test, by School

	0–9[1]		10–34		35+	
	N	%	*N*	%	*N*	%
Schools						
Glenhaven	12	12%	33	34%	52	54%
Highland[RF]	2	2%	17	19%	71	79%
Kemp mill	17	16%	57	54%	31	30%
Rosemont[RF]	6	8%	7	9%	64	83%
Summit Hall[RF]	3	4%	4	5%	67	91%
Twinbrook	7	9%	31	40%	40	51%
Weller road	9	12%	35	47%	31	41%
Wheaton Woods[RF]	3	3%	6	6%	98	92%
Non-Reading First Schools	45	13%	156	44%	152	43%
Reading First Schools	14	4%	34	10%	300	86%
Overall	59	8%	190	27%	454	65%

Note: [RF]Reading First school. Note: [1]0–9 Deficit, 10–34 = Emerging, 35+ Established.

Table D13
Kindergarten Student 2005–2006 DIBELS Nonsense Word Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and non-Reading First Schools

	Non-Reading First Schools						Reading First Schools					
	0–14 At Risk		15–24 Some Risk		25+ Low Risk		0–14 At Risk		15–24 Some Risk		25+ Low Risk	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Gender												
Female	24	15%	30	18%	111	67%	15	8%	29	16%	138	76%
Male	41	22%	32	17%	117	62%	14	8%	27	16%	125	75%
Race/Ethnicity												
African American	12	14%	15	17%	60	69%	3	5%	4	7%	48	87%
American Indian	0	0%	1	50%	1	50%	0	0%	1	50%	1	50%
Asian American	5	11%	7	15%	34	74%	6	17%	3	9%	26	74%
Hispanic	44	27%	31	19%	90	55%	19	8%	45	20%	160	71%
White	4	7%	8	15%	43	78%	1	3%	3	9%	28	88%
Special Services												
Current ESOL = No	15	9%	24	14%	132	77%	6	5%	11	9%	100	85%
Current ESOL = Yes	50	27%	38	21%	96	52%	23	10%	45	19%	163	71%
Current FARMS = No	9	6%	21	13%	126	81%	11	8%	15	11%	112	81%
Current FARMS = Yes	56	28%	41	21%	102	51%	18	9%	41	20%	151	72%
Special Education=No	57	18%	53	17%	209	66%	24	7%	50	16%	251	77%
Special Education=Yes	8	22%	9	25%	19	53%	5	23%	5	23%	12	55%

Note: [RF]Reading First school.

Table D14
Kindergarten Student 2005–2006 DIBELS Nonsense Word Fluency End-of-Year Test Performance, by School

	0-14		15-24		25+	
	N	%	*N*	%	*N*	%
Schools						
Glenhaven	13	13%	12	12%	72	74%
Highland[RF]	5	6%	16	18%	69	77%
Kemp mill	33	31%	18	17%	54	51%
Rosemont[RF]	7	9%	15	19%	55	71%
Summit Hall[RF]	5	7%	13	18%	56	76%
Twinbrook	10	13%	18	23%	50	64%
Weller road	9	12%	14	19%	52	69%
Wheaton Woods[RF]	12	11%	12	11%	83	78%
ReadingFirst Schools = No	65	18%	62	17%	228	64%
Reading First Schools = Yes	29	8%	56	16%	263	76%
Overall	94	13%	118	17%	491	70%

Table D15
Grade 1 Student 2004–2005 DIBELS End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools

	Phoneme Segmentation Fluency			Nonsense Word Fluency			Oral Reading Fluency		
	0–9[1] % (*N*)	10–34 % (*N*)	35+ % (*N*)	0–29[2] % (*N*)	30–49 % (*N*)	50+ % (*N*)	0–19[3] % (*N*)	20–39 % (*N*)	40+ % (*N*)
Gender									
Female	0% (0)	7% (11)	93% (143)	3% (5)	27% (42)	70% (107)	17% (27)	27% (41)	56% (86)
Male	1% (2)	16% (23)	83% (122)	9% (13)	30% (44)	61% (90)	26% (38)	24% (35)	50% (74)
Race/Ethnicity									
African American	2% (1)	10% (5)	88% (45)	4% (2)	25% (13)	71% (36)	16% (8)	22% (11)	63% (32)
American Indian	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Asian American	5% (1)	5% (1)	89% (17)	10% (2)	26% (5)	63% (12)	5% (1)	26% (5)	68% (13)
Hispanic	0% (0)	15% (28)	85% (163)	7% (14)	31% (59)	62% (118)	28% (53)	28% (54)	44% (84)
White	0% (0)	0% (0)	100% (40)	0% (0)	22% (9)	77% (31)	7% (3)	15% (6)	77% (31)
Special Services									
Current ESOL = No	1% (1)	8% (14)	91% (149)	2% (3)	25% (41)	73% (120)	12% (20)	17% (28)	71% (116)
Current ESOL=Yes	1% (1)	15% (20)	85% (116)	9% (13)	33% (46)	57% (78)	33% (45)	35% (48)	32% (44)
Current FARMS = No	1% (1)	5% (5)	94% (94)	5% (5)	18% (18)	77% (77)	14% (14)	21% (21)	65% (65)
Current FARMS = Yes	0.5% (1)	14% (29)	85% (171)	5% (11)	34% (69)	60% (121)	25% (51)	27% (55)	47% (95)
Special Education=No	1% (2)	11% (30)	88% (246)	6% (17)	28% (77)	66% (184)	19% (54)	24% (68)	56% (156)
Special Education=Yes	0% (0)	17% (4)	83% (19)	4% (1)	39% (9)	57% (13)	48% (11)	35% (8)	17% (4)
Overall	1% (2)	11% (34)	88% (265)	6% (18)	29% (86)	65% (197)	22% (65)	25% (76)	53% (160)

Note: [1]0–9 Deficit, 10–34 Emerging, 35+ Established. [2]0–29 Deficit, 30–49 Emerging, 50+ Established. [3]0–19 At Risk, 20–39 Some Risk, 40+ Low Risk.

Table D16
Grade 1 Student 2004–2005 DIBELS End-of-Year Test Performance, by School

	Phoneme Segmentation Fluency			Nonsense Word Fluency			Oral Reading Fluency		
	0–9[1] % (*N*)	10–34 % (*N*)	35+ % (*N*	0–29[2] % (*N*)	30–49 % (*N*)	50+ % (*N*)	0–19[3] % (*N*)	20–39 % (*N*)	40+ % (*N*)
Schools									
Rosemont	1% (1)	4% (3)	95% (79)	2% (2)	13% (11)	84% (70)	13% (11)	22% (18)	65% (54)
Summit Hall	1% (1)	9% (6)	89% (59)	6% (4)	27% (18)	67% (44)	15% (10)	35% (23)	50% (33)
Highland	0% (0)	25% (23)	75% (68)	13% (12)	40% (36)	47% (43)	36% (33)	29% (26)	35% (32)
Wheaton Woods	0% (0)	3% (2)	97% (59)	0% (0)	34% (21)	66% (40)	18% (11)	15% (9)	67% (41)
Overall	1% (2)	11% (34)	88% (265)	16% (5)	29% (87)	66% (198)	21% (65)	25% (76)	53% (160)

Note: No non-Reading First schools administered DIBELS in 2004–2005. [1]0–9 Deficit, 10–34 Emerging, 35+ Established. [2]0–29 Deficit, 30–49 Emerging, 50+ Established. [3]0–19 At Risk, 20–39 Some Risk, 40+ Low Risk.

Table D17
Grade 1 Student 2005–2006 DIBELS Phoneme Segmentation Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and non-Reading First Schools

	Non-Reading First Schools						Reading First Schools					
	0–9[1]		10–34		35+		0–9		10–34		35+	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Gender												
Female	3	2%	39	26%	108	72%	2	1%	9	5%	168	94%
Male	3	2%	58	39%	87	59%	0	0%	12	7%	163	93%
Race/Ethnicity												
African American	2	2%	20	23%	64	74%	1	2%	9	14%	55	85%
American Indian	0	0%	1	50%	1	50%	0	0%	0	0%	0	0%
Asian American	0	0%	13	36%	23	64%	0	0%	1	3%	34	97%
Hispanic	4	3%	57	42%	76	55%	1	1%	11	5%	199	94%
White	0	0%	6	16%	31	84%	0	0%	0	0%	43	100%
Special Services												
Current ESOL = No	3	2%	38	24%	118	74%	1	1%	6	4%	153	96%
Current ESOL = Yes	3	2%	59	42%	77	55%	1	1%	15	8%	178	92%
Current FARMS = No	2	2%	40	31%	89	68%	0	0%	10	8%	120	92%
Current FARMS = Yes	4	2%	57	34%	106	63%	2	1%	11	5%	211	94%
Special Education = No	3	1%	86	31%	185	67%	2	1%	16	5%	311	95%
Special Education =Yes	3	13%	11	46%	10	42%	0	0%	5	20%	20	80%

Note: [1]0–9 Deficit, 10–34 Emerging, 35+ Established.

