



# **Three-Year Report on Oregon Reading First *Impact and Implementation***

Prepared by  
**The Oregon Reading First Center**

Scott Baker, Ph.D.  
Jeanie Mercier Smith, Ph.D.  
Hank Fien, Ph.D.  
Janet Otterstedt, M.S.  
Rachell Katz, Ph.D.  
Doris Luft Baker, Ph.D.  
Carrie Thomas Beck, Ph.D.

For

The Oregon Department of Education and  
The Literacy Leadership State Steering Committee

**June, 2007**

## Table of Contents

Executive Summary .....	1
Background .....	1
Results .....	2
Introduction .....	5
Section I: Background and Context on Oregon Reading First .....	6
A Brief Overview of Oregon Reading First .....	6
Reading First Represents a Paradigm Shift in Reading Instruction.....	7
Scientifically Based Reading Research .....	8
A Brief History of Reading Reform.....	8
Determining the Impact of Previous Reforms .....	10
The Impact of Comprehensive School Reform Demonstration Project.....	11
Important Instructional Variations .....	12
Title I Reform .....	14
Reading First Reform .....	14
Section II: Implementation of Oregon Reading First .....	16
Major Components of the Federal Reading First Initiative.....	16
Oregon Reading First Implementation.....	17
Instructional Priorities and Practices.....	18
Reliable and Valid Assessment Used To Inform Instructional Practices.....	18
Protected and Sufficient Instructional Time .....	18
Data-Based Leadership and Administrative Involvement.....	19
High Quality Professional Development.....	20
Descriptions of the Districts, Schools, and Students .....	23
Number of Districts, Schools, and Students .....	23
Demographic Information .....	24
Teacher and Leadership Experience .....	27
Student Data Used to Determine Impact.....	28
Dynamic Indicators of Basic Early Literacy Skills (DIBELS).....	28
DIBELS Nonsense Word Fluency .....	29
DIBELS Oral Reading Fluency .....	30
Reading At Grade Level.....	31
Stanford Achievement Test-Tenth Edition [SAT-10] (Harcourt Brace Educational Measurement, 2002).....	31
Oregon Statewide Reading Assessment .....	33
Defining Grade Level Reading Performance on the Oregon Statewide Reading Assessment .....	33
Impact Questions and Analyses.....	34
Overview .....	34
Determining What Comparisons to Make.....	35
Four Ways of Examining Impact.....	35
Performance Across Years (Cohort A Over Time).....	35
Experienced Reading First Schools and Inexperienced Reading First Schools .....	36
Reading First Schools Versus Non-Reading First Schools .....	36
Performance of Intact Students Over Time .....	37

Analysis Measures and Performance Standards .....	37
Magnitude of Impact .....	38
Section III: Impact on Student Reading Achievement.....	40
Do Incoming Kindergartners Have Comparable Skills Across Years? .....	40
Has the Performance of Students in Cohort A Improved Over Time? .....	44
How Does the Reading Performance of Students in Cohorts A and B Compare? .....	53
What Is the Performance of Reading First Schools Compared to Non-Reading First Schools? .....	59
How Well do Students in Cohort A Perform If They Have Received Extensive Reading First Instruction.....	64
Section IV: Implications of the Three Year Findings .....	69
Summary of Findings.....	69
Background.....	69
Results .....	70
References .....	75

## Executive Summary

### Background

Oregon Reading First is part of the largest federal reading initiative ever undertaken. Although the scope of Reading First is vast, the goal of the initiative is transparent: *To ensure that every child reads at grade level or above by the end of third grade.* This report summarizes the impact of Oregon Reading First after three years of implementation. The primary focus is on student reading achievement.

Oregon Reading First has been implemented in 50 schools. In this report, we focus on two cohorts of Oregon Reading First schools, *Cohort A* (33 schools) and *Cohort B* (17 schools) and a cohort of non-Reading First comparison schools, *Cohort C* (6 schools). The percentages of English learners, minority students, and students eligible for free or reduced lunch prices are similar among the schools.

In Oregon, the *Schoolwide Beginning Reading Model*<sup>1</sup> is used to implement Reading First. This model prioritizes seven essential dimensions of reading instruction: (a) A schoolwide focus on essential reading content; (b) Regular use of reliable and valid assessments to inform instruction; (c) Protected and sufficient time allocated to reading instruction; (d) Data-based leadership devoted to sustained effective implementation and outcomes; (e) High-quality professional development that drives continuous improvement in the quality of reading instruction; (f) Research-based instructional programs and materials; (g) Differentiated instruction to optimize learning for all students.

A range of measures was used to estimate impact. DIBELS measures, used by Reading First schools to screen students for reading problems and monitor reading progress over time, were used to estimate the degree to which students met benchmark reading goals. Performance on two DIBELS measures was examined—Nonsense Word Fluency at the end of kindergarten, and Oral Reading Fluency at the end of grades 1, 2, and 3. These two measures were selected because they represent the most important DIBELS benchmarks that predict student performance on the Oregon Statewide Reading Assessment in third grade.

To determine grade level reading performance in kindergarten, first, and second grade, all Oregon Reading First students were administered the reading portion of the Stanford Achievement Test-10 (SAT-10) at the end of each year. Grade level performance on this *primary outcome measure* was defined as reading at the 40<sup>th</sup> percentile or above. Being at high risk for reading difficulties (i.e., well below grade level) was defined as reading below the 20<sup>th</sup> percentile.

*Grade level reading on the primary outcome measure in third grade was determined by student performance on the Oregon Statewide Reading Assessment (OSRA).* A score of 210 was used to define grade level reading on the OSRA rather than 201, which is defined as “meets proficiency,” according to state criteria. A score of 210 was selected because it corresponded to

---

<sup>1</sup> Kame’enui, Simmons, & Coyne (2000); Simmons, Kame’enui, Harn, Cole, & Braun (2002)

the 40<sup>th</sup> percentile in the first year of Oregon Reading First, and thus was comparable to the SAT-10 standard. A score of 201, in contrast, corresponded to the 16<sup>th</sup> percentile. *Across these measures, the evaluation targeted (a) mean performance scores, (b) the percentage of students reading at or above benchmark and grade level goals, and (c) the percentage of students at high risk for reading difficulties.*

The following four questions are highlighted in the report:

- Are Cohort A schools getting increasingly better reading outcomes each year of implementation?
- Are experienced Oregon Reading First schools (Cohort A, after three years of implementation) getting better reading outcomes than inexperienced Oregon Reading First schools (Cohort B, after one year implementation)?
- Is the performance of students in Oregon Reading First (Cohort A) better than the performance of students in non-Reading First comparison schools that were eligible for Reading First (Cohort C)?
- Are the outcomes for students in Cohort A who received three years of Reading First (i.e., kindergarten, first, and second grade) better than the outcomes of other groups of students in Cohort A who received less Reading First instruction?

Research on large-scale reading reform was used to anchor interpretations regarding the magnitude of impact of Oregon Reading First. For example, after 1-3 years of implementation, an impact corresponding to an effect size of approximately 0.15 is typical, according to research by Borman and colleagues.<sup>2</sup> In the most strongly established comprehensive school reform approaches, an effect size of around 0.20 was typical. The Institute for Education Sciences uses effect sizes of 0.25 to indicate meaningful impact. Thus, in Oregon Reading First, an impact in the range of 0.15 to 0.25 and above was considered to be educationally meaningful.

## Results

### **Are Cohort A schools getting increasingly better reading outcomes each year of implementation?**

In terms of the performance of Cohort A schools across years, the data are clear. In each grade, and on every measure, mean performance scores have increased consistently each year. In kindergarten, for example, the average score of students on *Nonsense Word Fluency* has more than doubled during Reading First. All of the effect sizes comparing Year 3 to Year 1 suggest the impact has been educationally meaningful.

In addition, across all measures, the percentage of children reaching benchmark or grade level goals has *increased* each year, and the percentage of children remaining at a high level of reading risk has *decreased*. This indicates that Cohort A schools are accomplishing two of the most important Reading First objectives: (a) *Cohort A schools are consistently increasing the*

---

<sup>2</sup> Borman, Hewes, Overman, & Brown (2003)

*percentage of children reading at grade level and (b) consistently decreasing the percentage of children at the highest levels of risk for reading difficulties.*

*The Cohort A analysis across years indicates that the impact of Reading First in all grades and on all measures has been moderate to large in magnitude thus far.*

**Are experienced Oregon Reading First schools (Cohort A) getting better reading outcomes than inexperienced Oregon Reading First schools (Cohort B)?**

In comparing the performance of Cohort A to Cohort B, after one year of implementation, reading outcomes were highly similar. Effect sizes were close to 0.0 and the small differences that did exist sometimes favored Cohort A and sometimes Cohort B. However, after three years of implementation in Cohort A schools (i.e., performance in 2005-2006) and one year of implementation in Cohort B schools (i.e., 2005-2006), differences are pronounced and favor Cohort A across the board. Higher scores for Cohort A occurred in all grades on both DIBELS measures and primary outcome measures. Effect sizes are consistently moderate to large in magnitude. This pattern suggests that Oregon Reading First is having a moderate to large impact on the reading skills of students in K-3. *The comparison between Cohort A after 3 years of implementation with Cohort B after 1 year of implementation is the best evidence of the value added of Oregon Reading First after multiple years of implementation.*

In the comparison between Cohorts A and B, impact was largest in kindergarten. Although somewhat smaller in grades 1, 2, and 3, the impact was still meaningful and roughly comparable across grades. In terms of the impact Oregon Reading First has had on the percentage of children reading at grade level and the percentage of children remaining at the highest level of reading risk, the outcomes favor Cohort A versus Cohort B. In every comparison, a *higher* percentage of children in Cohort A than Cohort B were reading at benchmark and grade level, and a *lower* percentage of children were at the highest level of reading risk. *In most cases, the odds of reaching benchmark or grade level were 1.5 times greater in Cohort A than Cohort B, and the odds of being at high risk for reading difficulties were 1.5 times greater in Cohort B than Cohort A.*

**Is the performance of students in Oregon Reading First (Cohort A) better than the performance of students in non-Reading First comparison schools that were eligible for Reading First (Cohort C)?**

The level of complexity increases when Cohort C schools (non-Reading First schools that were eligible for Reading First) are included in the analysis. The primary challenge is that when the performance of students in Cohorts B and C is examined prior to any Reading First instruction being delivered to Cohort B students (i.e., the beginning of the 2005-2006 school year), there is a performance difference favoring Cohort C. This suggests that Cohort C schools may be slightly higher achieving schools generally than Cohort B schools. Thus, outcome comparisons are complicated by the pre-existing differences in the student populations in these two groups of schools prior to any intervention. Given that Cohort A and B schools are highly comparable in terms of student populations, the population differences between Cohorts B and C are also likely to be relevant in comparisons between Cohorts A and C.

Despite these differences, the performance of students in Cohort A after 3 years of participation in RF is higher than the performance of students in Cohort C (or Cohort B). This conclusion applies to all four grades, and includes analysis of the mean performance score, the percentage of students reaching benchmark goals, and the percentage of students remaining at high risk for reading difficulties.

**Are the outcomes for students in Cohort A who received three years of Reading First (i.e., kindergarten, first, and second grade) better than the outcomes of other groups of students in Cohort A who received less Reading First instruction?**

When the performance of Cohort A students who received 3 years of Reading First instruction was compared to the performance of Cohort A students receiving less than 3 years of Reading First instruction, there is a difference at the beginning of kindergarten, prior to Reading First instruction. Cohort A students who had 3 years of Reading First instruction performed higher on early reading measures than students who had less than 3 years of instruction. Consequently, analyses at the end of the year are complicated by potential differences prior to the onset of Reading First instruction.

Despite this consideration, the benefit of more Reading First instruction is supported by the data. On every measure, students with 3 years of Reading First instruction performed better than students with less Reading First instruction. Mean performance scores were higher, the percentages of students reaching benchmark goals and reading at grade level was higher, and the percentage of students at high risk for reading difficulties was lower.

## Introduction

The purpose of this report is to summarize the impact of Oregon Reading First after three years of implementation (i.e., the end of the 2005-2006 school year). The primary focus is on student reading achievement. Issues regarding implementation are also addressed to provide context for interpreting impact and to consider potential implications that might be drawn from the findings.

As a federal reading initiative, Reading First was to serve as a catalyst for states to implement reading instruction based on scientific research. Reading First was to begin a process of reading reform that districts and schools, operating under state jurisdiction and with state support, would continue after the grant ended. As federal funding for Reading First is reduced or terminated, states are expected to assume more responsibility for supporting districts and schools to implement reading instruction based on scientific evidence. During Reading First, states are expected to systematically expand their capacity to accomplish the major goal of Reading First, which is getting all children to read at grade level or above by the end of third grade. In reading this report, it will be helpful to focus on the following two issues related to the goal of grade level reading performance for all students by the end of third grade.

- *To what extent is Oregon Reading First demonstrating a meaningful impact on student reading achievement?*
- *To what extent should principles of the Schoolwide Beginning Reading Model, which is implemented in Oregon Reading First, serve as a model for reading instruction for other districts and schools in the state of Oregon?*

To address these issues, this report is divided into four major sections.

1. *Section I* provides background on Reading First and previous reform efforts by the federal government to improve student achievement, particularly in reading.
2. *Section II* summarizes Reading First implementation in Oregon. The model schools have used to implement Oregon Reading First is described, as are participating districts and schools, and school personnel. Professional development activities related to the implementation of Oregon Reading First are summarized.
3. *Section III* is the presentation of student reading data. This is the heart of the report. Descriptive data are presented for the purpose of drawing initial conclusions regarding the strength of the impact of Oregon Reading First.
4. *Section IV* summarizes and interprets the impact data on improving student reading achievement in K-3.



## Section I: Background and Context of Oregon Reading First

### A Brief Overview of Oregon Reading First

Oregon Reading First is part of a national, federal reading initiative. Nationally, Reading First is the largest reading initiative ever undertaken by the federal government. Although the scope of Reading First is vast, the goal of the initiative is clear: *The purpose of Reading First is to make sure every child reads at grade level or above by the end of third grade.*

The focus on third grade reading achievement is highly purposeful. The end of third grade is widely viewed as a point where there is a shift in the focus of reading instruction from *building* foundational knowledge and skills (i.e., *learning to read*),<sup>3</sup> to using reading as a tool to *gain* knowledge and understanding about the world (i.e., *reading to learn*).<sup>4</sup> The essential link between these two priorities is strongly supported by evidence demonstrating that if students learn to read successfully in the first few years of school (e.g., by the end of third grade) their chances of doing well academically in school increase substantially, *precisely because they have the reading skills necessary to read independently.*<sup>5</sup> In other words, the reason for the substantial benefit if children exit third grade reading on grade level is that they have the reading skills necessary to read increasingly challenging texts on their own and to learn increasingly complex information contained in these texts.

Reading First is a K-3 initiative, and reading at grade level by the end of third grade is the critical standard by which the effectiveness of the reform will be evaluated. Equally important, however, is successful reading development in kindergarten, first grade, and second grade. As studies have demonstrated,<sup>6</sup> healthy reading development in K-2 is by far *the single best assurance schools and parents have that children are on track to read at grade level by the end of third grade and beyond.* Reading First legislation prioritizes reading achievement throughout K-3 for good reason. Schools are expected to take strong action instructionally when students are reading *below* grade level anywhere along the K-3 continuum.

In each grade K-3, Reading First schools determine whether students are reading at grade level. If so, students are considered to be on track to read at grade level by the end of third grade. When students are reading below grade level, they are considered to be at risk for *not* reading at grade level by the end of third grade. In this case, Reading First schools are expected to provide additional reading instruction, or more intense reading instruction, so that students are more likely to reach grade level reading performance by the end of third grade.

---

<sup>3</sup> See Adams (1990) for a thorough and very readable account of this difference. Although there is no precise point at which learning to read becomes reading to learn, the shift conceptually does represent a useful way of thinking about the *primary* purpose of reading instruction by grade.

<sup>4</sup> Adams (1990); Snow, Burns, & Griffin (1998)

<sup>5</sup> Juel, (1988) study with longitudinal in title. In her classic longitudinal study, Juel found the odds students who were strong readers at the end of grade 1 had about an 88% chance of remaining strong readers at the end of grade 4.

<sup>6</sup> Juel (1988)

## Reading First Represents a Paradigm Shift in Reading Instruction

Reading First is part of the No Child Left Behind (NCLB) Act of 2001.<sup>7</sup> Part B of NCLB appropriated approximately 1 billion dollars annually for Reading First to improve student reading performance in high poverty, underachieving schools.<sup>8</sup> In other words, Reading First is earmarked for those schools throughout the United States serving students who depend most heavily on high quality instruction to learn to read.

Oregon receives approximately 7 million dollars annually to implement Reading First. Federal guidelines targeting the use of Reading First funds are very specific. Throughout the country, the vast majority of Reading First schools use a significant portion of their funds to pay for a full time reading coach. Although Reading First has a significant price tag, hiring a full time reading coach is the only school level personnel expense routinely authorized at the federal, state, and district levels.

Reading First severely restricts the use of funds for personnel expenses for a reason. This reason illustrates a major purpose of Reading First and the degree of change expected under this program. The underlying premise of Reading First is that there is a scientific knowledge base detailing what is required to learn to read in an alphabetic writing system. That is, substantial and convergent research over many decades clearly points to a coherent set of specific instructional practices that are best suited for teaching children to read.<sup>9</sup> This type of instruction is particularly crucial for children most susceptible to reading difficulties, such as students living in high poverty environments, students from minority backgrounds, and students entering school without the knowledge and experiences that are the foundation for a great deal of early reading instruction.

Supporters of Reading First have maintained that if the scientific knowledge base were effectively translated into schoolwide reading practices, then substantially more students would be able to read at grade level. *The central argument is that the primary cause of reading difficulties is that reading instruction in schools is not aligned with the scientific knowledge base.*

The extension of this argument is that the implementation of a research-based reading approach, such as the one required under Reading First, should result in substantially higher reading outcomes, even in schools that are required to adhere to the fiscal constraints that operate in typical school settings. Consequently, Reading First funds cannot be used to hire new teaching faculty because this would be a short-term solution not viable under a normal school budget. Given that funds are not used to hire new staff (except for coaches), as well as the considerable expense of Reading First, it is important to know what the money is used for and why.