Table D18
Grade 1 Student 2005–2006 DIBELS Phoneme Segmentation Fluency End-of-Year Test Performance, by School

	0–9		10–34		35+	
	N	%	*N*	%	*N*	%
Schools						
Glenhaven	0	0%	25	41%	36	59%
Highland[RF]	1	1%	6	6%	87	93%
Kemp Mill	0	0%	25	28%	63	72%
Rosemont[RF]	1	1%	11	11%	84	88%
Summit Hall[RF]	0	0%	2	3%	69	97%
Twinbrook	3	4%	11	15%	57	80%
Weller Road	3	4%	36	46%	39	50%
Wheaton Woods[RF]	0	0%	2	2%	91	98%
Reading First Schools = No	6	2%	97	33%	195	65%
Reading First Schools = Yes	2	1%	21	6%	331	94%
Overall	8	1%	118	18%	526	81%

Note: [RF]Reading First school. [1]0–9 Deficit, 10–34 Emerging, 35+ Established.

Table D19
Grade 1 Student 2005–2006 DIBELS Nonsense Word Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and non-Reading First Schools

	Non-Reading First Schools						Reading First Schools					
	0–29[1]		30–49		50+		0–29		30–49		50+	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Gender												
Female	16	11%	45	30%	89	59%	4	2%	28	16%	147	82%
Male	29	20%	46	31%	73	49%	11	6%	36	21%	128	73%
Race/Ethnicity												
African American	14	16%	29	34%	43	50%	3	5%	13	20%	49	75%
American Indian	0	0%	1	50%	1	50%	0	0%	0	0%	0	0%
Asian American	0	0%	9	25%	27	75%	1	3%	2	6%	32	91%
Hispanic	31	23%	41	30%	65	47%	11	5%	44	21%	156	74%
White	0	0%	11	30%	26	70%	0	0%	5	12%	38	88%
Special Services												
Current ESOL = No	18	11%	48	30%	93	58%	2	1%	20	12%	138	86%
Current ESOL = Yes	27	19%	43	31%	69	50%	13	7%	44	23%	137	71%
Current FARMS = No	17	13%	30	23%	84	64%	5	4%	17	13%	108	83%
Current FARMS = Yes	28	17%	61	36%	78	47%	10	4%	47	21%	167	75%
Special Education = No	34	12%	85	31%	155	57%	13	4%	56	17%	260	79%
Special Education = Yes	11	46%	6	25%	7	29%	2	8%	8	32%	15	60%

Note: [RF]Reading First school. [1]0–29 Deficit, 20–49 Emerging, 50+ Established.

Table D20
Grade 1 Student 2005–2006 DIBELS Nonsense Word Fluency End-of-Year Test Performance, by School

	0–29[1]		30–49		50+	
	N	%	*N*	%	*N*	%
Schools						
Glenhaven	7	11%	17	28%	37	61%
Highland[RF]	8	9%	36	38%	50	53%
Kemp Mill	7	8%	33	38%	48	55%
Rosemont[RF]	4	4%	12	13%	80	83%
Summit Hall[RF]	2	3%	12	17%	57	80%
Twinbrook	4	6%	17	24%	50	70%
Weller Road	27	35%	24	31%	27	35%
Wheaton Woods[RF]	1	1%	4	4%	88	95%
Non-Reading First Schools	45	15%	91	30%	162	54%
Reading First Schools	15	4%	64	18%	275	78%
Overall	60	9%	155	24%	437	67%

Note: [RF]Reading First school. [1]0–29 Deficit, 20–49 Emerging, 50+ Established.