---

<sup>7</sup> No Child Left Behind Act of 2001

<sup>8</sup> <http://www.ed.gov/programs/readingfirst/funding.html>

<sup>9</sup> The National Reading Panel (2001) report provides the most authoritative summary of the research on beginning reading beginning with Adams' (1990) highly influential summary, extending to a report by the National Research Council (1998), and culminating in the National Reading Panel report (2001). The National Reading Panel report set the stage for parts of NCLB and outlines the content focus of Reading First.

Beyond hiring a reading coach, Reading First funds are used for two major purposes: (a) to purchase reading programs and other materials for reading instruction, and (b) for professional development and technical assistance for teachers, building and district leaders, coaches, and other personnel who have important responsibilities related to reading. Up to 20% of a state's Reading First budget be allocated to professional development.<sup>10</sup> Schools can also use discretionary funds in their Reading First budgets to purchase additional professional development. Most Reading First schools allocate a significant amount of their discretionary budget to professional development activities related directly to improving reading instruction in their K-3 classrooms.

Significant funding for professional development will be needed if the challenging goal of getting all children to read at grade level by third grade is to be met. Considerable funds are also needed for professional development in Oregon, and in all other states in the country, because the type of reading instruction required by Reading First is substantially different from the reading instruction most teachers practiced prior to Reading First. Reading First instruction is also substantially different from the reading instruction teachers were trained to provide in their teacher training programs.

The change required also extends to coaches, building principals, and district leaders. Coaches, for example, not only have to learn Reading First instruction but they also have to learn how to coach teachers who are not familiar with Reading First methods. Coaches have to deal with resistant teachers, the major changes expected under Reading First, and a whole host of new and demanding job responsibilities.

## Scientifically Based Reading Research

Scientifically-based reading research (SBRR) serves as the cornerstone for reading instruction required in Reading First. This research is described in a number of major syntheses including the National Research Council report called *Preventing Reading Difficulties In Young Children*,<sup>11</sup> and congressionally commissioned reports by Marilyn Adams<sup>12</sup> and the National Reading Panel.<sup>13</sup> These reports are consistent in their conclusions and recommendations regarding the skills and knowledge children need to become successful readers, and the kinds of teaching approaches that are most effective. Transferring this knowledge base into the everyday practice of schools via ongoing, high-quality professional development has been the primary vehicle for improving student reading achievement.<sup>14</sup>

## A Brief History of Reading Reform

Implementing specific reforms to increase student achievement is not new in education. The modern era of school reform was ushered in more than 20 years ago in response to the report, *A Nation At Risk*,<sup>15</sup> which argued the U.S. was in a state of crisis because of the

---

<sup>10</sup> [http://www.ed.gov/nclb/overview/intro/progsum/sum\\_pg3.html](http://www.ed.gov/nclb/overview/intro/progsum/sum_pg3.html)

<sup>11</sup> Snow, Burns, & Griffin (1998)

<sup>12</sup> Adams (1990)

<sup>13</sup> National Reading Panel (2000)

<sup>14</sup> Skindrud & Gersten (2006)

<sup>15</sup> National Commission on Excellence in Education (1983)

deterioration of the education system. School reform that emerged largely in response to this report, in the 1980s and 1990s, was characterized by large-scale, sweeping mandates. Earlier reform efforts were criticized for their piece-meal approach.<sup>16</sup> These new reforms were large and systematic, but ultimately criticized for not building capacity in schools to sustain reform efforts, and for targeting subpopulations of students within schools, rather than targeting achievement improvements for all students.<sup>17</sup> The net effect of these reforms was that student achievement *did not improve*, and teacher behaviors in the classroom *did not change*.<sup>18</sup>

In response to these failures, the current wave of school reform diverges from earlier reforms in two fundamental ways. First, reform emphasis is placed on the entire school, rather than on subpopulations within schools. Second, current reforms are concerned with building capacity within schools to improve student achievement and to sustain impact beyond the concentrated focus of the reform itself.

The current wave of school reform has focused on supporting schools (a) to implement curricular, instructional, and assessment technologies based on reliable research, and (b) by providing ongoing professional development and technical support to improve student outcomes.<sup>19</sup> The following three examples of recent schoolwide reform efforts were supported by the federal government and included major goals in the area of beginning reading instruction:

- The *Comprehensive School Reform Demonstration* Project (CSRD) in 1997
- The *Reading Excellence Act* of 1998 (P.L. 105-277), enacted by amending Title II of the Elementary and Secondary Education Act of 1965 (20 U.S.C. 6601 et seq.)
- No Child Left Behind Act of 2001 (P. L. No. 107-110, 2002), which reauthorized the CSRD (P. L. No. 107-110, Part F, Sec. 1601, 2002) and the *Reading First Legislation* (P. L. No.107-110, Part B, Subpart 1, 2002)

The *Comprehensive School Reform Demonstration* Project (CSRD) was enacted in 1997, and in 1998, over 1,800 schools across all 50 states received funds for implementation, totaling approximately \$145,000,000 dollars. The purpose of the program was to increase student achievement by assisting schools with the implementation of schoolwide reform models that were either based on research or were considered innovative and likely to result in improvements in student achievement. The program targeted low-achieving, high-poverty schools, particularly those receiving Title 1 funds. In No Child Left Behind (NCLB), funds allocated for CSRD increased to \$310,000,000 and the number of recipient schools increased to over 5,300 during the 2001-2002 school year.

NCLB added two important requirements in the reauthorization of CSRD. The first was that school reform models had to provide support for teachers, administrators, and staff in the form of high quality, ongoing professional development and technical support. The purpose of this requirement was to increase school capacity to improve student achievement. The second

---

<sup>16</sup> Hawley (1988)

<sup>17</sup> Fullan (1994)

<sup>18</sup> Desimone (2000)

<sup>19</sup> Office of Educational Research and Improvement (1999)

requirement was that school reform models needed to be based on methods and strategies that were proven to result in improvements in student academic achievement. The earlier clause allowing for schoolwide models that were innovative, but not based on solid science, was eliminated.

Whereas CSRD and NCLB targeted improvement in academic achievement generally, two major pieces of legislation specifically targeted improvements in reading outcomes: the *Reading Excellence Act*, and the *Reading First Initiative* (P.L. 107-110, Part B, Subpart 1, 2002), as part of No Child Left Behind.

The *Reading Excellence Act* had five main objectives:

- Provide children with the readiness skills and support needed in early childhood to learn to read once they enter school
- Teach every child to read by the end of third grade
- Improve the reading skills of students and the instructional practices of teachers and, as appropriate, other instructional staff
- Expand the number of high-quality family literacy programs
- Provide early literacy interventions to reduce the number of children who are inappropriately referred to special education

To accomplish these objectives, the federal government provided over \$260 million dollars to states to support: (a) professional development for teachers, (b) out-of-school tutoring for struggling students, and (c) family literacy programs.

The *Reading First Initiative* sought to replace and enhance the Reading Excellence Act with increased funding to state and local education agencies (LEAs). At over one billion dollars per year, Reading First represents the largest federal investment in reading reform ever. The central idea has been for states and LEAs to assist schools in establishing comprehensive reading programs based on SBRR. A major purpose of Reading First is to train teachers and schools in the essential components of beginning reading. These components are *phonemic awareness*, *phonics*, *fluency*, *vocabulary*, and *comprehension*. Another major purpose of Reading First is to train schools to use *screening*, *diagnostic*, and *instructional reading assessments* to identify children who may be at risk for reading difficulties.

## Determining the Impact of Previous Reforms

Given the extensive history of reading reform as well as the strong current emphasis on accountability in education, assessing the impact of Reading First is a top priority. The central questions are whether Reading First is having an impact on student reading achievement and if so, how large that impact is.

The most commonly used metric to determine magnitude of impact is effect size.<sup>20</sup> Specifically, *Cohen's d* is an effect size test statistic that indicates the difference in the mean outcome for students in a target group compared to the mean outcome for students in a comparison group. Cohen<sup>21</sup> recommends that researchers establish benchmarks to determine impact size within specific areas of study (e.g., large-scale educational interventions).<sup>22</sup> To get a sense of what magnitude of impact might be expected after three years of Reading First implementation in Oregon, it is helpful to examine the impact of other large-scale interventions on reading achievement. Below, we briefly review the impact of CSRD and Title I.

## The Impact of Comprehensive School Reform Demonstration Project

Recently, Borman and his colleagues conducted a meta-analysis on the Comprehensive School Reform Demonstration Project (CSRD) targeting impact on student reading achievement.<sup>23</sup> The meta-analysis summarized the effects of 29 widely implemented reform models. Reform models were categorized based on the following types of evidence:

- *Strongest* evidence of effectiveness
- *Highly promising* evidence of effectiveness
- *Promising* evidence of effectiveness
- *Greatest need* for additional research

The categories were based on the *quality* and *quantity* of the evidence, and *statistical tests* were used (*Cohen's d*) to evaluate the magnitude of impact.

Of the 29 popular reform models, 3 were determined to have the *strongest evidence of effectiveness*: *Direct Instruction*, *Success for All*, and the *School Development Program*. The largest number of studies were on *Direct Instruction*. Forty-nine studies were included, and 38 of these were third-party comparison studies. *Direct Instruction* also had the largest impact on reading achievement, with an overall effect size (Cohen's *d*) of 0.21 ( $Z = 11.61$ ,  $p < .01$ ). A 0.21 effect size means that a student reading at the 50<sup>th</sup> percentile in the comparison group (i.e., *did not* receive the *Direct Instruction* intervention approach) would likely have scored at the 58th percentile if that student had been in the *Direct Instruction* group. At the lower end of reading distribution, a student at the 20<sup>th</sup> percentile in the comparison group (a common designation for a student being at *high risk* for reading difficulties) would have read at the 25th percentile if he or she had been in the *Direct Instruction* group. In Oregon Reading First, where approximately 40% of students are at high risk for reading difficulties when they enter kindergarten, an impact of 0.21 would mean that of 100 students at high risk in the comparison group, 70 of those students would be at high risk if they had received *Direct Instruction*. In other words, 30% of students

---

<sup>20</sup> Cooper & Hedges (1994)

<sup>21</sup> Cohen (1988)

<sup>22</sup> Valentine & Cooper (2004)

<sup>23</sup> Borman, Hewes, Overman, & Brown (2003)

who would have been at the highest level of reading risk would no longer be in that category with an intervention with an effect size of 0.21.<sup>24</sup>

*Success for All* had an overall effect size of 0.18 ( $Z = 16.57$ ,  $p < .01$ ), with a total of 42 studies in the sample of which 25 were third-party comparison group studies. The *School Development Program* had an overall effect size of 0.15 ( $Z = 5.48$ ,  $p < .01$ ). There were 10 studies in the sample and 5 of these were third-party comparison group studies.<sup>25</sup>

One of the most important findings in the meta-analysis by Borman and his colleagues was that 19 of the 29 (66%) most popular and widely used school reform models received the lowest category designation of *greatest need for additional research*. Some examples of these models include *Audrey Cohen*, *Different Ways of Knowing*, and *High Scope*.

Equally important was the negligible impact many of these reform models have had on reading achievement. For example, one study examining *Audrey Cohen* had an overall effect size of  $-0.13$ . The study on *Different Ways of Knowing* resulted in an effect size was  $-0.04$ . Three studies on *High Scope* had an average effect size of  $-0.05$  ( $Z = -1.22$ ).<sup>26</sup> An important conclusion from this meta-analysis is that *there are extensive differences in the number of studies and the magnitude of achievement outcomes between the very few top reform programs and the other reform programs*.

## Important Instructional Variations

In addition to calculating how much impact the different models themselves had on reading outcomes, Borman and his colleagues also carefully examined whether variation in factors that cut across the different models accounted for achievement outcomes.<sup>27</sup> Not only was there considerable variation in reading achievement *among* program models, but there was also considerable variation in reading performance among students *within* each model. This variation is attributable to different factors. If factors associated with this variation that are under the control of schools and teachers can be determined, then we can extend the empirical basis on which instructional decisions are made.

Of particular interest in the analysis by Borman et al. was variation in academic achievement explained by the contexts in which the school reform models were implemented. Variables such as school poverty level and years implementing the reform model were examined. An important finding was that a school's poverty level *was not* a significant predictor of impact. That is, there was no association between school poverty and whether a particular reform model was successful. This is not to suggest there was no association between poverty and achievement. It is likely this association would have been found had it been investigated. What

---

<sup>24</sup> This is a hypothetical example and does not mean these patterns would occur in an intervention that had an impact of 0.21. This example assumes a normal distribution and that impact is equal at lower, middle and upper ends of the distribution. These assumptions must be checked in any particular implementation. It could be, for example, that impact was slightly larger with students at high risk, in which case the number of students at high risk who would benefit would be even larger.

<sup>25</sup> Borman et al. (2003)

<sup>26</sup> Borman et al. (2003)

<sup>27</sup> Borman et al. (2003)

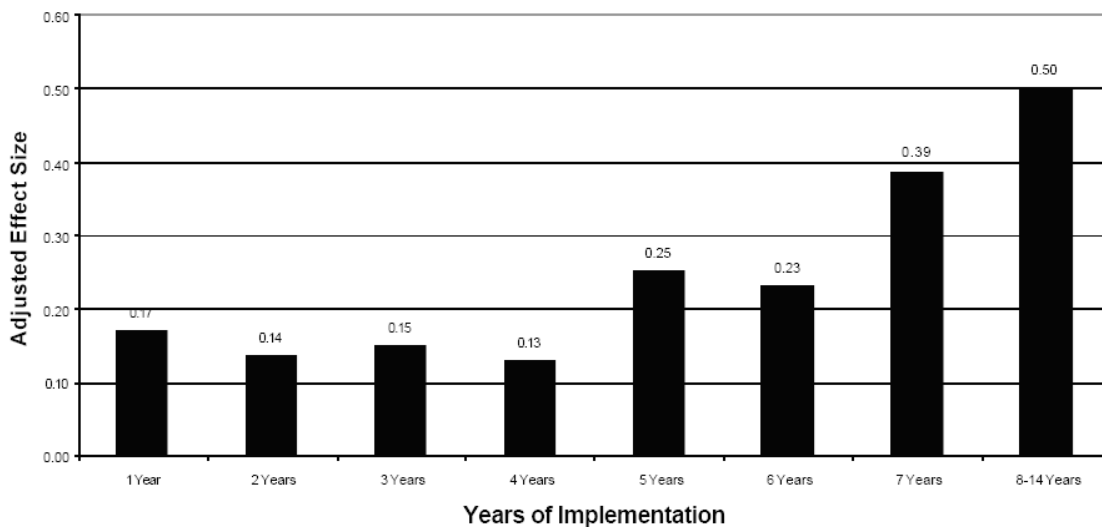
the finding means is that school reform worked equally well in high poverty schools and low poverty schools.<sup>28</sup>

What *was* a significant predictor of impact was *years of implementation*. This finding is extremely important for efforts associated with implementing solid reforms based on science, building school capacity, and sustainability. Figure 1 shows effect sizes based on years of implementation. The average effect size was 0.17 for schools involved in Year 1 of implementation. Overall, effect sizes hover around 0.15 during the first four years of implementation. Then, after 5, 7, and 8-14 years of implementation, average effect sizes increase to 0.25, 0.39, and 0.50, respectively.

From the 5-year point, the magnitude of the impact growth of mature reforms is consistent and impressive. An effect size of 0.50 for large-scale implementation efforts is considered very large. This would mean that in a non-reform school with 100 students at *high risk for reading failure* (e.g., below the 20<sup>th</sup> percentile on a comprehensive measure of reading), *only 40 students would be at high risk if they had been taught in a reform model school. This 60% reduction in the number of students at high risk for reading failure represents an extremely powerful intervention in a large-scale reform effort.*

Given other considerations, an effect size of 0.50 represents the highest empirically based impact we can expect from a large-scale reform effort. We can expect this degree of impact after an extensive implementation period.

**Figure 1. Average Impact of School Reform Efforts in the Comprehensive School Reform Demonstration Project by Year of Implementation (Borman, et al., 2003)<sup>1</sup>**



<sup>28</sup> It must be noted, however, that most of the schools involved in the CSRD tended to be higher poverty schools. Thus, it is possible that if a more expanded range of schools in terms of SES had been involved, more outcome variance might have been attributable to SES.



## Title I Reform

Starting with the late 1960s, long-term trend data from the National Assessment of Educational Progress (NAEP) indicate that U.S. schools have made progress in closing the achievement gap between children from middle class and high poverty backgrounds, and between African American and White children. Much of this progress occurred during the 1970s and 1980s with African American students gaining 2 grade levels on their White counterparts.<sup>29</sup> Some have asserted that these positive trends were due largely to programs like Title 1 and other “War on Poverty” programs first introduced during the mid 1960s.<sup>30</sup>

A meta-analysis by Borman and D’Agostino<sup>31</sup> supports the assertion that Title 1 had a positive impact during this period. They reviewed 17 federal evaluations conducted between 1966 and 1993 and concluded that the strongest improvements in reading and math occurred in the 1970s and 1980s, with very little improvement evident in the late 1960s.

Controlling for methodological differences across the 17 evaluations, Borman and D’Agostino<sup>32</sup> found that the relation between effect size and year of implementation was not consistent. There is a fairly consistent upward trend from 1966 through the early 1980s, from an effect size of 0.0 to approximately 0.15 during that time. These effects essentially leveled off in the 1980s and have remained at about 0.15 through the early 1990s. This is in direct contrast to the continued increase in impact represented in the CSRD analysis (see Figure 1 above).

Borman and D’Agostino suggest that an effect size of about 0.15 might be the best we can expect from a program such as Title I, given the level of federal funding. They also suggest that improvements in program scope and implementation might lead to stronger impact. In other words, stronger Title 1 effects might be attained if implementation procedures were reformed to promote higher achievement rates. One potential way to improve the impact of Title I, as suggested by the mature implementation efforts of the CSRD analysis, is to make sure that schools begin with strong schoolwide reading programs and sustain and improve their use of these programs over time.

## Reading First Reform

Recently, thousands of schools and over 100,000 teachers across the United States have engaged in a highly specified reform to improve the reading outcomes in high-poverty, low-achieving schools. Specifically, No Child Left Behind of 2001<sup>33</sup> established the *Reading First* Initiative (P.L. 107-110, Part B, Subpart 1, 2002), targeting reading reform in K-3.

In Reading First, State Education Agencies (SEAs) fund Local Education Agencies (LEAs) to implement reading programs based on scientifically based reading research (SBRR). Funding in Oregon began in 2002-2003. Thirty-three Cohort A schools began Oregon Reading First implementation in 2003-2004. By the end of 2005-2006, 33 of these schools completed

---

<sup>29</sup> Grissmer, Kirby, Berends, & Williamson (1994)

<sup>30</sup> Borman (2005); Grissmer, Kirby, Berends, & Williamson (1994); Smith & O’Day (1991)

<sup>31</sup> Borman & D’Agostino (1996, 2001)

<sup>32</sup> Borman & D’Agostino (1996, 2001)

<sup>33</sup> No Child Left Behind (2001)

their third year of Oregon Reading First.<sup>34</sup> A second cohort of 17 schools, Cohort B, began implementing Oregon Reading First in 2005-2006.

---

<sup>34</sup> One Reading First school in Cohort A dropped out of the program at the end of the second year.

## Section II: Implementation of Oregon Reading First

An important objective in understanding the impact of school reform on student achievement is to know in some detail what the reform is, beyond its nominal label.<sup>35</sup> In many large-scale evaluations, actual implementation varies from zero or few of the components being implemented to all of the components being implemented.<sup>36</sup> In the 1970s and 1980s, an era of many large-scale evaluations, the importance of assessing the extent to which an educational intervention or approach was actually implemented was stressed in texts and articles. Charters and Jones cautioned that evaluations that did not consider whether the program was actually implemented as intended could essentially be “evaluations of non-events.”<sup>37</sup>

In the context of Oregon Reading First, an important question is “What do Reading First schools do that separates them from non-Reading First schools?” The complexity of this type of question increases in the context of comprehensive reforms such as Reading First because there are many components in the program. How thoroughly each of these components is implemented, and the degree to which these components are used in Reading First schools, and not used in non-Reading First schools, is a challenge to measure precisely and accurately.

In complex reforms, what frequently occurs in considering implementation by target schools is that a thorough description of the reform is provided, and the schools funded to implement the reform are compared to other *comparable* schools *not* funded to implement the reform.<sup>38</sup> Thus, in the next section we describe Oregon Reading First, and suggest that the type of instruction provided in the context of Reading First is very different from the reading instruction provided in non-Reading First schools.

### Major Components of the Federal Reading First Initiative

At the federal level, Reading First is a K-3 reading initiative dedicated to preventing reading problems before they begin and remediating reading problems as soon as they occur. A premise of Reading First is that the vast majority of reading problems students currently experience could be *prevented* if effective instruction were providing early and intensely—that is, throughout the entire K-3 time span.

To provide intense, high quality reading instruction to prevent reading problems, schools must align their instruction with SBRR. The scientific basis for reading instruction, proponents argue, is substantial, convergent, and incontrovertible.<sup>39</sup> SEAs and LEAs are expected to play a

---

<sup>35</sup> Gersten, Baker, & Lloyd (2000)

<sup>36</sup> Miles & Huberman (1994); Willig (1985)

<sup>37</sup> Charters & Jones (1974). This would mean, for example, that if the impact of a new math program were being evaluated, and the math program consisted of 10 major components, none of which were actually implemented, then the impact “evaluation” of the program has not really occurred and the resources and funds that have gone into the evaluation have largely been wasted.

<sup>38</sup> The issue of implementation and implementation quality is one of the reasons the Oregon Reading First Center has worked hard to document that schools are implementing the Reading First program as intended. It provides a way to determine the extent to which schools that are implementing the program paying high attention to fidelity are getting outcomes different from schools that are implementing the program paying low attention to fidelity.

<sup>39</sup> Adams (1990); National Research Council (1998); National Reading Panel (2001)

major role in helping schools develop and implement *comprehensive reading programs* based on SBRR (P. L. No. 107-110, Part B, Subpart A, Sec. 1202 (d)(1), 2002). Comprehensive reading programs comprise the integration of four essential dimensions:

- Curricular materials (e.g., basal or core reading programs)
- Effective teaching practices and strategies
- Assessments with documented reliability and validity
- High-quality professional development to continuously improve instruction quality

### **Oregon Reading First Implementation**

In Oregon Reading First (as well as in many Reading First models used throughout the U.S.) the model used to address these essential components in an integrated fashion and to guide implementation in a comprehensive way is the *Schoolwide Beginning Reading Model*.<sup>40</sup> From its inception prior to Reading First, the developers of the Schoolwide Beginning Reading Model asserted that substantially improving reading instruction and student achievement could be accomplished through an integrated schoolwide effort. The result of this effort should be a system of beginning reading instruction that prioritizes the following essential elements<sup>41</sup>:

- Schoolwide priorities and practices focusing on the essential content in beginning reading development
- Regular use of reliable and valid assessment data to inform instructional practices
- Protected and sufficient time allocated to reading instruction to make sure students reach key reading goals and benchmarks
- High quality implementation of research-based instructional programs
- Differentiated instruction to meet the range of students' instructional needs
- Strong building leadership and administrative involvement based on data, and devoted to sustained effective implementation
- High-quality professional development that drives ongoing efforts to continuously improve the quality of reading instruction and student achievement<sup>42</sup>

These essential elements are fundamentally aligned with Reading First priorities established in NCLB. The specifications of Reading First in NCLB included additional requirements that were easily integrated into the existing framework of the Schoolwide Beginning Reading Model. The resulting Schoolwide Beginning Reading Model served as the foundation for the implementation of Oregon Reading First.

---

<sup>40</sup> Kame'enui, Simmons, & Coyne (2000); Simmons, Kame'enui, Harn, Cole, & Braun (2002)

<sup>41</sup> Kame'enui et al. (2000); Simmons et al. (2002)

<sup>42</sup> Kame'enui et al. (2000); Simmons et al. (2002)

In the next section, we describe how each of the seven elements of the Schoolwide Beginning Reading Model was operationalized in Oregon Reading First.

### **Schoolwide Priorities**

- Oregon Reading First has clearly defined, quantifiable **reading goals** that are linked to the essential elements of reading that guide instructional decisions.

### **Reliable and Valid Assessment Used To Inform Instructional Practices**

- Oregon Reading First uses the Dynamic Indicators of Basic Early Literacy Skills (**DIBELS**) as its primary assessment system for screening students for reading difficulties and monitoring reading progress over time. Oregon Reading First also uses DIBELS benchmark goals to establish important reading milestones, to help group students for instruction, and to inform instruction.
- All DIBELS measures are administered to students by school-based assessment teams. Each assessment team receives substantial and ongoing training on the administration and interpretation of DIBELS measures. In addition, reading coaches at each school conduct additional training sessions including calibration practice involving student participation. To maintain consistency across testers, coaches conduct individual checks with each assessment team member before data collection with students.
- Oregon Reading First also includes the administration of a primary reading measure, the SAT-10, at the end of the year to determine whether students are reading at grade level. This primary reading measure is used by all Reading First schools and is administered to all students in K-2. In third grade, the primary reading measure is the Oregon Statewide Reading Assessment (OSRA).

### **Protected and Sufficient Instructional Time**

- Each Oregon Reading First school dedicates at least 90 minutes to daily reading instruction for all K-3 students, including a minimum of 30 minutes of small-group, teacher-directed reading instruction.

### **Instructional Programs and Materials**

- Each Oregon Reading First school has adopted a research-based **core reading program** for K-3 and focuses on developing high-quality implementation of that program.
- Each Oregon Reading First school is implementing research-based **supplemental programs** that are intended to fill the gaps in the core program and to provide additional instruction and practice on the essential components for those students who need it.

- Each Oregon Reading First school is implementing research-based **intervention programs** to meet the needs of those students who do not benefit from instruction in the core reading program.

### **Differentiated Instruction**

- Each Oregon Reading First school provides additional reading instruction, beyond the 90 minutes, to those students who are not making adequate reading progress.
- Each Oregon Reading First school determines group instruction size, reading instructional time, and instructional programs according to student reading performance and instructional need.
- Each Oregon Reading First school has a schoolwide instructional map in reading for each grade, K-3, for the following groups of students:
  - Students on track for successful reading outcomes (*Benchmark Students*)
  - Students at moderate risk for reading difficulties (*Strategic Students*)
  - Students at high risk for reading difficulties (*Intensive Students*)

### **Data-Based Leadership and Administrative Involvement**

- Each Oregon Reading First school has leadership teams (e.g., Grade Level Teams, Early Reading Team, District Reading First Team) dedicated to analyzing student reading data and using those data to plan instruction. These teams ensure reading instruction is coordinated across grades and services.
- As part of the Schoolwide Beginning Reading Model, each school uses the *Outcomes-Driven Model*<sup>43</sup> to (a) identify students who need additional support and, (b) evaluate each student's response to the instructional support. The goal of the model is to prevent reading difficulties by providing necessary instruction as early as possible. The Outcomes-Driven Model includes the following series of decision-making steps:
  - Identify and validate the student's need for instructional support
  - Plan instructional support
  - Evaluate and modify as necessary the instructional support
  - Review student outcomes

---

<sup>43</sup> Good, Gruba, & Kaminski (2002)

## High Quality Professional Development

Professional development—largely targeting effective administrative support and effective classroom implementation—has been provided by Oregon Reading First to district Reading First leaders, building K-3 classroom teachers, principals, Reading First coaches, and instructional specialists (e.g., Title 1 teachers, special education teachers, English learner teachers, and speech and language specialists).

As part of Oregon Reading First, schools have engaged in several types of professional development activities:

- A major source of professional development for Oregon Reading First schools was the **Institutes on Beginning Reading (IBRs)**. IBRs are for teachers, mentor coaches, principals, and other personnel involved in Reading First implementation. Institute topics targeted assessing reading performance, teaching the five components of beginning reading, providing effective instruction within core and intervention programs, and setting goals to improve student reading outcomes.
- Each school was assigned a **regional coordinator** from the Oregon Reading First Center. Regional coordinators helped schools implement the Schoolwide Beginning Reading Model and build their capacity to implement all of the components of the model on their own. Regional coordinators worked closely with mentor coaches, principals, and district teams.
- Each school used **school-based teams** to oversee the day-to-day implementation of the Schoolwide Beginning Reading Model. Within grade and across grade teams worked together to coordinate systems of support, ensure that instructional programs were being delivered as intended, and that assessment information was guiding decisions about the overall system of reading instruction as well as decisions about individual students.
- Each school had a Reading First **mentor coach** who worked closely with classroom teachers and school-based teams to support effective reading instruction. Mentor coaches guided teams in coordinating instructional support and using data to make decisions. Coaches had three main responsibilities. First, they worked with teachers in the classroom on improving reading instruction. Second, they ensured that the major elements of the Schoolwide Beginning Reading Model were being implemented throughout K-3. Third, they made sure the school used student reading data to make decisions about instruction.
- Each Oregon Reading First school scheduled **ongoing high-quality professional development** to support teachers and instructional staff. This professional development included time for teachers to analyze, plan, and refine instruction.

Table 1 shows the number of hours of professional development provided to Oregon Reading First schools and districts in Cohorts A and B. The table highlights the “team”

orientation of the Schoolwide Beginning Reading Model and the involvement of district and school personnel.

The table does not show all of the professional development devoted to the implementation of Oregon Reading First. The table reflects the minimum hours of required professional development provided by the Oregon Reading First Center. Professional development activities in the table include leadership meetings, coaches meetings, and the Institutes on Beginning Reading. The table does not include optional trainings provided by the Oregon Reading First Center, site-based professional development activities (i.e., schools arranged site-based professional development on effective implementation of their chosen instructional programs and research-based instructional practices), or Core Enhancement Training. The calculated professional development hours reflect six hours of training per professional development day. The state Reading First director provided oversight to make sure additional Reading First funds were used for professional development aligned with Oregon Reading First objectives.

One trend in professional development was the shift over time from off-site professional development in the first years of the grant to more on-site professional development in later years, organized by the coaches. In Table 1, this trend is reflected in the decrease in the number of hours of direct professional development contact provided to schools by the Oregon Reading First Center. For example, in the first year of Oregon Reading First, districts, principals, coaches, and teachers received between 60 and 78 hours of training focusing primarily on the implementation of the Schoolwide Beginning Reading Model. By year three, the districts, principals, coaches and teachers received between 6 and 48 hours of training.

A second trend in the professional development provided by the Oregon Reading First Center shifted from a focus on the entire school team (i.e., administrators, principal, coach, and teachers) to a focus on the coaching role. In the first year of the project, district leaders, principals, coaches, and teachers received similar amounts of professional development provided by the Oregon Reading First Center (i.e., 60 to 78 hours). In year three, the coaches received more hours of professional development (i.e., 48 hours) compared to the other team members (i.e., 6 to 18 hours).



**Table 1. Number of Professional Development Hours Provided by the Oregon Reading First Center for Staff in Cohorts A and B**

	Teachers	Coach	Principal & District RF Team Leader
<i>Cohort A</i>			
Year 1	60	78	72
Year 2	18	66	36
Year 3	6	48	18
Total Years 1 through 3	84	192	126
<i>Cohort B</i>			
Year 1	42	108	60

Another important aspect of Table 1 is the large number of hours on professional development devoted to the Reading First coaches, relative to other positions. The professional development provided by the Oregon Reading First Center shifted from a focus on the entire school team (i.e., administrators, principal, coach, and teachers) in the first year to an increased focus on the coaching role in the years that followed. This intense focus reflects the nature, novelty, and importance of the coaching position. Mentor coaches have played a critical role in Oregon Reading First schools. They have organized and provided refresher trainings for assessments, trained teachers and instructional assistants on effective ways to deliver instruction, modeled new teaching strategies in the classrooms, observed program implementation, and led grade level team meetings that focused on using data to inform instruction. Based on staff needs, they also planned professional development from outside providers on core, supplemental, and intervention programs. Coaches have been responsible for following up on all professional development from the Center and outside providers, and have assisted staff with implementation in classrooms. Mentor coaches worked closely with building principals to ensure successful Reading First implementation.

In short, the primary responsibility of coaches has been to provide staff development, support, and direction. They provided extensive explicit training and feedback to teachers about the implementation of scientifically-based reading programs, instructional strategies, and reading assessments. Prior to implementation, most of the Oregon Reading First schools did not have

reading coaches. Consequently, there was an extensive amount of training required to prepare coaches for their positions.

## Descriptions of the Districts, Schools, and Students

In this impact evaluation, there are two cohorts of Oregon Reading First schools, *Cohort A* and *Cohort B*, and a third group of schools, *non-Reading First comparison schools*, which we refer to as *Cohort C*. Cohort C schools were eligible for Reading First and either did not apply for the program or applied and were not funded. When discussing these cohorts, it is important to note that all three were eligible for Oregon Reading First, which sets them apart in important ways from other elementary schools in the state. Districts were eligible to apply for Oregon Reading First if they had at least one school in the district that met criteria addressing high rates of child poverty and low rates of reading achievement as measured by the Oregon Statewide Reading Assessment (OSRA) in third grade. The Oregon Department of Education awarded Reading First funds to districts, and districts were responsible for monitoring the implementation of Reading First within participating schools.

The implementation schedule of Oregon Reading First for Cohorts A, B, and C is presented in Table 2. Following this table we present descriptions of schools in these three cohorts.

**Table 2. Oregon Reading First Implementation by Cohort and Year**

Cohort	Year 2003-2004	Year 2004-2005	Year 2005-2006
Cohort A, 33 Schools <sup>44</sup>	Implementation Year 1	Implementation Year 2	Implementation Year 3
Cohort B, 17 Schools	<i>No Implementation</i>	<i>No Implementation</i>	Implementation Year 1
Cohort C, 6 Schools	<i>No Implementation</i>	<i>No Implementation</i>	<i>No Implementation</i>

## Number of Districts, Schools, and Students

Table 3 presents information on the number of districts, schools, and students in Cohorts A, B, and C. Information on the number of students is based on the 2004-2005 school year.<sup>45</sup> In Cohort A, Reading First grants were awarded to 14 independent school districts where the

<sup>44</sup> Thirty-four schools started in Cohort A. Prior to the 2005-2006 school year, one school dropped out of the program. Data analysis is based on the 33 schools that completed all three years of Oregon Reading First.

<sup>45</sup> Data for one Cohort B school is not included in this table because the school opened in 2005-2006, the year after data were collected to describe the schools in Cohorts A, B, and C. Consequently, the number of students in Cohort B is underestimated by approximately 300.

schools were located. Approximately half of the schools were located in large urban areas (16 schools), and the remaining schools were approximately equally divided between mid-size cities with populations between 50,000 and 100,000 (8 schools) and rural areas (9 schools).

Cohort B included 17 funded schools in 12 independent school districts located in most regions of the state. Just less than half of the schools were in large urban communities (7 schools); one school was located in a mid-size city with a population between 50,000 and 100,000, and 9 schools were located in rural areas.

Cohort C included 6 schools from 4 school districts. Two schools were located in large urban areas, one in a mid-size city with a population between 50,000 and 100,000, and three in rural areas.

## Demographic Information

Demographic information on participating schools by cohort is presented in Table 4. This table presents averages and ranges for individual schools based on the 2004-2005 school year.<sup>46</sup> The 2004-2005 school year was Cohort A's second year of implementation and it was one year prior to the first year of implementation by Cohort B. The percentages reported in Table 4 are based on information for the entire school, not just K-3. This is the most reliable way to calculate demographic information and we believe it provides a stable estimate for K-3 specifically.

The most important thing about Table 4 is that Cohorts A and B are very similar in terms of key demographic variables (although Cohort A is approximately twice as large as Cohort B). It is particularly important that the percentage of English learners, the percentage of minority students, and the percentage of students on free or reduced lunch (an index of poverty) are very similar in Cohorts A and B. The percentages in Cohort C are also highly comparable in most of these categories. Cohort C schools do seem to have fewer English learners than Cohorts A and B, and a slightly larger African American population. Although not reflected in the table, it is also important to note that the percentages of English learners (especially Cohorts A and B), minority students, and high-poverty students are well above elementary school averages in the state.

It is also important to note the significant variability among schools in Oregon Reading First on demographic variables. This is relevant in terms the Schoolwide Beginning Reading Model because although both this model and Reading First are precise in specifying the components of comprehensive beginning reading programs, an equally important principle is that *one size does not fit all* when it comes to establishing a program for all students. The variables in Table 4 (e.g., mobility, special education, English learners, etc.) should be considered when schools make decisions about their beginning reading program. Schools where 95% of the population is on free or reduced lunch prices, for example, may make different decisions about their program than schools where the poverty rate is less than 50%.

---

<sup>46</sup> 2004-2005 was selected as the year for summarizing demographic information because it represents the most recent available data from the Oregon Department of Education for all participating Oregon Reading First schools.