Table D21
Grade 1 Student 2005–2006 DIBELS Oral Reading Fluency End of Year Test Performance by Demographics and Service Provision Characteristics in Reading First and non-Reading First Schools

	Non-Reading First Schools						Reading First Schools					
	0–19[1]		20–39		40+		0–19		20–39		40+	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Gender												
Female	11	7%	43	29%	96	65%	17	9%	33	18%	129	72%
Male	33	22%	36	24%	79	53%	36	21%	36	21%	103	59%
Race/Ethnicity												
African American	11	13%	18	21%	57	66%	12	18%	16	25%	37	57%
American Indian	0	0%	2	100%	0	0%	0	0%	0	0%	0	0%
Asian American	2	6%	5	14%	29	81%	2	6%	4	11%	29	83%
Hispanic	27	20%	47	34%	63	46%	36	17%	43	20%	132	63%
White	4	11%	7	19%	26	70%	3	7%	6	14%	34	79%
Special Services												
Current ESOL = No	16	10%	36	23%	107	67%	15	9%	24	15%	121	76%
Current ESOL = Yes	28	20%	43	31%	68	49%	38	20%	45	23%	111	57%
Current FARMS = No	15	11%	25	19%	91	69%	11	9%	17	13%	102	78%
Current FARMS = Yes	29	17%	54	32%	84	51%	42	19%	52	23%	130	58%
Special Education = No	34	12%	72	26%	168	61%	42	13%	64	19%	223	68%
Special Education =Yes	10	42%	6	29%	7	29%	11	44%	5	20%	9	36%

Note: [1]0–19 At Risk, 20–39 Some Risk, 40+ Low Risk.

Table D22
Grade 1 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by School

	0–19[1]		20–39		40+	
	N	%	*N*	%	*N*	%
Schools						
Glenhaven	6	10%	16	26%	39	64%
Highland[RF]	15	16%	26	28%	53	56%
Kemp Mill	9	10%	20	23%	59	67%
Rosemont[RF]	20	21%	15	16%	61	64%
Summit Hall[RF]	12	17%	16	23%	43	61%
Twinbrook	8	11%	17	24%	46	65%
Weller Road	21	27%	26	33%	31	40%
Wheaton Woods[RF]	6	6%	12	13%	75	81%
Reading First Schools = No	44	15%	79	26%	175	59%
Reading First Schools = Yes	53	15%	69	19%	232	66%
Overall	97	15%	148	23%	407	62%

Note: [RF]Reading First school. [1]0–19 at Risk, 20–39 Some Risk, 40+ Low Risk.

Table D23
Grade 2 Student 2004–2005 DIBELS Oral Reading Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools

	0–69[1]		70–89		90+	
	%	*N*	%	*N*	%	*N*
Gender						
Female	33%	56	18% 30	*N*	49%	84
Male	42%	76	18% 32	*N*	41%	74
Race/Ethnicity						
African American	36%	22	13%	8	51%	31
American Indian	20%	1	40%	2	40%	2
Asian American	10%	4	19%	8	71%	29
Hispanic	46%	97	18%	39	36%	76
White	24%	8	15%	5	61%	20
Special Services						
Current ESOL = No	29%	70	16%	37	55%	131
Current ESOL = Yes	54%	62	22%	25	24%	27
Current FARMS = No	27%	31	14%	16	59%	67
Current FARMS = Yes	42%	101	19%	46	38%	91
Special Education = No	33%	103	18%	57	49%	153
Special Education =Yes	74%	29	13%	5	13%	5

Note: [1]0–69 = At Risk, 70–89 = Some Risk, 90+ Low Risk

Table D24
Grade 2 Student 2004–2005 DIBELS Oral Reading Fluency End-of-Year Test Performance, by School

	0–69[1]		70–89		90+	
	%	*N*	%	*N*	%	*N*
Schools						
Rosemont	33%	20	16%	10	51%	31
Summit Hall	47%	41	17%	15	36%	31
Highland	35%	35	28%	28	37%	37
Wheaton Woods	35%	36	9%	9	57%	59
Overall	38%	132	18%	62	45%	158

Note: [1]0–69 = At Risk, 70–89 = Some Risk, 90+ Low Risk

Table D25
Grade 2 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First and non-Reading First Schools

	Non-Reading First Schools						Reading First Schools					
	0–69[1]		70–89		90+		0–69		70–89		90+	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Gender												
Female	39	28%	20	14%	80	58%	42	23%	33	18%	108	59%
Male	66	36%	47	26%	70	38%	52	30%	33	19%	87	51%
Race/Ethnicity												
African American	28	35%	17	21%	36	44%	11	17%	10	16%	42	67%
American Indian	1	50%	0	0%	1	50%	0	0%	0	0%	0	0%
Asian American	4	11%	8	22%	25	68%	3	14%	3	14%	16	73%
Hispanic	60	40%	35	23%	56	37%	74	34%	45	20%	101	46%
White	12	24%	7	14%	32	63%	6	12%	8	16%	36	72%
Special Services												
Current ESOL = No	58	26%	46	20%	123	54%	30	13%	45	20%	161	70%
Current ESOL = Yes	47	50%	21	22%	27	28%	64	51%	27	22%	34	27%
Current FARMS = No	33	23%	25	18%	83	59%	19	15%	20	16%	86	69%
Current FARMS = Yes	72	40%	42	23%	67	37%	75	33%	46	20%	109	47%
Special Education = No	81	29%	60	21%	141	50%	72	22%	61	19%	188	59%
Special Education =Yes	24	60%	7	18%	9	23%	22	65%	5	15%	7	21%

Note: [1]0–69 At Risk, 70–89 Some Risk, 90+ Low Risk.