**Table 3. Number of Districts, Schools, and Students in Cohorts A, B, and C**

	Cohort A <sup>a</sup>			Cohort B <sup>a</sup>			Cohort C <sup>b</sup>		
	Number	Mean <sup>c</sup>	Range <sup>d</sup>	Number	Mean	Range	Number	Mean	Range
Number of Districts	14			12			4		
Number of Schools	33			17			6		
Number of Students K-3	9493	288	153 - 477	4429	277	214 - 429	1481	247	152 - 405
Kindergarten	2429	74	40 - 121	1022	64	37 - 112	387	65	31 - 116
Grade 1	2394	73	36 - 124	1140	71	56 - 93	367	61	30 - 108
Grade 2	2363	72	28 - 113	1147	72	50 - 115	382	64	39 - 95
Grade 3	2307	70	44 - 129	1120	70	50 - 113	345	58	37 - 86

*Note.* Information on the number of students is based on the 2004-2005 school year. <sup>a</sup>Reading First schools. <sup>b</sup>Non-Reading First comparison schools. <sup>c</sup>The mean represents the mean number of students per school or grade. <sup>d</sup>The range indicates the lows and highs across schools.

**Table 4. Demographic Information by Cohort for the 2004-2005 school year**

	Cohort A <sup>a</sup>		Cohort B <sup>a</sup>		Cohort C <sup>b</sup>	
	Mean <sup>c</sup>	Range <sup>d</sup>	Mean	Range	Mean	Range
Percent of Students in Major Categories						
Student Mobility <sup>e</sup>	23.2	10.4 - 32.6	18.1	7.7 - 27.8	NA	NA
Special Education	12.1	1.8 - 19.3	14.9	6.8 - 26.0	14.2	8.23 - 21.03
Free & Reduced lunch	74.8	54.5 - 94.5	68.3	47.2 - 92.0	75.1	47.2 - 87.8
English Language Learners	31.8	0 - 82.6	29.0	0 - 60.3	19.7	3.8 - 47.72
Minority	52.7	17.2 - 97.1	51.3	8.8 - 92.4	50.1	26.4 - 90.2
Percent of Students by Race/Ethnicity						
Black (not Hispanic)	9.2	0 - 67.7	9.6	0 - 69.5	14.7	0.8 - 60.8
American Indian/Alaskan Native	6.1	0 - 95.1	3.2	0 - 9.4	4.4	0.4 - 10.7
Asian/Pacific Islander	5.4	0 - 19.0	7.1	0.3 - 28.6	7.7	0.7 - 20.7
Hispanic	30.5	0.8 - 78.0	29.0	2.9 - 68.2	20.5	12.6 - 28.7
White	45.8	3.1 - 82.3	49.7	6.7 - 91.4	51.5	11.5 - 75.3

*Note.* <sup>a</sup>Reading First schools. <sup>b</sup>Non-Reading First comparison schools. <sup>c</sup>The mean represents the mean percentage of students per school. <sup>d</sup>The range indicates the low and high percentage of students per school. <sup>e</sup>Mobility is based on the number of students taking DIBELS tests in the fall and spring of the 2004-2005 school year. Students who were not assessed at both time periods were defined as mobile. The percentage represents the proportion of mobile students out of all students who participated in DIBELS testing. All other calculations were based on ODE summary statistics as reported on ODE's website.

## Teacher and Leadership Experience

The goal of Oregon Reading First has been to improve student reading achievement by focusing on improving reading instruction through professional development and training. Given the essential role of school personnel, it is useful to consider issues related to the experience and training of the teaching and leadership staff.

Information on the professional experience of Oregon Reading First school personnel was collected for Cohort A in the spring of 2005 by an external evaluation team.<sup>47</sup> Interview and survey data formed the basis of a report on teaching and leadership in Cohort A. Data from Cohort B schools are not available, because they did not participate in the 2005 external evaluation.

The external evaluation included the participation of 34 Oregon Reading First principals, 36 coaches (two schools had an additional half-time coach due to size of building), and 404 K-3 teachers. Table 5 reports the average years of experience for these staff members. On average, school personnel in Cohort A have been in the field of education for over ten years. Years of teaching experience vary considerably, ranging from 0-37 years. Some teachers received their teaching credentials the year prior to working in a Reading First school, and therefore lacked professional experience teaching in the classroom.

**Table 5. Experience of Professional Staff (Cohort A)**

<i>Experience</i>	Principals		Coaches		Teachers	
	Mean	Range	Mean	Range	Mean	Range
Years as a professional educator	20.73	10-35	18.95	6-41	11.69	0-37
Years at current position category	5.77	10-23	1.82	0-6 <sup>a</sup>	8.9	0-35
<i>Highest Level of Education</i>	n	%	n	%	n	%
Bachelors	0	0	10	28	202	50
Masters	33	97	26	72	202	50
Doctorate	1	3	0	0	0	0
Reading Specialist Certification	7	23	20	59	40	10

*Note.* <sup>a</sup>Some coaches reported working in coaching positions prior to Oregon Reading First.

The Oregon Reading First Center also gathered information in the fall of 2006 (after 3 years of implementation) from Cohort A coaches on the number of years that the reading

<sup>47</sup> Edmonds, Lemons, & Roberts (2005)

teachers in their buildings had participated in Oregon Reading First. The purpose was to gauge teacher turnover, with the assumption being that on average the longer that teachers were part of Reading First, the higher the quality of reading instruction they would provide.

Twenty-four coaches provided this information about their schools. In the reported schools, 65% of teachers had taught in Reading First for the duration of the program, 15% had been in the same building for 2 years, and 18% had taught in the building for one year. Of the teachers who taught in the building for less than three years, 14% transferred from another Oregon Reading First school.

## Student Data Used To Determine Impact

The following section describes the three different types of measures used to estimate the reading impact of Oregon Reading First. The first type of measure, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), was used to determine if students met key benchmark goals (indicators) in reading. The second, the Stanford Achievement Test-10<sup>th</sup> Edition (SAT-10) was used to determine if students read at grade level at the end of grades kindergarten, 1, and 2. The third, the Oregon Statewide Reading Assessment (OSRA), was used to determine if students were reading at grade level by the end of third grade.

### Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

The DIBELS measures are indicators of critical beginning reading skills that predict reading success. DIBELS measures are designed to estimate the degree of reading risk of students and to measure performance over time. That is, they are used to screen students for reading problems and monitor reading progress. Screening with all students is done three times per year (this is frequently called benchmarking) to get an indication of each student's overall reading "health." Because the measures are brief, they can be used to efficiently monitor students' reading progress on an ongoing basis. For students on track for reading success, it is recommended that three assessments per year on DIBELS measures are sufficient to monitor whether students are on track for strong reading outcomes. For students who are somewhat below benchmark levels of reading performance, the recommendation is that their progress be monitored once per month, in addition to the benchmark assessments conducted three times per year. For students well below benchmark levels of reading performance—that is, they are at high risk for reading problems—it is recommended that their progress be monitored two times per month, in addition to the benchmark assessments conducted three times per year.

DIBELS measures can be used to evaluate the outcomes of beginning reading instruction at the individual student level (screening and progress monitoring) or at a systems level (e.g., class, grade, school, district, or state). At the systems level, for example, a school can determine the percentage of students who began the year at *moderate risk* for reading difficulties but ended the year *meeting the benchmark reading goal* for being on track for successful reading outcomes. The school can then determine if the percentage of students who reached this higher standard of performance represents a strong or weak system of instruction overall for these students.

In our analysis of impact, we examined performance on two DIBELS measures specifically, *Nonsense Word Fluency* (NWF) at the end of kindergarten, and *Oral Reading*

*Fluency* (ORF) at the end of grades 1, 2, and 3. We used these two measures because they represent the most important DIBELS outcomes at the ends of the four Oregon Reading First grades (i.e., K-3), and they are direct measures of two of the five essential components of beginning reading instruction (phonics and reading fluency).

The DIBELS Data System—which all Oregon Reading First schools use to keep track of how well students are reading and how strong their outcomes are as a school—provides benchmark target goals for performance on NWF at the end of kindergarten, and on ORF at the end of first, second, and third grade. We use the term *benchmark performance goals* for scores and time points indicating if students are *on track for successful reading outcomes* (i.e., they are at *low risk* for future reading problems). Students who score below these goals at specific points in time (e.g., end of first grade) are considered to be at *moderate risk* or *high risk* for reading problems. The degree of risk (i.e., moderate or high) provides a general indication of the intensity of reading instruction students require to reduce the risk they face for reading problems. Generally, students at low risk for reading problems should be able to stay on track for successful reading achievement if they are provided with the school’s *core reading program*, and students at high risk are going to require an *intensive intervention* if they are going to reduce the risk they face and catch up to their grade level peers.

The specific benchmark goals and time points for the DIBELS measures are:

- 25 letter-sound segments read correctly at the end of Kindergarten on Nonsense Word Fluency (NWF)
- 40 words read correctly per minute in grade level material at the end of First Grade on Oral Reading Fluency (ORF)
- 90 words read correctly per minute in grade level material at the end of Second Grade on ORF
- 110 words read correctly per minute in grade level material at the end of Third Grade on ORF

### **DIBELS Nonsense Word Fluency (NWF)**

The DIBELS Nonsense Word Fluency (NWF)<sup>48</sup> measure is a standardized, fluency-based measure of students’ knowledge of the alphabetic principle or phonics. Students are presented with *cv* and *cvc* nonsense words arranged in a random order and asked to read the “words” one at a time. The nonsense word item pool was selected to represent the most frequently occurring letter sounds in the English language.<sup>49</sup> For example, probes include only short vowel sounds, and the letter “c” occurs only in the final position of a word where it always corresponds to the /k/ sound.

Students are directed to provide the sounds of the letters or to read the whole word. For example, students can say the sounds in the word *tob*, /t/ /o/ /b/ or they can read the whole word

<sup>48</sup> Good & Kaminski (2002)

<sup>49</sup> Carnine, Silbert, Kame’enui, & Tarver (2004)



“tob.” Because the measure is fluency-based, students that read the whole nonsense word are generally able to read more letter-sound segments in one minute (and therefore obtain a higher score) than students that sound out each letter.

NWF is administered for the first time to all students in the middle of kindergarten, and for the second time at the end of kindergarten. Based on their performance on this measure and other DIBELS measures, kindergarten students are placed into one of three risk categories. At the end of kindergarten, students reading correctly 25 or more letter-sound segments on NWF are considered to be at *low risk* for reading difficulties (that is, on track for successful reading outcomes). Students scoring between 15 and 25 segments correctly are considered to be at *moderate risk* for reading difficulties, and students below 15 are considered to be at *high risk* for reading difficulties.

The end of kindergarten risk categories are based on the performance of thousands of students participating in the DIBELS Data System in the 2001 – 2002 academic year. The decision rules are described in detail in Good, Kaminski, Simmons, Kame'enui, and Wallin (2002)<sup>50</sup> and complete descriptive statistics are available in Good, Wallin, Simmons, Kame'enui, and Kaminski (2002).<sup>51</sup> As described in Good, Simmons, and Kame'enui (2001),<sup>52</sup> the NWF risk categories at the end of kindergarten are anchored to a score of 50 or more on NWF by the middle of first grade, which represents an important, attainable, and meaningful terminal goal on this type of phonics measure. The goal of 50 is not intended to be the goal for high achieving students, or even the goal for an average student, but rather the goal for the lowest performing student in the middle of first grade in order to be considered on track for successful reading outcomes.

## DIBELS Oral Reading Fluency (ORF)

The DIBELS measure of ORF was developed following procedures used in the development of Curriculum-Based Measurement (CBM).<sup>53</sup> DIBELS ORF measures are one-minute fluency measures that take into account accuracy and speed of reading connected text. The difficulty level of the DIBELS ORF passages was calibrated for grade level difficulty.<sup>54</sup> In the standard administration protocol, students are administered 3 passages at each of three benchmark assessment points during the year (beginning, middle, and end of the year) and the median score at each point is used as the representative performance score. Reliability is very high on this measure, consistently above .90 for alternate-form and test-retest.<sup>55</sup> Criterion-related validity estimates with comprehensive measures of reading performance, including direct measures of reading comprehension, are generally in the .70 to .90 range.<sup>56</sup> The correlation

---

<sup>50</sup> Good, Kaminski, Simmons, Kame'enui, & Wallin (2002)

<sup>51</sup> Good, Wallin, Simmons, Kame'enui, & Kaminski (2002)

<sup>52</sup> Good, Simmons, & Kame'enui (2001)

<sup>53</sup> Deno (1989); Shinn (1989)

<sup>54</sup> Good & Kaminski (2002)

<sup>55</sup> Good & Kaminski (2002)

<sup>56</sup> Marston (1989)

between third-grade DIBELS ORF passages and the Oregon Statewide Reading Assessment was estimated at .67.<sup>57</sup>

The Oral Reading Fluency measure builds on the foundational work of Stan Deno and colleagues at the University of Minnesota Institute for Research on Learning Disabilities who developed CBM procedures.<sup>58</sup> CBM is a set of procedures to assess students in the areas of reading, math, spelling, and written expression. All CBM measures are fluency based and generally take no more than 1-3 minutes to administer. The most common CBM reading measure is oral reading fluency.

The DIBELS ORF passages are distinguished from other CBM reading procedures primarily by the set of generic passages that have been developed for progress monitoring assessments. Passages used to monitor student progress went through a readability analysis to reach comparable levels of difficulty at each grade.<sup>59</sup>

On DIBELS ORF, students reading 40 or more words correctly per minute by the end of first grade are on track to achieve second and third grade literacy goals and consequently are considered to be at *low risk* for reading difficulties. At the end of first grade, students who read between 21 and 39 words correctly in grade level material are at *moderate risk* for reading difficulty and students who read below 20 words correct per minute are at *high risk* for reading difficulty.<sup>60</sup>

### Reading At Grade Level

The most important indicator of Reading First's impact nationally, and of Oregon Reading First's impact specifically, is whether students are reading at grade level at the end of third grade. Of nearly equal importance is whether students are reading at grade level at the end of kindergarten, first, and second grade.

#### **Stanford Achievement Test-Tenth Edition [SAT-10] (Harcourt Educational Measurement, 2002)**

To determine grade level reading performance in kindergarten, first, and second grade, all Oregon Reading First students were administered the entire reading portion of the Stanford Achievement Test-10 (SAT-10)<sup>61</sup> at the end of the year. Grade level reading performance was defined as reading at the 40<sup>th</sup> percentile or above, based on grade level norms. Being at moderate risk for reading difficulties (below grade level) was defined as reading between the 20<sup>th</sup> and 40<sup>th</sup> percentiles. Being at high risk for reading difficulties (well below grade level) was defined as scoring below the 20<sup>th</sup> percentile on the SAT-10.

In May, all students in K-2 were administered the SAT-10, which is group administered. The measure is not timed, although guidelines with flexible time recommendations are given.

---

<sup>57</sup> Baker, Smolkowski, Katz, Fien, Seeley, Kame'enui, et al. (in press); Good, et al. (2001)

<sup>58</sup> Deno (1985); Shinn (1989)

<sup>59</sup> Good & Kaminski (2002)

<sup>60</sup> Good et al. (2001)

<sup>61</sup> Harcourt Brace Educational Measurement (2002)

Test materials are in full color and are considered easy to navigate. Reliability and validity data are strong. Kuder-Richardson reliability coefficients for total reading score were .97 at grade 1 and .95 at grade 2. The correlations between the SAT-10 Total Reading score and the Otis-Lennon School Ability Test<sup>62</sup> ranged from .61 to .74. The edition of the SAT-10 is the most comprehensive and psychometrically studied of any previous edition. Test content is aligned with state and national standards, including standards defined by the National Assessment of Educational Progress (NAEP) and the National Research Council. The normative sample is representative of the U.S. student population.

In kindergarten, the reading subtests of the SAT-10 were administered to all students at the end of the year. The subtests include *Sounds and Letters*, *Word Reading*, and *Sentence Reading* and typically administered in four sessions totaling approximately 120 minutes. The teacher-led *Sounds and Letters* subtest asks children to match words beginning or ending with the same sounds, recognize letters, and match letters with their corresponding sounds. During *Word Reading*, students need to choose the word, or group of words, corresponding to a picture or spoken word. The teacher administers approximately half of the items in this subtest, and the student completes half independently. *Sentence Reading* requires students to read a sentence then choose a picture corresponding to the printed sentence. Eight of these items complete a short story spoken by the examiner. The remaining 21 items are completed independently.

All four of the SAT-10 subtests were administered at first grade: *Word Study Skills*, *Word Reading*, *Sentence Reading*, and *Reading Comprehension*. The entire battery takes approximately 155 minutes to complete. On the *Word Study Skills* subtest, students have to identify compound words, words with similar endings, contractions, and words with a particular sound. *Word Reading* required students to independently select printed words that match a picture. On the *Sentence Reading* subtest, students selected a picture that matched a sentence. Five of the 30 items in this subtest were teacher led and students completed the remaining items independently at their own pace. The more difficult items in the subtest had two sentences. *Reading Comprehension* included items that required students to choose a picture that went with the story and to choose words missing from a story that went with a picture. The final portion of the *Reading Comprehension* subtest required students to choose an answer to a comprehension question about a passage. Six different stories were presented with three items per story.

The second grade version of the SAT-10 included subtests for *Word Study Skills*, *Reading Vocabulary*, and *Reading Comprehension*. The entire test takes approximately 110 minutes to complete. The *Word Study Skills* subtest in the second grade test was similar to that of the first grade test. The reading vocabulary subtest required students to choose the correct definition of the word used in a sentence, identify the sentence in which a target word had the same meaning as it had in a sample sentence, and identify a word that meant the same as the target word used in a sentence. The *Reading Comprehension* subtest required the student to identify correct answers to comprehension questions about a reading selection. This subtest included a total of forty items from nine reading selections across a variety of genres.