Table D26
Grade 2 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by School

	0–69[1]		70–89		90+	
	N	%	*N*	%	*N*	%
Schools						
Glenhaven	16	21%	15	20%	44	59%
Highland[RF]	45	48%	22	23%	27	29%
Kemp mill	23	29%	22	28%	35	44%
Rosemont[RF]	16	18%	15	17%	56	64%
Summit Hall[RF]	16	25%	11	17%	38	58%
Twinbrook	37	45%	15	18%	30	37%
Weller road	29	34%	15	18%	41	48%
Wheaton Woods[RF]	17	16%	18	17%	74	68%
Reading First Schools = No	105	33%	67	21%	150	47%
Reading First Schools = Yes	94	26%	66	19%	195	55%
Overall	199	29%	133	20%	345	51%

Note: [RF]Reading First school. [1]0–69 At Risk, 70–89 Some Risk, 90+ Low Risk.

Table D27
Grade 3 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by Demographics and Service Provision Characteristics in Reading First Schools

	0–79		80–109		110+	
	N	%	*N*	%	*N*	%
Gender						
Female	38	23%	47	28%	82	49%
Male	42	25%	60	36%	66	39%
Race/Ethnicity						
African American	20	34%	15	26%	23	40%
American Indian	0	0%	1	33%	2	67%
Asian American	1	3%	7	19%	29	78%
Hispanic	56	27%	78	38%	73	35%
White	3	10%	6	20%	21	70%
Special Services						
Current ESOL = No	32	14%	74	32%	126	54%
Current ESOL = Yes	48	47%	33	32%	22	21%
Current FARMS = No	19	16%	35	29%	66	55%
Current FARMS = Yes	61	28%	72	34%	82	38%
Special Education = No	56	19%	91	31%	145	50%
Special Education = Yes	24	56%	16	37%	3	7%

Note: 0–79 = At Risk, 80–109 = Some Risk, 110+ = Low Risk

Table D28
Grade 3 Student 2005–2006 DIBELS Oral Reading Fluency End-of-Year Test Performance, by School

	0–79		80–109		110+	
	N	%	*N*	%	*N*	%
Schools						
Rosemont	15	25%	17	28%	28	47%
Summit Hall	16	24%	20	29%	32	47%
Highland	30	30%	39	39%	32	32%
Wheaton Woods	19	18%	31	29%	56	53%
Overall	80	24%	107	32%	148	44%

Note: 0–79 = At Risk, 80–109 = Some Risk, 110+ = Low Risk

Appendix E. Results of Alternative Analyses of Student Achievement Data

In addition to the propensity score matched model, three other models were used to assess the effect of the intervention. First, an *unadjusted model* was run which is equivalent to an independent groups t-test using school type (Reading First versus non-Reading First) as the grouping variable and achievement as the outcome. This model was used to provide a reference point for comparing the other three models that adjust for demographic and service provision characteristics. Second, a *covariate-adjusted model* was used with the following covariates: gender, race, FARMS, ESOL, special education and proportion of school days in attendance. Third, a model containing all the covariates plus a propensity score adjustment was included to help determine the utility of propensity score adjustment relative to unadjusted and standard approaches to educational analyses. This model was referred to as the *covariate-adjusted and stratified on propensity score model*. All models were run using multiple regression analyses. Results for all four types of analyses are reported.

Differences between Reading First and non-Reading First schools were expressed as effect sizes (Cohen's *d*; Cohen, 1988). For all but the propensity score matched analysis this was an independent groups effect size estimate calculated from *t*, df, and sample sizes in each group ($ES = t(n_1 + n_2)/[\sqrt{df}(n_1 n_2)]$). For the propensity score matched analysis this is an effect size estimate based on a one-sample t-test for matched pairs ($ES = t/\sqrt{df}$). In all cases the Propensity Score Matched analysis was interpreted, with the remaining models providing a reference for comparison.