---

<sup>62</sup> Harcourt Assessments (2003)

## Oregon Statewide Reading Assessment

Student performance on the Oregon Statewide Reading Assessment (OSRA) was used to determine grade level reading proficiency in third grade. The OSRA is an untimed, multiple-choice test administered yearly to all students in Oregon beginning in third grade. Reading passages representing literary, informative, and practical selections are included in the third grade test. These passages are intended to represent selections that students might encounter in both school settings and in other daily reading activities. The OSRA assesses seven essential reading skills:

- Understanding word meanings in the context of a selection
- Locating information in common resources
- Answering literal comprehension questions
- Answering inferential comprehension questions
- Answering evaluative comprehension questions
- Recognizing common literary forms such as novels, short stories, poetry, and folk tales
- Analyzing the use of literary elements and devices such as plot, setting, personification, and metaphor

Test items are updated regularly. Oregon teachers are trained to write items aligned with the Oregon content standards. Items are reviewed before field-testing by assessment experts for content validity and grade level appropriateness. Items are then field tested and calibrated for difficulty. Tests include items that have passed field-testing, and once students complete a version of the test with new items, assessment specialists analyze student performance to be sure the new items are conforming to result specifications for difficulty and response patterns. The Oregon Department of Education reports that the correlation between OSRA and the California Achievement Test was .75 and the correlation between the OSRA and the Iowa Test of Basic Skills was .78.<sup>63</sup> The four alternate forms used in the OSRA demonstrated an internal consistency reliability (KR-20) of .95,<sup>64</sup> which is high.

## Defining Grade Level Reading Performance on the Oregon Statewide Reading Assessment

A score of 210 was used to define grade level reading proficiency on the Oregon Statewide Reading Assessment (OSRA). This standard was used instead of 201, which is defined as “meets proficiency” according to state criteria, because *meets proficiency* may mean something different from *reading at grade level*. Because *reading at grade level* is the national goal of Reading First outlined in NCLB, we had to determine whether *meets proficiency* was an

---

<sup>63</sup> Oregon Department of Education (2005)

<sup>64</sup> Oregon Department of Education (2000)

acceptable standard for *reading at grade level*. In the first year of Oregon Reading First, a score of 201 (i.e., meets proficiency) corresponded to the 16<sup>th</sup> percentile for all students who were administered the OSRA. We used the 40<sup>th</sup> percentile and above on the SAT-10 in K, 1, and 2, as the standard for grade level reading, and below the 20<sup>th</sup> percentile as the standard for being at *high risk* for reading difficulties. Thus, our SAT-10 standard for *high risk* is roughly comparable to the OSRA standard for *meets proficiency* on the third grade test.

The correlation between the SAT-10 at the end of second grade and the OSRA at the end of third grade is .72 in our Reading First dataset, providing evidence that these two reading assessments provide comparable estimates of overall reading proficiency. Using a definition of 201 as grade level reading in third grade could result in interpretation problems. For example, a student who performed at the 16<sup>th</sup> percentile in grade 2 would be at the highest level of reading risk. If that student made normal reading growth in third grade and scored at the 16<sup>th</sup> percentile on the OSRA, that student would be considered to be reading at grade level (because the 16<sup>th</sup> percentile corresponds to a score of 201). Even though the growth the student made was unremarkable (the performance gap with peers was not reduced), the conclusion in terms of the degree of reading risk the student faced could be easily at odds with a conclusion about unremarkable growth. In terms of risk status, we would say the student went from being at the highest level of risk at the end of grade 2 to reading at grade level at the end of third grade. The strong temptation would be to conclude that third grade provided a powerful reading intervention that worked remarkably well with this student. In fact, the student's actual reading performance remained at the 16<sup>th</sup> percentile compared to grade level peers.

In this hypothetical case, it was the change in the standard used to define grade level reading performance that lead to the potential interpretation problems. Thus, rather than using 201 as the grade level reading standard in third grade, we concluded a better option was to generate a third grade standard for grade level reading performance that was comparable to the consistent standard we used in grades K, 1, and 2 (i.e., the 40<sup>th</sup> percentile). The score on the OSRA in 2003-2004 that was at the 40<sup>th</sup> percentile was 210. Consequently, we used this performance level to determine whether Reading First students *were reading at grade level in third grade*.

## Impact Questions and Analyses

### Overview

Preparing the analytic framework for evaluating the impact of Oregon Reading First involved answering three specific decisions:

1. What comparisons should be made?
2. What measures should be used for the comparisons?
3. What would constitute meaningful impact in terms of student reading achievement?

## **Determining What Comparisons to Make**

Oregon Reading First is designed to have a direct, immediate, and measurable impact on student reading achievement. In theory, the impact of any particular student's participation in Oregon Reading First could be positive, negative, or neutral. We are hypothesizing a positive impact so that on average a student will read at a higher level after receiving instruction in an Oregon Reading First school than that student would have had he or she attended a non-Reading First school.

Obviously, no student can be taught to read in both a Reading First and non-Reading First school simultaneously. Consequently, it was necessary to compare a group of students taught in Reading First schools to a comparable group of students taught in non-Reading First schools, or some variation on this general design.

The best design would have been to randomly assign schools to Reading First and non-Reading First conditions and examine student performance at the end of one year or more. This type of design is not possible in Reading First because Reading First is not the kind of reform that could be offered to some schools and not others on a strictly random basis. Thus, it is necessary to find other comparison groups and comparison standards against which the performance of Reading First students and schools can be analyzed to determine if performance in Oregon Reading First schools is greater than performance we would have expected had they not participated in Reading First.

We believe four types of comparisons represent a valid set of comparisons for the analysis of the performance of Reading First students and schools:

- Performance across years (Cohort A overtime)
- Performance of students in experienced Reading First schools (Cohort A) versus students in less experienced Reading First schools (Cohort B)
- Performance of students in Reading First schools (Cohorts A and B) versus non-Reading First schools (Cohort C)
- Performance of students who have attended Reading First schools over multiple years versus students who have attended Reading First schools for less time

## **Four Ways of Examining Impact**

### **Performance Across Years (Cohort A Over Time)**

One of the success indicators of Oregon Reading First is continuous improvement across years. That is, the federal government uses as an indicator of Reading First impact whether reading achievement is increasing across successive years of implementation. Because we have three complete years of Oregon Reading First data with Cohort A, we are able to evaluate the impact of Reading First over successive years of implementation.

*Consequently, in the following section, we evaluate whether a school that has been in Oregon Reading First for three years is getting increasingly better reading outcomes each year.*

## **Experienced Reading First Schools Versus Inexperienced Reading First Schools**

Two cohorts of schools have been funded in Oregon Reading First. Cohort A began its first year of implementation in 2003-2004 and finished its third year of implementation in 2005-2006. Cohort B began Reading First implementation in 2005-2006. Thus, Cohort B schools completed their first year of implementation the same year that Cohort A schools completed their third year of implementation. Comparing the performance of students in Cohort A to the performance of students in Cohort B allows for a comparison of schools that have attained considerable experience with the implementation of Oregon Reading First with schools that are in the early stages of implementation.

*We evaluate whether relatively experienced Oregon Reading First schools (Cohort A) are obtaining better reading outcomes than relatively inexperienced Oregon Reading First schools (Cohort B).*

## **Reading First Schools Versus Non-Reading First Schools**

The state of Oregon receives funds to implement Reading First according to precise specifications. Each state awards subgrants within the state to districts that identify schools eligible for Reading First. Not all districts or schools eligible for Reading First apply for Reading First funds. In Oregon, for example, 21 districts and 78 schools were eligible for Reading First funds in the initial funding cycle. Fifteen districts and 44 schools applied for funding and 14 districts and 34 schools received funding. Six districts representing 34 schools did not apply for Reading First.

Comparing the performance of students in Oregon Reading First to students in schools that were eligible for Reading First but never participated in the program is potentially advantageous because we know two important things about this group of comparison schools. First, they have relatively high rates of student poverty and second they have relatively low rates of student reading achievement. These two factors determined their eligibility for Oregon Reading First. A potential disadvantage, however, is that funded schools and non-funded schools may be dissimilar in ways that make the comparison problematic. For example, schools that apply for Reading First may place a higher priority on improving reading performance than schools that do not apply. Perhaps a school that does not apply is implementing a schoolwide behavior plan or a new mathematics reform. In this case, the comparison would be influenced by the priority the school places on improving reading outcomes. The comparison between schools that did and did not apply (or applied and did not get funded) can be made, but it is necessary to keep in mind that differences may be influenced by factors other than whether Reading First was implemented, or by apparent differences between the groups of schools on measurable demographic variables such as poverty rates or ethnicity.

*We evaluate whether the performance of students in Oregon Reading First (Cohorts A and B) differs from the performance of students in non-Reading First schools that were eligible for Reading First (Cohort C).*

## Performance of Intact Students Over Time

Many hypothesize that Reading First will have a differential impact on students in relation to length of participation. That is, the impact of Reading First will be largest for students who have attended a Reading First school regularly throughout K-3. Although we cannot easily integrate attendance data into our impact analysis, we can examine students in Cohort A who have participated in Oregon Reading First schools for one, two, or three years. In particular, there is a group of students who began kindergarten in Cohort A in Implementation Year 1 and have completed three full years of Oregon Reading First. That is, these students have received Reading First instruction in kindergarten, first, and second grade. Carefully examining the performance of these students in relation to the performance of other groups of students that have had less exposure to Oregon Reading First instruction can provide information about the value added of Reading First over multiple years of exposure. The potential difficulties with this analysis are similar to the difficulties discussed above involving the analysis of Cohort C schools. Intact students may be dissimilar from non-intact students in ways besides amount of Reading First instruction they receive. One obvious potential difference is mobility. Intact students would appear to be less mobile than non-intact students. Given that achievement is negatively associated with student mobility, it is important to consider potential reading differences between intact and non-intact Reading First students in this light.

*We evaluate whether the performance of students in Cohort A who received three years of Reading First (i.e., kindergarten, first, and second grade) is higher than the performance of students who received fewer years of reading instruction in Oregon Reading First schools.*

## Analysis Measures and Performance Standards

A clear strength of Oregon Reading First is that there is a great deal of high-quality data on student reading performance. In Oregon Reading First, students are assessed on DIBELS, the SAT-10, and the Oregon Statewide Reading Assessment (OSRA). All of these measurement instruments have strong documented reliability and validity. In Oregon Reading First, DIBELS are used to screen students for reading problems, monitor progress, and estimate overall reading proficiency by examining performance in relation to benchmark goals. *As one indication of impact, we will examine student performance in relation to DIBELS benchmark goals.*

We will examine performance on Nonsense Word Fluency (NWF) at the end of kindergarten, and on Oral Reading Fluency (ORF) at the end of grades 1, 2, and 3, on three performance indicators: (a) mean scores, (b) the percentage of students at or above benchmark goals, and (c) the percentage of students at high risk for reading difficulties.<sup>65</sup>

---

<sup>65</sup> In this report, we will not analyze the percentage of students that scored at the *moderate risk* category. The reason for excluding this middle group of students for this part of the analysis (they are included in the mean score analyses) is because it is difficult to say whether these students are on track for reading success or not. Some



Two types of measures were used as primary outcome measures to determine if students were reading at grade level: the SAT-10 and the OSRA. These measures were administered to students at the end of each year. The SAT-10 was administered to students in kindergarten, first, and second grade, and the OSRA was administered to students in third grade.

In summary, we analyzed student performance on reading measures in three ways: (a) overall mean performance score, (b) the percentage of students determined to be reading at grade level at the end of the year, and (c) the percentage of students determined to be at high risk for reading difficulties at the end of the year. As with the DIBELS analysis, we did not analyze the percentage of students who were at *moderate risk* for reading difficulties at the end of the year (but they were included in the mean score analyses).

### Magnitude of Impact

One reason for reviewing the literature on large-scale reading interventions and reform efforts was to determine the impact of previous reforms comparable to the Oregon Reading First. Returning to Figure 1 in Section I, for example, we see that after 1-3 years of implementation an impact of approximately  $d = 0.15$  was typical. In the analysis by Borman and his colleagues,<sup>66</sup> we also see that the two strongest approaches, *Direct Instruction* and *Success for All*, achieved an impact in the 0.18 to 0.21 range. **Thus, in Oregon Reading First, we would interpret an impact in the range of 0.15 to 0.25 to be educationally meaningful.**

In these analyses the student was used as the unit of analysis. That is, all of the students in Cohort A might be compared to all of the students in Cohort B. In calculating an effect size, the two means are subtracted from each other and divided by the pooled standard deviation of scores of students in Cohorts A and B. We also examined performance differences using the school as the unit of analysis. In this case, the mean of each school is calculated. Then the overall mean of school means in Cohort A would be compared to the overall mean of school means in Cohort B. The difference between the means calculated this way will closely approximate the mean difference when students were used as the unit of analysis. What will differ when the school is the unit of analysis is the standard deviation used to divide the difference between the two means. When the school is used as the unit of analysis, the standard deviation between school means is calculated (when the student was the unit of analysis the standard deviation between students' scores was used). The standard deviation of school mean scores is systematically smaller than the standard deviation of students' scores, and consequently, if there is an impact on reading performance at the school level the effect sizes will generally be systematically larger.<sup>67</sup>

---

students in this middle group may achieve later benchmark goals on DIBELS, and others may fall into the high-risk category and may need additional intensive support in order to increase their reading performance.

<sup>66</sup> Borman et al. (2003)

<sup>67</sup> Evaluations of student performance in school settings invariably involve what researchers refer to as hierarchical structures. What they mean is that students are taught in classrooms and classrooms are situated in schools. Both of these factors – classrooms and schools – can exert an influence on achievement that is independent of the particular intervention being studied. The clearest example of this hierarchical impact is teacher effects. Two teachers might deliver the same intervention but get different results with students because of factors unrelated to the intervention. Teacher 1, for example, may be a master teacher with 15 years of experience and Teacher 2 may be in his first year of teaching. Other things being equal, including student factors, we would expect students in the classroom of

The primary focus in this report will be using the student as the unit of analysis. One advantage of this level of analysis is that we will be able to compare our estimates of impact to similar analyses that have been conducted (e.g., the Borman meta-analysis on CSDR and Title I reform). Presenting effect sizes based on the school as the unit of analysis provides an additional way to consider impact. We will present effect sizes with the school as the unit of analysis and in a subsequent report we will conduct formal statistical analyses examining the impact of Oregon Reading First, accounting for student level and school level factors.

---

Teacher 1 to get better outcomes than students in the classroom of Teacher 2. The impact of teacher/classroom factors and school level factors can be determined using statistical procedures such as hierarchical linear modeling (HLM). HLM is particularly powerful when there are data available that can be used to help estimate the potential teacher/classroom level effects and school level effects. In Oregon Reading First, we have data we can use to estimate the influence of factors at the school level. Estimating teacher/classroom level effects is more difficult because students are not always in just one classroom for reading instruction. The school level data we have are still in the process of being coded and analyzed. When the data are ready, we will conduct HLM analyses examining student level and school level factors.

## Section III: Impact on Student Reading Achievement

In this section on impact, we first review data to determine performance comparability between Cohorts A and B. Then we examine the performance of Cohort A over time to determine if performance for these schools improved from Implementation Year 1 to Implementation Year 3. Next, we examine reading difference among three groups of schools: Cohort A, Cohort B, and Cohort C. Cohorts A and B are Reading First schools and Cohort C is non-Reading First schools, but schools that were eligible for Reading First. The focus of these comparisons is on performance in 2005-2006, which was the third year of implementation for Cohort A and the first year of implementation for Cohort B. In the final part of Section III, we examine the performance of a specific group of students in Cohort A. We are interested in the reading outcomes of those students who participated in Reading First during all three years of implementation: In kindergarten, first, and second grade. These students have had the most exposure to Reading First instruction, and it will be useful to determine how much benefit, if any, this additional instructional exposure provided above and beyond the benefit derived by students in the same schools but who received less exposure to Reading First instruction.

### Do Incoming Kindergartners in Cohort A and B Have Comparable Skills?

The data in Table 6 addresses the question of comparability between Cohort A and Cohort B, *prior* to the implementation of Reading First instruction. Conceptually, there are two ways to collect data prior to the implementation of a specific large-scale intervention program such as Oregon Reading First. One way is to present data in years prior to formal implementation (Oregon Reading First implementation began in 2003-2004). The difficulty with this is that not all of the schools were administering DIBELS measures prior to Oregon Reading First. Thus, our comparison would be incomplete. We could rely on OSRA data in third grade, but prior to Reading First, variability among schools in testing policies and procedures on the OSRA makes comparisons problematic. For example, prior to 2003-2004, not all students in third grade were tested on the OSRA. The percentage of students that schools exempted from this test varied considerably prior to 2003-2004. In 2003-2004, criteria for exemptions were tightened considerably.

A second way to examine data prior to Reading First instruction is to examine the performance of all children at the beginning of kindergarten, because at this point they have not received any formal Reading First instruction. After the beginning of kindergarten, children began to receive Reading First instruction and from that point on some degree of their reading achievement was influenced by the instruction they received in Reading First. In previous analyses, we have shown that the percentage of children entering kindergarten who are at risk for reading difficulty provides a valid estimate of the school context in which reading instruction occurs and the degree of challenge the school faces in helping all students become successful readers.<sup>68</sup> For example, given equally effective reading programs generally, it will be easier for

---

<sup>68</sup> There is a consistent negative association between the percentage of students entering kindergarten at high risk for reading difficulties and the student reading performance. In other words, the more students a school has who enter kindergarten at high risk for reading problems the lower the overall reading performance. We have used percent of students entering school at high risk as an estimate of the difficulty level the school faces in reaching high reading outcomes, such as all children reading at grade level by grade 3.

schools with relatively few children entering school at risk for reading difficulties to help students read at grade level by third grade than it will be for schools with a relatively high percentage of students entering at risk.

In Table 6, we present information on the comparability of Cohorts A and B in two ways. In the first row of Table 6 we present the average (mean) scores on Letter Naming Fluency (LNF) for Cohort A schools in Implementation Years 1, 2, and 3, and for Cohort B schools in Implementation Year 1. The reason we present performance on letter knowledge is because this measure has consistently proven to be one of the single best predictor of reading achievement in the early grades.<sup>69</sup> It is a potent predictor of success, even when reading achievement is measured many years into the future.<sup>70</sup>

The reason knowledge of letter names provides such a strong predictor of reading achievement is that it seems to function as an overall proxy for a host of proficiencies and knowledge attributes that are *causally* related to reading achievement. For example, children who enter school knowing letter names are also more likely to have rich vocabularies that make it easier for them to learn to read,<sup>71</sup> and it is also more likely that they have internalized that stories are structured in a predictable way (for example with a beginning, middle, and end),<sup>72</sup> which also makes it easier for them to learn to read. Letter knowledge also makes children more sensitive to the sound structure of language (phonological awareness), which is an essential skill in learning to read. Table 6 shows that *performance on Letter Naming Fluency across years and cohorts is highly similar at the beginning of kindergarten.*

We present additional comparability data between Cohorts A and B in the second row of Table 6 by showing the percentage of students who enter school in kindergarten *at high risk* for reading difficulties. The percentage is based on a composite score consisting of knowledge of phonological awareness and letter names. These two indicators are from the DIBELS measures administered to all K-3 children in Reading First schools at the beginning of the year. Table 6 shows there is a small degree of variability among incoming kindergarten students in Cohort A schools across the three years (the percentage of at risk kindergarten children is a bit lower in Year 2). Overall, however, the percentages are highly similar and there are no discernable differences between Cohort A and Cohort B. In addition to establishing this general comparability, the other important factor in Table 6 is the high percentage of children entering Reading First schools with *very limited early literacy skills that are predictive of reading success.* This variable represents the significant challenge schools face in getting all children to read at grade level in the early grades.

The challenge these schools face should not be minimized. When close to half of the entering kindergarten “class” begins school essentially unable to recognize letters or the sounds in words, it means that the school has to have a very strong system of reading instruction in place. The school must be able to address the significant needs of many children with little literacy knowledge, as well as the needs of children at the opposite end of the spectrum, who enter kindergarten already knowing how to read. The percentage in Table 6 means that 4 out of 10

<sup>69</sup> Baker, Gersten, & Keating (2000)

<sup>70</sup> Snow, Burns, & Griffin, (1998)

<sup>71</sup> Stanovich (1986)

<sup>72</sup> Idol, 1987; Idol & Croll (1987)

students are at the *highest* level of reading risk, even before they have received any public school instruction whatsoever.

In the right hand section of Table 6 we present *effect sizes* and *odds ratios*. These are statistics that help sort out whether comparisons between two quantities are meaningful or trivial. In the case of Table 6, we have comparison scores on a specific measure, Letter Naming Fluency, which is best summarized by an effect size statistic. We also have comparisons of percentages of students in specific categories, which is best summarized as an odds ratio statistic. Effect sizes are typically used with scores on a continuous distribution and odds ratios are typically used with percentages that reflect dichotomous categories such as the percentage of students that read above (or below) grade level.

The effect size comparisons show performance differences of incoming kindergarten students between two years—Cohort A Year 2 compared to Year 1, Year 3 compared to Year 2 etc. In addition, we have Cohort A Year 1 compared to Cohort B Year 1 and Cohort A Year 3 compared to Cohort B Year 1. *In all comparisons of incoming kindergarten students, the effect sizes are extremely small*—very close to 0.0. In other words, differences at this point are trivial, not meaningful. If one of the effect sizes was 0.10, for example, we would talk about a difference of that magnitude as still being relatively small but potentially meaningful depending on a variety of factors. *The effect sizes in Table 6 confirm that entering kindergarten children were comparable between Cohorts A and B and across years within Cohort A.*

In Table 6, we also present odds ratios. Odds ratios show the odds or chances of being in a group or not in that group based on a specific variable, such as location or time or intervention condition. For example, the difference in the odds of being a democrat in California versus New York would be the odds ratio. Similarly, the difference in odds of being “at risk” of reading difficulties in Year 1 of Oregon Reading First versus Year 2 or Oregon Reading First would be the odds ratio. If there was no difference in the odds of being in one group versus another, the odds ratio would be 1.0. For example, if 25% of children in Cohort A were at high risk at the beginning of kindergarten and 25% were at high risk in Cohort B, the odds ratio would be 1.0. In both cohorts the chances that a child would start the kindergarten year at high risk would be one in four. If the odds were higher in Cohort A than Cohort B (in this example), the odds ratio will be greater than 1.0. Odds ratios of approximately 1.20 or higher are said to be meaningful. In Table 6, the odds ratio of being at high risk at the beginning of kindergarten in Cohort A Year 3 are 1.14 times greater than they are of being at high risk at the beginning of Year 2. This would be considered a small difference.

When the odds of children at high risk is lower in the first group than the second, the odds ratio will be less than 1.0. An odds ratio of about 0.85 or less is considered to be meaningful. In Table 6, 4 odds ratios are less than 1.0. Of these, only one is below 0.85 (the odds ratio for Cohort A Year 2 vs. Cohort A Year 1 is 0.79). The odds ratio 0.79 is still relatively small but it shows that the odds of being at high risk in Cohort A Year 2, 36%, was somewhat lower than the odds of being at risk in Cohort A Year 1, 41.5%. This means that in two schools of the same size, if 100 children were a high risk in Cohort A Year 1, 79 would be at high risk in Cohort A Year 2.

**Table 6. Kindergarteners' Early Literacy Knowledge and Percent at High Risk for Reading Difficulties at the Beginning of the Year Across Reading First Cohorts and Years**

Performance Indicator	Cohort A <sup>a</sup>			Cohort B <sup>b</sup>	Effect sizes				
	Year 1	Year 2	Year 3	Year 1					
	<i>Means and Standard Deviations</i>				A2 – A1 <sup>c</sup>	A3 – A2	A3 – A1	A1 – B1	A3 – B1
DIBELS Letter Naming Fluency	7.5 (11.0)	8.0 (11.7)	7.4 (11.1)	7.0 (10.6)	0.04	-0.05	-0.01	0.05	0.04
					<i>Odds Ratios</i>				
<i>Percent at High Risk</i>					A2 v. A1	A3 v. A2	A3 v. A1	A1 v. B1	A3 v. B1
DIBELS Instructional Recommendation <sup>d</sup>	41.5	36.0	39.1	42.5	0.79	1.14	0.91	0.96	0.87

*Note.* <sup>a</sup>For Cohort A, year 1 of implementation took place during the 2003-04 school year; year 2 during the 2004-05 school year; year 3 during the 2005-06 school year. <sup>b</sup>For Cohort B, year 1 of implementation took place during the 2005-06 school year. <sup>c</sup>A1 represents the first year of implementation for Cohort A. B1 represents the first year of implementation for Cohort B etc. <sup>d</sup>The DIBELS Instructional Recommendation in kindergarten is a composite measure of Phonetic Awareness and Alphabetic Knowledge.

## Has the Performance of Students in Cohort A Improved Over Time?

Table 7 shows basic mean performance data of Cohort A students in Reading First in Years 1, 2, and 3. The table shows the amount of change that has occurred in mean scores over time in each grade, K-3, on the grade appropriate DIBELS measure and primary outcome measures. In the right part of the table, the amount of change that has occurred is presented as an *effect size*, which we discuss and interpret below. *This is the impact Reading First has had over time in Cohort A.*

Four types of measures are depicted in Table 7. There are two DIBELS measures, NWF administered in kindergarten and ORF administered in grades 1, 2, and 3. The third measure, the SAT-10, is the primary outcome measure administered at the end of grades K, 1, and 2. The final measure is the Oregon Statewide Reading Assessment (OSRA), which is the primary outcome measure administered at the end of third grade.

The table shows that in each grade and on each measure, mean performance scores increased every year. In kindergarten, for example, the average score of students on NWF at the end of the first year of implementation increased from 28.6 to 31.1 to 39.0 in Years 1, 2, and 3, respectively. On the SAT-10 in grade 2, average scores increased from 581.6 to 585.4 to 589.9 across Years 1, 2, and 3, respectively.

It is important to note that this is cross sectional data, not longitudinal data.<sup>73</sup> If the data were longitudinal it would mean that the same students would be assessed over time. For example, the data presented for grades 1 and 2 would include the same students. With cross sectional data the students tested are not *necessarily* the same students at each assessment point. Some of the students assessed in grade 2 will be the same students assessed in grade 1, but many of the students assessed in grade 2 will not have been assessed in grade 1. In Table 7, a reasonable assumption is that part of the reason children in grade 2 performed better in Year 3 than they did in Year 1 is because of *better* instruction in kindergarten and first grade. However, the purpose of Table 7 is not to present the cumulative value of Reading First instruction across years. This is the difference between a longitudinal analysis and a cross sectional analysis.

However, we *can* get an indirect indication of how much instruction has improved each year in an absolute sense by examining the kindergarten data. In kindergarten there is no prior Reading First instruction to provide a “value added” component from the preceding year. Further, the fact that Table 6 revealed comparability of entering kindergarten cohorts across years increases our confidence that Reading First instruction in kindergarten is improving systematically each year during the first three years.

Regardless of what actually contributes to performance outcomes at the end of each year, the performance scores allow us to determine *overall* whether children in Reading First schools

---

<sup>73</sup> This cross sectional data means that all children were tested at the end of each year and the mean score is presented. It is certainly true that some of the children tested in kindergarten in Year 1 will be tested again as first graders in Year 2 and second graders in Year 3, and these students’ scores are included in Table 7. However, this table does not tell us which children, or how many children, are represented multiple times (that is, longitudinally), or what the performance is of these children specifically. This table shows the average performance of all children tested at the end of each Reading First year.

are reading better each year. Table 7 suggests that reading performance is improving. There are a total of 16 change scores in Table 7—two increases between years (i.e., Year 1 to Year 2 and Year 2 to Year 3) across four grades (i.e., K-3) on two measures per grade (i.e., a DIBELS measure and a primary outcome measure). *On each change score there was a positive increase from one year to the next.* The strength of these increases—in other words, how large these changes were—is indicated by the effect sizes.

For our purposes, we assume an effect size of 0.10 to be small but *potentially* meaningful educationally. An effect size of 0.15 to 0.25 we think of as educationally meaningful, warranting focus and attention. As effect sizes increase above 0.25, the importance of the impact becomes even more apparent. The interpretation of effect sizes is based on the analysis by Borman et al.<sup>74</sup> in their examination of large-scale reforms in reading,<sup>75</sup> and standards described by the Institute for Education Sciences regarding intervention impact. In Table 7, the effect sizes on DIBELS are generally larger than the effect sizes on the primary outcome measures, a finding that is not unexpected. DIBELS measures are collected regularly throughout the year and Reading First schools focus on making instructional adjustments that should result in performance increases on the progress monitoring measures. Because the primary outcome measures are only collected at the end of the year, teachers are not making instructional adjustments on the basis of ongoing performance on these measures.

A related issue is that the primary outcome measures are comprehensive measures of reading with multiple components. Instructional adjustments may have a smaller impact on comprehensive measures of reading than they have on measures of specific skills. Although the impact may be less, there is an important relation between the formative measures used to monitor progress and the primary outcome measures. If students are improving on the progress monitoring measures, the chances are they will also be improving in their overall reading proficiency, as measured by performance on the primary outcome measures.<sup>76</sup>

The third thing to note about the effect sizes is that the largest impact is between Year 3 and Year 1, and all of the effect sizes are meaningful in size. *The evidence indicates that the impact of Reading First across all grades and on all four measures is educationally meaningful.* This is a very important finding in the Year 3 analysis.

A final interesting point in Table 7 is that the smallest effect sizes are in third grade. There are several considerations here. The first is that Reading First may have a smaller impact on third grade than other grades. A rival explanation concerns the fact that after the first three years of implementation, none of the students in third grade had any Reading First instruction in kindergarten. To the degree that kindergarten reading instruction is important in establishing healthy reading development, the lack of a strong program in kindergarten may have reduced the benefit derived in subsequent grades. This hypothesis can be examined more closely at the end of

---

<sup>74</sup> Borman et al. (2003)

<sup>75</sup> The primary effect sizes we report are based on the student as the unit of analysis. This is comparable to what Borman et al. did in examining the impact of large schoolwide reform efforts. We also report effect sizes when the school is used as the unit of analysis, but we limit the comparison to Year 3 vs. Year 1 and do not interpret these effect sizes as small, moderate, or large. These effect sizes will tend to be systematically larger and we present them to offer a second way of considering impact based on effect size.

<sup>76</sup> Baker, et al. (in press)



Year 4, when a cohort of third grade students that has been taught all four years in the Oregon Reading First program, will be assessed.

**Table 7. Performance of Cohort A Schools Over Time on Key DIBELS Measures and on Primary Outcome Measures**

Reading Performance Measure	Implementation Year						Effect Sizes		
	1 (2003 - 04)		2 (2004 - 05)		3 (2005 - 06)				
	Mean	SD	Mean	SD	Mean	SD	Y2 – Y1 <sup>a</sup>	Y3 – Y2	Y3 – Y1 <sup>b</sup>
<i>DIBELS Measures</i>									
Nonsense Word Fluency									
Kindergarten	28.6	19.0	31.1	19.9	39.0	20.9	0.13	0.39**	0.52**/1.50
Oral Reading Fluency									
Grade 1	43.1	32.3	48.0	33.9	52.9	34.1	0.15*	0.14	0.30**/1.15
Grade 2	77.8	39.6	83.8	38.7	90.6	37.6	0.15*	0.18*	0.33**/1.19
Grade 3	98.7	38.2	100.2	37.6	106.6	35.2	0.04	0.18*	0.22*/0.94
<i>Primary Outcome Measures</i>									
SAT-10									
Kindergarten	475.9	44.3	478.6	46.1	487.4	47.9	0.06	0.19*	0.25*/0.70
Grade 1	538.0	44.4	543.8	46.7	549.4	48.1	0.13	0.12	0.25*/0.82
Grade 2	581.6	43.4	585.4	43.1	589.9	41.2	0.09	0.11	0.20*/0.57
Oregon Statewide Assessment									
Grade 3	208.9	11.8	209.9	10.4	210.9	10.1	0.09	0.10	0.18*/0.58

*Note.* <sup>a</sup>Y1 represents the first year of implementation; Y2 the second year; and Y3 the third year. <sup>b</sup>The first effect size is calculated from the mean and standard deviation of the performance of all students in the group. The second effect size is based on the means and standard deviations of school means of all schools in the group. Student-level effects range from 0.15 - 0.25\*, or above 0.25\*\*.

Table 8 presents the percentage of students meeting key benchmark goals on DIBELS and grade level performance goals on the SAT-10 and the OSRA. This addresses a key objective of Reading First: to help all students read at grade level by the end of third grade. Consequently, the most important information in Table 8 is the percentage of students in third grade who are reading at grade level. In Year 1 this was 47.1% of students; in Year 2 it was 51.4% of students; and in Year 3 it was 53.4% of students. A score of 210 or above on the OSRA was needed to reach this goal.

An important pattern in Table 8 is that across measures and years, the percentage of children reaching benchmark goals on DIBELS or reading at grade level increased each year. This would be predicted given that in Table 7 the mean scores also increased systematically each year. However, the translation of mean score increases into higher percentages of students reaching grade level goals is not a certainty. Thus, the across-the-board percentage increases from Year 1 to Year 2 to Year 3 is an important pattern and different from the pattern of mean score increases presented in Table 7.

Table 8 also shows a fairly consistent decrease in the percentage of students reaching benchmark goals and reading at grade level as students move up in grade. In most cases, these decreases are relatively minor but in a couple of cases, they are quite large. In trying to interpret these grade level changes, it is valuable to keep in mind that the measures being used in this determination change in kindergarten to grade 1, from NWF to ORF, and in grade 2 to third grade, from the SAT-10 to OSRA. Thus, in these cases the changes in the percentage of students reaching grade level may be partly attributable to the fact that different measures were used to calculate grade level reading.

It is important to note that the decrease in the percentage of students reaching reading goals as grade level increases does *not* suggest that students are consistently losing ground each year. In examining the data in Table 8 across years, we see that except for moving from kindergarten to grade 1, there are systematic increases in the percentage of students reaching reading goals from one year to the next. For example, in examining the performance of first grade students in Year 1, and second grade students in Year 2, and third grade students in Year 3, we see that there is an increase each year in the percentage of students reaching both benchmark goals on DIBELS and grade level performance on the primary outcome measures.

Part of the reason for the increases by grade across years beginning in grade 1 may be because the reading skills of students are improving above and beyond attaining one year of growth for one year of instruction. A possible explanation for the lack of growth from kindergarten to grade 1 is that the skills being assessed are changing substantially, and perhaps it is easier for students to reach the skills measured in kindergarten versus grade 1. For example, on DIBELS, students are assessed on the alphabetic principle at the end of kindergarten and on reading connected text fluently in grades 1, 2, and 3. On the SAT-10, although the measure nominally remains the same in kindergarten and grade 1, the content of the test changes substantially. There is much more reading required in grades 1 and 2. In kindergarten, the assessment focuses more on reading preskills such as phonological awareness.

In Table 8, we have the odds ratio statistic for children reaching benchmark goals and grade level performance in Year 1 versus Year 2 versus Year 3. Odds ratios greater than 1.0 indicate a positive impact for Oregon Reading First. That is, the odds for reaching the benchmark goals and grade level performance are increasing each year. Odds ratios below 1.0 indicate a negative impact. *All of the odds ratios in Table 8 are above 1.0, indicating a consistent pattern of positive impact of Oregon Reading First in terms of the percentage of children reaching desired levels of reading performance.* In some cases the improvement odds are low. For example, the odds of students reaching the grade level score of 210 on the OSRA are only 1.08 times greater in Year 3 than Year 2, a very modest increase. In other cases the improvements are substantial. For example, the odds of children reaching the NWF benchmark goal in kindergarten are 2.78 times greater in Year 3 than Year 1. In most cases, the odds of reaching benchmark goals and grade level standards are larger than 1.5 times greater in Year 3 than Year 1. This represents substantial improvement. For example, in a hypothetical case involving two schools with the same number of children, if in Year 1, 100 students reached grade level reading, in Year 3, 150 students would reach grade level performance.<sup>77</sup>

---

<sup>77</sup> This is a hypothetical case. It assumes the same number of students in Years 1 and 3.

**Table 8. Percent of Students in Cohort A Reaching Benchmark Goals and Reading at Grade Level Across Years**

Reading Performance Measure	Implementation Year			Odds Ratios		
	1 (2003 - 04)	2 (2004 - 05)	3 (2005 - 06)	Y2 v. Y1	Y3 v. Y2	Y3 v. Y1
<i>DIBELS: Percent Reaching Benchmark Goal</i>						
Nonsense Word Fluency						
Kindergarten	55.6	60.6	77.7	1.23	2.27	2.78
Oral Reading Fluency						
Grade 1	44.6	50.3	59.0	1.26	1.42	1.79
Grade 2	39.9	46.0	54.5	1.28	1.41	1.80
Grade 3	37.9	41.8	50.0	1.18	1.39	1.64
<i>Primary Outcome Measure: Percent Reading at Grade Level (40<sup>th</sup> Percentile)</i>						
SAT-10						
Kindergarten	58.2	60.4	67.9	1.10	1.39	1.52
Grade 1	47.0	52.5	58.0	1.25	1.25	1.56
Grade 2	47.4	50.1	55.3	1.11	1.23	1.37
Oregon Statewide Reading Assessment						
Grade 3	47.1	51.4	53.4	1.18	1.08	1.29
<i>Primary Outcome Measure: Percent Meeting or Exceeding Current ODE Grade 3 Reading Standard (score of 201)</i>						
Grade 3	79.2	83.2	87.7	1.30	1.44	1.87

In Table 9, the percentage of students at high risk for reading problems (substantially below grade level performance) is presented. A major goal of Reading First is to systematically reduce the percentage of students in this category. On the primary outcome measure, these are students who are reading below the 20<sup>th</sup> percentile. Table 9 indicates that *Oregon Reading First schools in Cohort A consistently reduced the percentage of students at the highest level of reading risk*. There was a clear reduction in the percentage of students at high-risk from Year 1 to Year 2, and from Year 2 to Year 3. The reduction of students at high risk was greatest on the DIBELS measures, but the decreases were also impressive on the primary outcome measures.