In all analyses the propensity score matched model yielded the most conservative effect size (i.e., closest to 0), as much as half the effect size obtained from the other analyses on the DIBELS outcomes for kindergarten and Grade 1 students. The check for balance among covariates in the matched sample showed that in all cases the matched samples used for the propensity score matched model were similar with respect to the covariates.

Table E1
Differences Between Reading First and Comparison Schools for Grade 3 MSA Reading, 2003–2004, 2004–2005 and 2005–2006

	Mean Difference	*SE*	*t*	*Df*	*ES*
2003–2004 (*N*=691 339/352[1])					
Unadjusted	-6.05	2.81	-2.16	689	-0.16
Covariate-Adjusted	-2.31	2.56	-0.90	679	-0.07
Covariate-Adjusted and Stratified on Propensity Score	-2.25	2.56	-0.88	675	-0.07
Propensity Score-Matched and Covariate-Adjusted	-3.62	2.90	-1.25	675	-0.05
2004–2005 (*N*=718 344/374)					
Unadjusted	-4.26	2.67	-1.60	716	-0.12
Covariate-Adjusted	-0.93	2.47	-0.38	706	-0.03
Covariate-Adjusted and Stratified on Propensity Score	-0.95	2.48	-0.38	702	-0.03
Propensity Score-Matched and Covariate-Adjusted	-1.71	2.82	-0.60	702	-0.02
2005–2006 (*N*=661 345/316)					
Unadjusted	-1.65	2.53	-0.65	659	-0.05
Covariate-Adjusted	0.94	2.35	0.40	651	-0.03
Covariate-Adjusted and Stratified on Propensity Score	1.17	2.36	0.49	647	0.04
Propensity Score-Matched and Covariate-Adjusted	1.32	2.51	0.53	647	0.02

Note: [1]Sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *Df*=Degrees of Freedom. *ES*=Effect size (Cohen's *d*). For 2003–2004, before matching KS test statistic *KS* = 0.20, bootstrapped *p*-value <0.0001 and after matching *KS* = 0.04, bootstrapped *p*-value = 0.99. For 2004–2005, before matching *KS* = 0.17, bootstrapped *p*-value=0.02 and after matching *KS* = 0.02, bootstrapped *p*-value = 0.78. For 2005–2006, before matching *KS* = 0.12, bootstrapped *p*-value = 0.51 and after matching the *KS* = 0.004, bootstrapped *p*-value = 0.80. This suggests that the matching algorithm resulted in well-balanced groups for all three years, although for 2005–2006 groups were initially reasonably well matched.

Table E2
Differences Between Reading First and Comparison Schools for Grade 3 MSA Reading 2003–2004, 2004–2005 and 2005–2006 with Grade 2 CTBS as Covariate

	Difference	*SE*	t	*Df*	*ES*
2003–2004 (*N*=623 319/304[1])					
Unadjusted	-6.66	2.99	-2.22	621	-0.18
Covariate-Adjusted	-5.84	2.12	-2.75	610	-0.22
Covariate-Adjusted and Stratified on Propensity Score	-5.90	2.13	-2.77	606	-0.22
Propensity Score-Matched and Covariate-Adjusted	-7.56	2.36	-3.21	606	-0.13
2004–2005 (*N*=584 282/302)					
Unadjusted	-6.95	2.91	-2.39	582	-0.20
Covariate-Adjusted	0.01	2.24	0.00	571	0
Covariate-Adjusted and Stratified on Propensity Score	0.05	2.24	0.02	567	0
Propensity Score-Matched and Covariate-Adjusted	-2.16	2.45	-0.88	567	-0.04
2005–2006 (*N*=356 270/86)					
Unadjusted	6.12	4.18	1.46	354	0.18
Covariate-Adjusted	7.95	3.39	2.35	345	0.29
Covariate-Adjusted and Stratified on Propensity Score	7.78	3.41	2.28	341	0.29
Propensity Score-Matched and Covariate-Adjusted	9.28	4.19	2.22	341	0.12

Note: [1]Sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *Df*=Degrees of Freedom. *ES*=Effect size (Cohen's *d*). For 2003–2004, before matching KS test statistic *KS* = 0.19, bootstrapped *p*-value 0.01 and after matching *KS* = 0.07, bootstrapped *p*-value = 0.26. For 2004–2005, before matching *KS* = 0.18, bootstrapped *p*-value=0.07 and after matching *KS* = 0.09, bootstrapped *p*-value = 0.996. For 2005–2006, before matching *KS* = 0.22, bootstrapped *p*-value = 0.10 and after matching the *KS* = 0.10, bootstrapped *p*-value = 0.77. This suggests that the matching algorithm resulted in well-balanced groups for all three years, although for 2005–2006 groups were well matched with respect to covariates prior to matching.