The percentage drops were remarkably consistent on the primary outcome measure. There was greater variability in the decreases on DIBELS measures, with the largest drops occurring in kindergarten and grade 2, and with consistently higher reductions occurring from Year 2 to Year 3 than from Year 1 to Year 2.

Perhaps the most important thing about Table 9 is that the reductions in the percentage of children moving out of the high risk category was greater than the increases in the percentage of children reading at benchmark or grade level performance (cf. Table 8). On all measures and grades, except the SAT-10 in grade 1, the odds of *not* being in the highest risk category were at least 1.5 times greater in Year 3 than Year 1. This means that in a school with the same number of students across years, if 150 students were at high risk for reading difficulties in a specific grade in Year 1, 100 would be at high risk in that same grade in Year 3.

In summary, the data in Tables 8 and 9 indicate that *Cohort A schools are accomplishing two of the most important Reading First objectives: increasing the percentage of children reading at grade level and decreasing the percentage of children at the highest levels of risk for reading difficulties*.

**Table 9. Percent of Students in Cohort A at the Highest Level of Risk Across Years**

Reading Performance Measure	Implementation Year			Odds Ratios		
	1 (2003 - 04)	2 (2004 - 05)	3 (2005 - 06)	Y2 v. Y1	Y3 v. Y2	Y3 v. Y1
<i>DIBELS: Percent at High Risk</i>						
Nonsense Word Fluency						
Kindergarten	23.0	20.2	8.7	1.18	2.66	3.13
Oral Reading Fluency						
Grade 1	27.6	23.2	17.4	1.26	1.43	1.81
Grade 2	42.7	35.4	27.1	1.36	1.47	2.00
Grade 3	29.3	27.9	20.6	1.07	1.49	1.60
<i>Primary Outcome Measure: Percent Significantly Below Grade Level (&lt;20<sup>th</sup> Percentile)</i>						
SAT-10						
Kindergarten	20.6	18.7	14.3	1.13	1.38	1.55
Grade 1	28.2	24.5	21.4	1.21	1.19	1.44
Grade 2	31.3	26.6	23.1	1.26	1.21	1.52
Oregon Statewide Reading Assessment						
Grade 3	28.1	22.6	18.2	1.34	1.31	1.76
<i>Primary Outcome Measure: Percent Not Meeting the Current ODE Grade 3 Reading Standard (score of 201)</i>						
Grade 3	20.8	16.8	12.3	1.30	1.44	1.87

## How Does the Reading Performance of Students in Cohorts A and B Compare?

In Table 10, the comparison of Cohorts A and Cohort B begins by presenting student reading data after both cohorts completed one year of implementation. In Table 6, we showed that the mean performance of children in Cohorts A and B was highly comparable at the beginning of kindergarten, that is, prior to the implementation of Oregon Reading First instruction. In Table 10, after one year of implementation the mean performance of Cohorts A and B schools remains highly similar. Effect sizes are close to 0.0, meaning that the mean performance scores for both cohorts are about the same, and the small differences that do exist sometimes favor Cohort A and sometimes favor Cohort B. This means that if Cohort A and B are similar prior to the implementation of Oregon Reading First instruction (Table 6), and if they are achieving similar outcomes at the end of their first year of implementation (Table 10), then a reasonable conclusion is that the overall impact or value of the Reading First program can be said to be roughly comparable for the two cohorts of schools after one year of implementation.

The assertion of comparability after one year of implementation is further supported in Table 11, where the odds of reaching benchmark goals or reading at grade level, and not being in the high risk group, are presented for Cohorts A and B. Similar to Table 10, we see a pattern of comparability. Not only are the odds ratios fairly low but also there is no consistent pattern among them. In some cases, the odds slightly favor Cohort A, in other cases they slightly favor Cohort B. The largest odds ratio was students reading at grade level on the OSRA at the end of third grade. In Cohort A the percentage was 47.1% and in Cohort B this percentage was 40.5%.

Table 12 presents the mean score differences between Cohort A after 3 years of implementation and Cohort B after 1 year of implementation. Data for this table is based on performance in 2005-2006. *This is one of the most important tables in the report because it shows the value added of Oregon Reading First after multiple years of implementation.* The strength of the interpretations depend on establishing a sense of comparability between Cohorts A and B *prior to* the implementation of Oregon Reading First (Table 6) and *comparability of performance* at the end of one year of implementation (Tables 10 and 11).

In Table 12, two things stand out. First, there are consistent score differences between Cohort A after 3 years of implementation and Cohort B after 1 year of implementation *favoring* Cohort A. Higher scores for Cohort A occur at all grades on both DIBELS measures and primary outcome measures.

The second noteworthy pattern in Table 12 is that the effect sizes are meaningful in magnitude. *This means that Oregon Reading First is having a meaningful impact on the reading skills of students in K-3.* This impact is largest in kindergarten, and although smaller in grades 1, 2, and 3, the impact is highly meaningful and roughly comparable across grades. In relation to the Borman et al.<sup>78</sup> analysis of large-scale reading reform, the effect sizes in Table 12 are consistently higher than effect sizes presented in the Borman et al. analysis, for implementation lengths less than 5 years.

---

<sup>78</sup> Borman et al. (2003)



**Table 10. Performance of Cohorts A and B at the End of Implementation Year 1 on Key Reading Measures**

Reading Performance Measure	Cohort A		Cohort B		Effect Sizes
	Mean	SD	Mean	SD	Cohort A - Cohort B
<i>DIBELS Measures</i>					
Nonsense Word Fluency					
Kindergarten	28.6	19.0	27.6	18.6	0.05
Oral Reading Fluency					
Grade 1	43.1	32.3	44.0	31.5	-0.03
Grade 2	77.8	39.6	78.7	39.6	-0.02
Grade 3	98.7	38.2	95.5	36.7	0.09
<i>Primary Outcome Measures</i>					
SAT-10					
Kindergarten	475.9	44.3	472.1	45.1	0.09
Grade 1	538.0	44.4	541.5	45.7	-0.08
Grade 2	581.6	43.4	580.1	43.5	0.03
Oregon Statewide Reading Assessment					
Grade 3	208.9	11.8	207.7	10.6	0.11

**Table 11. Students in Cohorts A and B Reaching Reading Goals and at High Risk at the End of Implementation Year 1**

Reading Performance Measure	Percent Reaching Goal			Percent At High Risk		
	Cohort A	Cohort B	Odds Ratio A v. B	Cohort A	Cohort B	Odds Ratio B v. A
<b>DIBELS</b>	<i>Lowest Risk</i>			<i>At High Risk</i>		
Nonsense Word Fluency						
Kindergarten	55.6	54.9	1.03	23.0	25.7	1.16
Oral Reading Fluency						
Grade 1	44.6	47.4	0.89	27.6	25.8	0.91
Grade 2	39.9	41.6	0.93	42.7	40.7	0.92
Grade 3	37.9	34.8	1.14	29.3	30.4	1.05
<b>Primary Outcome Measure</b>	<i>At or Above 40<sup>th</sup> percentile</i>			<i>Below 20<sup>th</sup> percentile</i>		
SAT-10						
Kindergarten	58.2	54.4	1.17	20.6	23.8	1.20
Grade 1	47.0	50.9	0.86	28.2	25.7	0.88
Grade 2	47.4	45.6	1.08	31.3	32.6	1.06
Oregon Statewide Reading Assessment						
Grade 3	47.1	40.5	1.31	28.1	29.7	1.08
	<i>Met Current Standard<sup>a</sup></i>			<i>Did Not Meet Current Standard<sup>a</sup></i>		
Grade 3	79.2	77.8	1.09	20.8	22.2	1.09

*Note.* <sup>a</sup>The current grade three Oregon State Reading Assessment standard is a score of 201.

**Table 12. Performance in Cohort A at the End of Implementation Year 3 and Cohort B at the End of Implementation Year 1**

Reading Performance Measure	Cohort A Year 3		Cohort B Year 1		Effect Sizes <sup>a</sup>
	Mean	SD	Mean	SD	Cohort A Y3 – Cohort B Y1
<i>DIBELS Measures</i>					
Nonsense Word Fluency					
Kindergarten	39.0	20.9	27.6	18.6	0.58**/1.45
Oral Reading Fluency					
Grade 1	52.9	34.1	44.0	31.5	0.27**/0.92
Grade 2	90.6	37.6	78.7	39.6	0.31**/1.08
Grade 3	106.6	35.2	95.5	36.7	0.31**/1.42
<i>Primary Outcome Measure</i>					
SAT-10					
Kindergarten	487.4	47.9	472.1	45.1	0.33**/0.93
Grade 1	549.4	48.1	541.5	45.7	0.17*/0.56
Grade 2	589.9	41.2	580.1	43.5	0.23*/0.74
Oregon Statewide Reading Assessment					
Grade 3	210.9	10.1	207.7	10.6	0.31**/1.18

*Note.* <sup>a</sup>The first effect size is calculated from the mean and standard deviation of the performance of all students in the group. The second effect size is calculated based on the mean and standard deviations of school means of all schools in the group. Student-level effects range from 0.15 - 0.25\*, or above 0.25\*\*.

Also important is the impact of Oregon Reading First on increasing the percentage of children reading at grade level and decreasing the percentage of children at the highest levels of reading risk. The percentage of students reaching benchmark goals and reading at grade level after 3 years of implementation in Cohort A versus 1 year of implementation for Cohort B is presented in Table 13. The percentage of students at the highest level of risk in both cohorts is also provided.

The most important pattern in Table 13 is that in every case a *higher* percentage of children in Cohort A are reading at benchmark and grade level, and a *smaller* percentage of children remain at the highest level of risk. Odds ratios are above 1.0 confirms this pattern. In most cases, the percentage differences are close to or exceed 10%. The differences are meaningful on both the DIBELS measures and the primary outcome measures. In most cases, the odds ratios exceed 1.50, meaning the odds of reaching benchmark or grade level performance goals are 1.5 times greater in Cohort A than Cohort B, and the odds of being at high risk for reading difficulties are 1.5 times greater in Cohort B than Cohort A.

**Table 13. Students Reaching Goals and at High Risk at the End of Year 3 in Cohort A and the End of Year 1 in Cohort B**

Reading Performance Measure	Percent Reaching Goals			Percent At High Risk		
	Cohort A Year 3	Cohort B Year 1	Odds Ratio A3 v. B1	Cohort A Year 3	Cohort B Year 1	Odds Ratio B1 v. A3
<b>DIBELS</b>	<i>Lowest Risk</i>			<i>Highest Risk</i>		
Nonsense Word Fluency						
Kindergarten	77.7	54.9	2.86	8.7	25.7	3.63
Oral Reading Fluency						
Grade 1	59.0	47.4	1.60	17.4	25.8	1.65
Grade 2	54.5	41.6	1.68	27.1	40.7	1.85
Grade 3	50.0	34.8	1.87	20.6	30.4	1.68
<b>Primary Outcome Measure</b>	<i>At or Above 40<sup>th</sup> Percentile</i>			<i>Below 20<sup>th</sup> Percentile</i>		
SAT-10						
Kindergarten	67.9	54.4	1.77	14.3	23.8	1.87
Grade 1	58.0	50.9	1.33	21.4	25.7	1.27
Grade 2	55.3	45.6	1.48	23.1	32.6	1.61
Oregon Statewide Reading Assessment						
Grade 3	53.4	40.5	1.68	18.2	29.7	1.90
	<i>Met Current Standard<sup>a</sup></i>			<i>Did Not Meet Current Standard</i>		
Grade 3	87.7	77.8	2.03	12.3	22.2	2.03

*Note.* <sup>a</sup>The current grade three Oregon State Reading Assessment standard is a score of 201.

## What Is the Performance of Reading First Schools Compared to Non-Reading First Schools?

In Tables 14 – 17 the performance of students in Cohort C is included to determine the impact of Reading First schools compared to non-Reading First schools that were eligible for Reading First but did not participate. Comparability among cohorts is presented in Table 14 and 15, where Cohort B is compared to Cohort C during the 2005-2006 school year. At the beginning of the year, neither Cohort B nor Cohort C had participated in Reading First (Cohort A had completed two full years of implementation and consequently are not included in these tables). Cohort C schools were generally higher performing than Cohort B schools. The effect sizes in Table 14 all favor Cohort C and 3 of the 4 effect sizes are meaningful in magnitude. In Table 15 the percentage of children beginning the year on track for reading success is consistently higher in Cohort C than Cohort B. Three of the four odds ratios are meaningful. In terms of the percentage of children at high risk at the beginning of the year, the percentages are greater in Cohort B than Cohort C in every case, and two of the four odds ratios are meaningful in magnitude.

It is not clear why Cohort C would have systematically higher reading performance than Cohort B, but one explanation seems plausible. There were a total of 22 potential Cohort C schools but 6 were selected because they used DIBELS regularly in their schools. It was necessary to choose Cohort C schools that were using DIBELS so that we would have reading data for the analysis. The fact these schools were using DIBELS data without being required to as part of Reading First may indicate that as a group, practices in these schools may have been more in line with scientifically-based reading research (SBRR) than schools that were not using DIBELS, except when required to by Reading First.

Although we do not have comparability contrasts between Cohorts A and C, given that the starting points between Cohorts A and B were similar (Table 6) and impact after one year of Reading First was also similar (Table 10), it is reasonable to assume that Cohort C schools would have higher reading achievement scores than Cohort A prior to RF implementation. In Tables 16 and 17 we present performance data for Cohorts A, B, and C at the end of the 2005-2006 school year on DIBELS measures.<sup>79</sup> Table 16 shows that in all grades the performance of Cohort A is higher than the performance of Cohort C on the DIBELS measures. In kindergarten and third grade, these effects are clearly meaningful. The higher performance of Cohort A is particularly noteworthy given that Cohort C schools seem to have a higher achieving group of students generally. In Table 16, the performance differences between Cohorts B and C are generally minor, confirming that the impact of Reading First seems to grow with multiple years of implementation.

Table 17 compares Reading First and non-Reading First schools on the percentages of children reaching benchmark reading goals and being at high risk for reading difficulties at the end of the year. All of the comparisons between Cohorts A and C result in meaningful differences favoring Cohort A. Substantially *more* students in Cohort A reach benchmark

---

<sup>79</sup> Note that Cohort C does not administer the SAT-10 in K-2. Once the OSRA data are available for Cohort C we will include this in the analysis.

reading goals at the end of the year, and substantially *fewer* students in Cohort A are at high risk for reading difficulties. Examining the actual percentages helps illustrate the magnitude of this advantage. In comparing Cohorts B and C, we see that half of the odds ratios favor Cohort B and half favor Cohort C. In virtually all cases, the odds ratios indicate the magnitude of the differences is not meaningful.

*In summary, Tables 16 and 17 suggest that the reading skills of students in Cohort A are superior to the reading skills of students in Cohorts B or C.*

**Table 14. Performance of Students in Cohorts B and C on DIBELS Measures at The Beginning of the 2005 – 06 School Year**

	Cohort B Year 1		Cohort C		
Reading Performance Measure	Mean	SD	Mean	SD	Effect Size
Letter Naming Fluency					
Kindergarten	7.0	10.6	10.6	15.5	-0.28
Nonsense Word Fluency					
Grade 1	16.6	19.0	27.1	24.4	-0.48
Oral Reading Fluency					
Grade 2	34.1	30.7	41.9	29.0	-0.26
Grade 3	56.9	33.9	59.2	34.7	-0.06

**Table 15. Percent of Students in Cohort B and Cohort C Reaching Benchmark Goals and at High Risk at the Beginning of the 2005-2006 School Year on DIBELS Measures**

Grade Level	Percent Reaching Benchmark Goal			Percent at High Risk		
	Cohort		Odds Ratios	Cohort		Odds Ratios
	B	C	C v. B	B	C	B v. C
Kindergarten	18.5	22.7	1.29	42.5	39.6	1.13
Grade 1	27.9	47.3	2.32	44.5	24.4	2.48
Grade 2	30.8	40.9	1.55	49.6	33.1	1.99
Grade 3	25.7	26.3	1.03	47.8	47.4	1.02

*Note.* For Cohort B, the 2005-2006 school year represents Reading First Implementation Year 1.



**Table 16. Performance of Students Cohorts A, B, and C on Key DIBELS Measures at the End of the 2005-06 School Year**

Reading Performance Measure	Cohort A, Year 3		Cohort B, Year 1		Cohort C		Effect Sizes	
	Mean	SD	Mean	SD	Mean	SD	A - C	B - C
Nonsense Word Fluency								
Kindergarten	37.22	20.86	26.17	19.17	24.64	19.51	0.62	0.08
Oral Reading Fluency								
Grade 1	50.30	34.56	41.76	31.28	45.72	36.22	0.13	-0.12
Grade 2	87.48	38.73	75.58	40.48	81.25	35.47	0.17	-0.15/0.16 <sup>a</sup>
Grade 3	102.37	38.03	91.99	38.51	91.14	38.50	0.29	0.02/0.09 <sup>a</sup>

*Note.* <sup>a</sup>Effect sizes are adjusted to account for pretest performance differences.<sup>80</sup>

<sup>80</sup> Wortman (1994)

**Table 17. Percent of Students Reaching Benchmark Goals and at High Risk in Cohorts A, B, and C at the End of the 2005 – 06 School Year**

Reading Performance Measure	Percent Reaching Benchmark					Percent at High Risk				
	Cohort			Odds Ratios		Cohort			Odds Ratios	
	A Year 3	B Year 1	C	A v. C	B v. C	A Year 3	B Year 1	C	C v. A	C v. B
Nonsense Word Fluency										
Kindergarten	73.7	50.8	48.5	2.98	1.10	11.4	29.5	32.3	3.71	1.14
Oral Reading Fluency										
Grade 1	55.4	44.6	45.5	1.49	0.96	21.2	29.1	26.3	1.33	0.87
Grade 2	51.9	38.9	41.6	1.51	0.89	29.9	43.8	37.2	1.39	0.76
Grade 3	46.5	32.9	31.6	1.88	1.06	24.6	34.2	36.7	1.78	1.12

## **How Well Do Students in Cohort A Perform If They Have Received Extensive Reading First Instruction**

In Tables 18 – 21, we present information on the reading performance of students in Cohort A who were in Reading First schools during kindergarten, first grade, and second grade. This group of students received the greatest amount of Reading First instruction because they had this instruction for the first three years they attended school. Other students in second grade at the same time received less Reading First instruction because they did not attend the Reading First school in kindergarten through grade 2. They may have entered the school in first grade, for example, and received only two years of Reading First instruction. The performance of these 3-year students in Cohort A was compared to all other students in Cohort A who received Reading First instruction for less than three years.

In Table 18, we see that at the beginning of kindergarten, the 3-year Reading First students scored higher on Letter Naming Fluency and are less likely to be at high risk than students who received less than 3 years of Reading First instruction. This makes the subsequent comparisons more difficult to interpret because the groups are not the same at the beginning of the comparison. The comparison is still useful, however, because it shows the actual performance of students when they receive Reading First instruction for an extended period.

In Table 19, performance at the end of each year for 3-year students is compared to other students with less than 3 years of Reading First instruction. At the end of each year in K-2, the performance of the 3-year group is higher than the performance of the less than 3-year group on both the DIBELS measures and the SAT-10. This shows the potential impact of Reading First for students who are in the program longer. However, the differences can also be partly attributed to performance differences between the two groups of students that existed prior to Reading First instruction.

In Tables 20 and 21 we present the odds ratios for reaching benchmark-reading goals and for being at high risk for reading difficulties. These odds ratios show that students in the 3-year group are more likely to reach benchmark reading goals and are less likely to be in the high risk reading category. These odds ratios are probably too large to be explained solely on the basis of differences between these groups that existed prior to the delivery of any Reading First instruction. However, differences prior to Reading First do make it difficult to determine how much is attributable to Reading First and how much is due to other factors that distinguished the groups prior to Reading First. It is probably true that both more Reading First instruction and other factors are responsible for the large differences in Tables 20 and 21. Even in the context of these considerations, it is important and meaningful that students in Cohort A who had received Reading First instruction for three years read substantially better than students who received less than 3 years of Reading First instruction.

**Table 18. Percent of Students at High Risk at the Beginning of Kindergarten for Students with 3 Years or Less than 3 Years of Reading First Instruction on Knowledge of the Alphabet**

Reading Performance Measure	Length of Participation				Effect Size
	3 Years		Less than 3 Years		
	Mean	SD	Mean	SD	
DIBELS Letter Naming Fluency	9.03	12.19	6.60	10.28	0.22
	Percent at High Risk				Odds Ratio
DIBELS Instructional Recommendation <sup>a</sup>	35.6		45.1		1.49

*Note.* <sup>a</sup>The DIBELS Instructional Recommendation in kindergarten is a composite measure of Phonetic Awareness and Alphabetic Knowledge as defined by Good et al.<sup>81</sup>

<sup>81</sup> Good et al. (2002)

**Table 19. Performance of Students in Reading First for Three years and Students in Reading First for Less than Three Years on Key Reading Measures**

Reading Performance Measure	Length of Participation				Effect Sizes
	3 Years		Less than 3 Years		
	Mean	SD	Mean	SD	
<i>DIBELS Measures</i>					
Nonsense Word Fluency					
Kindergarten	29.55	19.68	24.46	18.91	.26
Oral Reading Fluency					
Grade 1	51.39	33.02	41.14	33.59	.31
Grade 2	93.55	36.14	82.09	40.16	.30
<i>Primary Outcome Measures</i>					
SAT-10					
Kindergarten	479.20	44.28	469.11	43.38	.23
Grade 1	550.06	44.50	536.49	47.54	.29
Grade 2	594.16	39.32	583.19	43.19	.27

**Table 20. Percent Reaching Benchmark Goals and Reading at Grade Level for Students in Reading First for Three Years and Students in Reading First for Less Than Three Years**

Reading Performance Measure	Length of Participation		Odds Ratios
	3 Years	Less than 3 Years	
<i>Percent at Reaching Benchmark Goal</i>			
Nonsense Word Fluency			
Kindergarten	56.7	46.2	1.52
Oral Reading Fluency			
Grade 1	55.3	41.5	1.74
Grade 2	57.4	47.1	1.51
<i>Percent Reading at Grade Level (40<sup>th</sup> Percentile)</i>			
SAT-10			
Kindergarten	62.9	50.8	1.64
Grade 1	59.2	45.5	1.74
Grade 2	59.6	48.7	1.55

**Table 21. Percent at High Risk for Students in Reading First for Three Years and Students in Reading First for Less Than Three Years**

Reading Performance Measure	Length of Participation		Odds Ratios
	3 Years	Less than 3 Years	
<i>Percent at High Risk</i>			
Nonsense Word Fluency			
Kindergarten	21.8	32.9	1.76
Oral Reading fluency			
Grade 1	17.5	33.3	2.35
Grade 2	23.7	35.4	1.76
<i>Percent Significantly Below Grade Level (&lt;20th Percentile)</i>			
SAT-10			
Kindergarten	17.4	27.0	1.76
Grade 1	18.4	31.4	2.03
Grade 2	19.0	29.4	1.78

## Section IV: Implications of the Three Year Findings

### Summary of Findings

#### Background

Three cohorts of schools were included in this evaluation: two cohorts of Oregon Reading First schools, *Cohort A* (33 schools) and *Cohort B* (17 schools), and non-Reading First comparison schools, *Cohort C* (6 schools). A range of measures was used to estimate impact of the Oregon Reading First after three years of implementation. DIBELS measures, which are used by Oregon Reading First schools to screen students for reading problems and monitor reading progress over time, were analyzed to determine the percentage of students who met benchmark-reading goals. Performance on two DIBELS measures were examined—Nonsense Word Fluency at the end of kindergarten, and Oral Reading Fluency at the end of grades 1, 2, and 3. These two measures were selected because they represent the most important DIBELS measures administered at the ends of the four Reading First grades. Furthermore, these DIBELS measures are also predictive of essential reading skills such as word reading proficiency and reading comprehension.

Two additional primary outcome measures were used to determine grade level reading performance. Grade level reading performance in kindergarten, first, and second grade was determined by performance on the Stanford Achievement Test-10 (SAT-10). Grade level performance on this primary outcome measure was defined as reading at or above the 40<sup>th</sup> percentile. High risk for reading difficulties (i.e., well below grade level) was defined as reading below the 20<sup>th</sup> percentile.

Grade level reading performance in third grade was determined by student performance on the Oregon Statewide Reading Assessment (ORSA). A score of 210 was used to define grade level reading on the OSRA rather than the more commonly used score of 201. A score of 210 was selected because it corresponded to the 40<sup>th</sup> percentile in the first year of Oregon Reading First, and thus, was comparable to the SAT-10 standard. In contrast, a score of 201 corresponded to the 16<sup>th</sup> percentile. Across these measures, the analysis targeted mean performance scores, the percentage of students reading at benchmark and grade level goals, and the percentage of students at high risk for reading difficulties.

The following four questions were the major focus of the report.

- Are Cohort A schools getting increasingly better reading outcomes each year of implementation?
- Are experienced Oregon Reading First schools (Cohort A) getting better reading outcomes than inexperienced Oregon Reading First schools (Cohort B)?
- Is the performance of students in Oregon Reading First (Cohorts A and B) better than the performance of students in non-Reading First comparison schools that were eligible for Reading First (Cohort C)?



- Are the outcomes for students in Cohort A who received three years of Reading First (i.e., kindergarten, first, and second grade) better than the outcomes of students in Cohort A who received less than 3 years of Reading First instruction?

Research on large-scale reading reform was used to anchor interpretations regarding the magnitude of impact of the Oregon Reading First. Effect sizes in the 0.15 to 0.25 range were considered to be educationally meaningful.

## Results

The analysis was clear regarding the performance of Cohort A schools across years. In each grade and on every measure, mean performance scores increased systematically each year. Out of a total of 16 change scores for Cohort A—two increases between years (i.e., Year 1 to Year 2 and Year 2 to Year 3) across four grades (K-3) on two measures per grade (i.e., a DIBELS measure and a primary outcome measure)—there was a positive increase on all 16 scores. All of the effect sizes comparing Year 3 to Year 1 are educationally meaningful.

In addition, across all measures and years, the percentage of children reaching benchmark or grade level goals increased systematically over time, and the percentage of children remaining at a high level of reading risk decreased systematically. Thus, in addition to increases in mean performance, Cohort A schools are accomplishing two of the most important Reading First objectives: (a) Cohort A schools are *systematically increasing* the percentage of children reading at grade level and (b) they are *systematically decreasing* the percentage of children at the highest levels of risk for reading difficulties.

*The analysis of the performance of Cohort A across years indicates that in all grades and on all reading measures the impact of Oregon Reading First has been meaningful over the course of the reform thus far.*

In comparing the performance of Cohort A to Cohort B after one year of implementation, reading outcomes were highly similar. Effect sizes were close to 0.0 and the small differences that did exist sometimes favored Cohort A and sometimes favored Cohort B. At the end of three years of implementation in Cohort A schools (i.e., performance in 2005-2006) and one year of implementation in Cohort B schools (i.e., 2005-2006), *differences were meaningful and consistently favored Cohort A*. Higher scores for Cohort A occurred in all grades on both DIBELS measures and primary outcome measures. Effect sizes were consistently meaningful in magnitude. *This pattern suggests that Oregon Reading First is having a meaningful impact on the reading skills of students in K-3. The comparison between Cohort A after 3 years of implementation and Cohort B after 1 year of implementation is the best evidence of the value added of Oregon Reading First after increased years of implementation.*

In comparing Cohort A with Cohort B, impact was largest in kindergarten. The impact was somewhat smaller in grades 1, 2, and 3, but still meaningful in these grades and roughly comparable across grades. In terms of the impact Oregon Reading First has had on the percentage of children reading at grade level and the percentage of children remaining at the highest level of reading risk, the outcomes were clear. In every comparison, a *higher* percentage of children in Cohort A than Cohort B were reading at benchmark and grade level, and a *lower*

percentage of children were at the highest level of reading risk. In most cases, the odds of reaching benchmark or grade level were more than 1.5 times greater in Cohort A than Cohort B, and the odds of being at high risk for reading difficulties were more than 1.5 times greater in Cohort B than Cohort A.

Analysis and interpretation involving Cohort C schools (non-Reading First schools that were eligible for Reading First) was more complex. In examining the performance of students in Cohorts B and C prior to any Reading First instruction being delivered to Cohort B students (i.e., the beginning of the 2005-2006 school year), differences favored Cohort C. This suggests that Cohort C schools may have been slightly higher achieving schools generally than Cohort B schools (for example, they may serve a less at risk student population). Thus, outcome comparisons were complicated by potential differences in the student populations in these two groups of schools prior to the delivery of Reading First instruction.

Given that Cohorts A and B appeared to have highly comparable student populations—reading skills of entering students are highly similar, and the percentages of students attending these schools who are English learners, high-poverty students, and students from minority backgrounds are roughly the same—the student population differences between Cohorts B and C were also likely to be relevant in comparisons between Cohorts A and C.

Despite these differences, it is still informative to compare the performance of students in Cohorts A, B, and C. Similar to comparisons between Cohorts A and B, when Cohort C is compared to Cohort A, the performance of students in Cohort A was superior to the performance of students in Cohort C. This pattern was true at all four grades, and it was true in terms of mean performance scores, the percentage of students reaching benchmark and grade level goals, and the percentage of students remaining at high risk for reading difficulties.

In the final comparison, the performance of Cohort A students who received 3 years of Reading First instruction (in kindergarten, first grade, and second grade) was compared to the performance of students in Cohort A who received less than 3 years of Reading First instruction. Similar to the challenge faced in the comparison with Cohort C, it appears that at the beginning of Reading First, Cohort A students who received 3 years of instruction were different from Cohort A students who received less than three years of Reading First instruction. At the beginning of kindergarten, prior to Reading First instruction, the 3-year students performed higher on an early reading measure than students who received less than 3 years of Reading First instruction. Consequently, interpreting performance differences after the implementation of Reading First instruction was complicated by the apparent differences prior to the onset of Reading First instruction.

Despite this consideration, the benefit of more Reading First instruction is supported by the data. On every measure and on every indication of impact, students with 3 years of Reading First instruction performed better than students with less Reading First instruction. Mean performance scores were higher, the percentages of students reaching benchmark goals and reading at grade level was higher, and the percentage of students at high risk for reading difficulties was lower.

## References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, 35(4), 494-519.
- Baker, S., Smolkowski, K., Katz, R., Fien, H., Seeley, J., Kame'enui, E., et al. (in press). Reading fluency as a predictor of reading proficiency in low performing high poverty schools: Going to scale with Reading First.
- Borman, G. D. (2005). Efforts to bring reform to scale in high-poverty schools: Outcomes and implications. In L. Parker (Ed.), *Review of Research in Education* (pp. 1-28). Washington, DC: American Educational Research Association.
- Borman, G. D., & D'Agostino, J. V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*, 18(4), 309-326.
- Borman, G. D., & D'Agostino, J. V. (2001). Title I and student achievement: A quantitative synthesis. In G. D. Borman, S. C. Stringfield & R. Slavin (Eds.), *Title I: Compensatory education at the crossroads* (pp. 25-57). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and student achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125-230.
- Carnine, D., Silbert, J., Kame'enui, E., & Tarver, S. (2004). *Direct instruction reading* (4 ed.). Upper Saddle River, NJ: Pearson.
- Charters, W. W., Jr., & Jones, J. E. (1974, February). *On neglect of the independent variable in program evaluation*. Paper presented at the Project MITT, Eugene, OR.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Cooper, H., & Hedges, L. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219-232.
- Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York, NY: Guilford.

- Desimone, L. (2000). *The role of teachers in urban school reform. ERIC/CUE Digest Number 154*. (No. EDO-UD-00-3). New York: ERIC Clearinghouse on Urban Education, New York, NY.
- Edmonds, M., Lemons, C., & Roberts, G. (2005). *Oregon Reading First evaluation: Final report*. Austin, Tx: University of Texas.
- Fullan, M. (1994). *Change forces: Probing the depths of educational reform. School development and the management of change series: 10*. Pennsylvania: Falmer Press, Taylor & Francis Inc.
- Gersten, R., Baker, S., & Lloyd, J. W. (2000). Designing high-quality research in special education: Group experimental design. *Journal of Special Education*, 34(1), 2-18.
- Good, R. H., Gruba, J., & Kaminski, R. (2002). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. IV, pp. 679-700). Washington, DC.: National Association of School Psychologist.
- Good, R. H., Kaminski, R., Simmons, D., Kame enui, E. J., & Wallin, J. U. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene, OR.: University of Oregon.
- Good, R. H., & Kaminski, R. a. (2002). Nonsense word fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic Indicators of Basic early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.
- Good, R. H., Simmons, D. C., & Kame'enuei, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5(3), 257-288.
- Good, R. H., Wallin, J. U., Simmons, D., Kame enui, E. J., & Kaminski, R. (2002). *System-wide percentile ranks for DIBELS benchmark assessment* (Technical Report No. 9). Eugene, OR.: University of Oregon.
- Grissmer, D. W., Kirby, S. N., Berends, M., & Williamson, S. (1994). *Student achievement and the changing American family*. Santa Monica, CA: RAND.
- Harcourt Assessment Inc. (2003). *Otis-Lennon School Ability Test [OLSAT]* (8th ed.). San Antonio, TX.
- Harcourt Educational Measurement. (2002). *Stanford Achievement Test [SAT-10]*. San Antonio, TX.
- Hawley, W. D. (1988). Missing pieces of the educational reform agenda: Or, why the first and second waves. *Educational Administration Quarterly*, 24(4), 416-437.

- Idol, L. (1987). Group story mapping: A comprehension strategy for both skilled and unskilled readers. *Journal of Learning Disabilities*, 20, 196-205.
- Idol, L. & Croll, V. J. (1987). Story-mapping training as a means of improving reading comprehension. *Learning Disability Quarterly*, 10, 214-229.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80, 437-447.
- Kame'enui, E. J., Simmons, D. C., & Coyne, M. D. (2000). Part II: Intervention issues: Schools as host environments: Toward a schoolwide reading improvement model. *Annals of Dyslexia*, 50, 31-51.
- Marston, D. (1989). Curriculum-based measurement: What is it and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford Press.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Beverly Hills, CA: Sage.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education*. District of Columbia: The Commission.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: National Institute of Child Health and Human Development.
- National Reading Panel. (2001). *Put reading first: The research building blocks for teaching children to read. Kindergarten through grade 3*. District of Columbia: National Inst. for Literacy, Washington, DC.
- National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- No Child Left Behind Act of 2001. (PL 107-110). *Title I, part B, student reading skills improving grants, subpart I, Reading First*.
- Office of Educational Research and Improvement. (1999, Spring). From at-risk to excellence [Electronic Version], Vol. 1 from <http://www.ed.gov/offices/OERI/At-Risk/newsletter/spring99/index.html>.
- Oregon State Department of Education. (2000). *Report cards - school and district*. Salem, OR: Oregon Department of Education.
- Oregon State Department of Education. (2005). *Closing the achievement gap: Oregon's plan for Success for All Students*. Salem, OR: Oregon State Department of Education.

- Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.
- Simmons, D., Kame'enui, E., Harn, B., Cole, C., & Braun, D. (2002). Building, implementing, and sustaining a beginning reading improvement model: Lessons learned school by school. In M. R. Shinn, H. M. Walker & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 537-570). Washington, DC: National Association of School Psychologists.
- Skindrud, K., & Gersten, R. (2006). An evaluation of two contrasting approaches for improving reading achievement in a large urban district. *Elementary School Journal*, 106(5), 389-407.
- Smith, M., & O'Day, J. (1991). *Putting the pieces together: Systemic School Reform*. New Brunswick, NJ: Eagleton Institute of Politics.
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(3), 360-406.
- Valentine, J. C., & Cooper, H. (2004). *What works clearinghouse study design and implementation assessment device (Version 1.1)*. Washington, DC: US Department of Education.
- Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, 55(3), 269-318.
- Wortman, P. M. (1994). Judging research quality. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York, NY: SAGE.