Table E3
Differences Between Reading First and Comparison Schools for Grade 3 MAP-R RIT Scores, 2004–2005 and 2005–2006 (Winter)

	Difference	*SE*	*t*	*Df*	*ES*
2003–2004 (*N*=448 341/107[12])					
Unadjusted	-0.19	1.62	-0.12	446	-0.01
Covariate-Adjusted	0.40	1.44	0.28	436	0.03
Covariate-Adjusted and Stratified on Propensity Score	0.37	1.44	0.26	432	0.03
Propensity Score-Matched and Covariate-Adjusted	-0.52	1.60	-0.32	432	-0.01
2005–2006 (winter *N*=549 336/213[3])					
Unadjusted	-5.53	1.34	-4.13	547	-0.36
Covariate-Adjusted	-3.94	1.17	-3.38	539	-0.30
Covariate-Adjusted and Stratified on Propensity Score	-3.86	1.17	-3.29	535	-0.29
Propensity Score-Matched and Covariate-Adjusted	-3.51	1.27	-2.76	535	-0.12

Note: [1]Sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *Df*=Degrees of Freedom. *ES*=Effect size (Cohen's *d*). [2]Wheaton Woods was the only Reading First school that supplied 2004–2005 MAP-R test scores. [3]Highland was the only Reading First school that did not supply 2005–2006 MAP-R test scores. MAP-R data from 2003–2004 are not available. For 2004–2005, before matching *KS* = 0.25, bootstrapped *p*-value < 0.0001 and after matching *KS* = 0.08, bootstrapped *p*-value = 0.72. For 2005–2006, before matching *KS* = 0.11, bootstrapped *p*-value = 0.41 and after matching *KS* = 0.01, bootstrapped *p*-value = 0.97, suggesting matching reduced covariate bias between samples for both years.

Table E4
Grade 2 Differences Between Reading First and Comparison Schools for CTBS Reading Scale Scores, 2003–2004, 2004–2005, and 2005–2006

	Difference	*SE*	*t*	*df*	*ES*
2003–2004 (*N*=738 357/381[1])					
Unadjusted	-6.22	2.93	-2.12	736	-0.16
Covariate-Adjusted	-4.61	2.67	-1.72	726	-0.13
Covariate-Adjusted and Stratified on Propensity Score	-4.55	2.69	-1.69	722	-0.13
Propensity Score-Matched and Covariate-Adjusted	-5.58	3.01	-1.89	722	-0.07
2004–2005 (*N*=438 338/100[2])					
Unadjusted	-0.55	4.51	-0.12	436	-0.01
Covariate-Adjusted	2.56	4.17	0.61	426	0.07
Covariate-Adjusted and Stratified on Propensity Score	2.32	4.19	0.55	422	0.06
Propensity Score-Matched and Covariate-Adjusted	3.68	5.27	0.70	422	0.03
2005–2006 (*N*=648 315/333)					
Unadjusted	3.84	2.67	1.44	646	0.11
Covariate-Adjusted	5.19	2.60	2.00	639	0.16
Covariate-Adjusted and Stratified on Propensity Score	4.97	2.59	1.92	635	0.15
Propensity Score-Matched and Covariate-Adjusted	4.08	2.57	1.58	635	0.06

Note: [1]Sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *DF*=Degrees of Freedom. *ES*=Effect size (Cohen's *d*). [2]Only Wheaton Woods had CTBS/TerraNova data available for analysis in Grade 2. For 2003–2004, before matching *KS* = 0.17, bootstrapped *p*-value < 0.03 and after matching *KS* = 0.03, bootstrapped *p*-value = 0.42. For 2004–2005, before matching *KS* = 0.22, bootstrapped *p*-value < 0.0001 and after matching *KS* = 0.11, bootstrapped *p*-value = 0.38. For 2005–2006, before matching *KS* = 0.17, bootstrapped *p*-value = 0.02 and after matching *KS* = 0.001, bootstrapped *p*-value = 1, suggesting matching reduced covariate bias between samples for all three years.

Table E5
Differences Between Reading First and Comparison Schools for Kindergarten DIBELS Subtests, 2005–2006

Effect	Difference	*SE*	*t*	*Df*	*ES*
Letter Naming Fluency (*N*=699 353/346[1])					
Unadjusted	-0.85	1.25	-0.69	697	-0.05
Covariate-Adjusted	0.31	1.21	0.26	688	0.02
Covariate-Adjusted and Stratified on Propensity Score	0.35	1.21	0.29	684	0.02
Propensity Score-Matched and Covariate-Adjusted	0.53	1.26	0.42	684	0.02
Phoneme Segmentation Fluency (*N*=699 353/346)					
Unadjusted	16.36	1.17	14.15	697	1.07
Covariate-Adjusted	17.09	1.10	15.55	688	1.18
Covariate-Adjusted and Stratified on Propensity Score	17.28	1.11	15.59	684	1.19
Propensity Score-Matched and Covariate-Adjusted	16.90	1.17	14.46	684	0.55
Nonsense Word Fluency *(N*=699 353/346)					
Unadjusted	4.97	1.48	3.35	697	0.25
Covariate-Adjusted	6.77	1.41	4.79	688	0.37
Covariate-Adjusted and Stratified on Propensity Score	6.90	1.43	4.83	684	0.37
Propensity Score-Matched and Covariate-Adjusted	6.92	1.55	4.47	684	0.17

Note: [1]Sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *Df*=Degrees of Freedom. *ES*=Effect size (Cohen's *d*). For 2005–2006, before matching *KS* = 0.21, bootstrapped *p*-value = <0.001 and after matching *KS* = 0.009, bootstrapped *p*-value = 0.43, suggesting matching reduced covariate bias between samples.

Table E6
Differences Between Reading First and Comparison Schools for Grade 1
DIBELS Subtests, 2005–2006

Effect	Difference	*SE*	*t*	*Df*	*ES*
Phoneme Segmentation Fluency (*N*=650 296/354[1])					
Unadjusted	12.30	1.01	12.17	648	0.96
Covariate-Adjusted	12.03	1.00	12.00	639	0.95
Covariate-Adjusted and Stratified on Propensity Score	12.00	1.01	11.91	635	0.95
Propensity Score-Matched and Covariate-Adjusted	12.41	1.08	11.52	635	0.45
Nonsense Word Fluency (*N*=650 296/354)					
Unadjusted	14.47	2.53	5.72	648	0.45
Covariate-Adjusted	14.56	2.39	6.08	639	0.48
Covariate-Adjusted and Stratified on Propensity Score	14.30	2.41	5.94	635	0.47
Propensity Score-Matched and Covariate-Adjusted	14.54	2.45	5.93	635	0.23
Oral Reading Fluency (*N*=650 296/354)					
Unadjusted	3.67	2.68	1.37	648	0.11
Covariate-Adjusted	4.77	2.43	1.96	639	0.16
Covariate-Adjusted and Stratified on Propensity Score	4.73	2.44	1.94	635	0.15
Propensity Score-Matched and Covariate-Adjusted	5.16	2.54	2.03	635	0.08

Note: [1]Sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *Df* =Degrees of Freedom. *ES*=Effect size (Cohen's d). For 2005–2006, before matching *KS*= 0.20, bootstrapped *p*-value = 0.002 and after matching *KS* = 0.006, bootstrapped *p*-value = 0.46, suggesting matching reduced covariate bias between samples.

Table E7
Differences Between Reading First and Comparison Schools for Grade 2 DIBELS Oral Reading Fluency 2005–2006

Effect	Difference	*SE*	*t*	*Df*	*ES*
Oral Reading Fluency (*N*=675 320/355[1])					
Unadjusted	4.36	2.87	1.52	673	0.12
Covariate-Adjusted	5.84	2.57	2.27	664	0.18
Covariate-Adjusted and Stratified on Propensity Score	5.51	2.57	2.14	660	0.17
Propensity Score-Matched and Covariate-Adjusted	5.87	2.73	2.15	660	0.08

Note: [1]Sample sizes for non-Reading First and Reading First schools respectively. *SE*=Standard Error, *Df*=Degrees of Freedom. *ES*=Effect size (Cohen's d). For 2005–2006, before matching *KS*= 0.20, bootstrapped *p*-value = 0.002 and after matching *KS* = 0.006, bootstrapped *p*-value = 0.46, suggesting matching reduced covariate bias between samples.