

# COLORADO READ TO ACHIEVE Data Summary Report for 2010–2011



October 2011

# **COLORADO READ TO ACHIEVE**

## **Data Summary Report for 2010–2011**

Ann E. Davis, M.Ed.

October 2011



101 SW Main Street, Suite 500  
Portland, Oregon 97204  
[www.educationnorthwest.org](http://www.educationnorthwest.org)

## EXECUTIVE SUMMARY

The 2010–2011 school year marked the first year of the fourth cohort of Colorado Read to Achieve students. A total of 47 schools enrolling almost 2,000 students in kindergarten through third grade participated in the 2010–2011 Colorado Read to Achieve Program. Of the 47 schools, 28 of them (60%) were part of one of this year’s four consortia. Schools implemented the program in kindergarten through third grade or in a configuration of these grade levels. Almost all of schools implemented the program in second and third grades.

Overall, 64 percent of matched students who were enrolled in Colorado’s Read to Achieve Program met or exceeded the program benchmark goals by spring 2011. The vast majority of kindergarten students (95%) achieved their benchmark program goal, while 84 percent of first-grade students did. Second- and third-grade students lagged behind the younger students. Slightly less than one-half of them reached their program goals—49 percent for second grade and 44 percent for third grade.

The percentage of students in different demographic subgroups, e.g., ethnicity, gender, English language learners, and special education, generally clustered around the overall percentage of 64 percent reaching the program benchmark goal. The exceptions were the American Indian/Native American group which surpassed the overall percentage by 11 percentage points and black and Asian/Pacific Islander students and special education students who fell from five to seven percentage points below the 64 percent. It should be noted that the American Indian/Native American subgroup contained only 11 students.

Two indicators of program success are how well the program helped intensive and strategic students to progress in their reading and how well the program kept benchmark students at benchmark. From fall 2010 to spring 2011, Read to Achieve was extremely successful in maintaining almost all benchmark students across time in kindergarten and first grade, but slightly less successful in second and third grades. Also, schools with kindergarten and first-grade programs succeeded in moving almost all of their intensive students to either the strategic or benchmark levels. Unfortunately, over 60 percent of second-grade, intensive students remained in intensive while 40 percent of third-grade students in the intensive group stayed behind. Overall, schools helped their kindergarten and first-grade students to become successful readers, but their second- and third-grade students continued to struggle by the end of the year and showed little improvement.

Very modest correlations between the Oral Reading Fluency (ORF) measure of the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) and the reading test of the Colorado Student Assessment Program (CSAP) were found in the third grade. Scoring at benchmark on the ORF is not a particularly good predictor of meeting proficiency on the CSAP.

# CONTENTS

	<b>Page</b>
Executive Summary .....	iii
Contents .....	iv
Figures .....	v
Tables .....	vi
Introduction.....	1
Methods .....	1
Results .....	4

# FIGURES

	<b>Page</b>
Figure 1. Kindergarten—Percentage of Students at Each Level on the PSF (N=264) .....	6
Figure 2. Grade 1—Percentage of Students at Each Level, by DIBELS Measure and Overall Program Goal (N=461) .....	6
Figure 3. Grade 2—Percentage of Students at Each Level on the ORF (N=543) .....	7
Figure 4. Grade 3—Percentage of Students at Each Level on the ORF (N=488) .....	8
Figure 5. Scatterplot of Spring 2011 CSAP Scaled Scores and DIBELS ORF Scores .....	14

# TABLES

	<b>Page</b>
Table 1. DIBELS Measures Administered at Which Testing Intervals, by Grade Level .....	2
Table 2. DIBELS Scores Needed to Meet the Fall and Spring Program Benchmark Goals, by Grade Level .....	3
Table 3. Percentage (n) of Students in Each Matching Group, by Grade Level – 2010–2011.....	3
Table 4. Number of Students <i>Not</i> Matched on All Measures at Testing Points, by Grade Level – 2010–2011.....	4
Table 5. Number and Percentage of Schools with Program Implemented, by Grade Level .....	4
Table 6. Number and Percentage of Schools Implementing Program in Different Grade Configurations.....	5
Table 7. Percentage of Students in Project for Full Cycle and for Less Than Full Cycle, by Grade Level (N=1,901).....	5
Table 8. Overall Percentage of Matched Students at Program Benchmark Goal in Spring 2011, by Demographics, Grades K-3 Combined .....	9
Table 9. Kindergarten—Changes in ISR Status from Winter 2010 to Spring 2011.....	10
Table 10. Grade 1—Changes in ISR Status from Fall/Winter 2010 to Spring 2011, by Measure.....	11
Table 11. Grade 2 and Grade 3—Changes in ISR Status from Fall/Winter 2010 to Spring 2011 .....	12
Table 12. Relationship between Grade 3 ORF and CSAP Performance Levels, Spring 2011 (N=487).....	13

# INTRODUCTION

The Colorado Department of Education contracted Education Northwest to analyze the program's 2010–2011 data and to submit a Data Summary Report of the results as part of its external evaluation. This report summarizes the assessment results from the fall/winter 2010 and spring 2011 administrations of three measures of the *Dynamic Indicators of Basic Early Literacy Skills*, or DIBELS. In addition, Education Northwest analyzed the results of the reading test of the Colorado Student Assessment Program (CSAP) of third-grade, Read to Achieve students.

## Overview of Read to Achieve Program

In 2010–2011, the fourth cohort of schools implemented its first year of the Read to Achieve Program. The first cohort cycle was from 2000–2004; the second cohort cycle was from 2004–2007; and the third cohort cycle was from 2007–2010. Each school in cohort four screened students for eligibility into the Read to Achieve program. All schools were required to use the DIBELS to select students for the Read to Achieve intervention. In addition to the DIBELS, some schools used a variety of other assessments such as the Direct Reading Assessment 2 (DRA2). The Read to Achieve Program was designed for “strategic,” but schools could enroll low “benchmark” and high “intensive” students.

State implementation guidelines are quite flexible. The structure of the program, intensity, and curriculum varied widely from school to school depending on what each school wrote in their proposals. This year the majority of schools used Leveled Literacy Intervention (LLI), a new program from Fountes and Pinnell which required small groups to be about 3 students. To increase the number of students receiving services, some schools decided to implement two groups of students—one group from fall to winter and the other group from winter to spring.

Schools administered the DIBELS in the fall for screening as well as for the fall benchmark test, progress monitoring, and for the winter and spring benchmark tests. At schools where the LLI curriculum was implemented, students who participated in winter to spring intervention group first took the DIBELS in the winter. The personnel involved in assessment administration varied by school. Depending on the school, paraprofessionals, teachers and/or Read to Achieve consultants might have given the tests.

## METHODS

### Data Collection

Depending on the grade level, Read to Achieve schools administered different measures of the DIBELS in fall, winter, and spring of 2010–2011. While the DIBELS is administered three times a year, the focus of the data analyses is on the fall 2010 and spring 2011 DIBELS assessment results. In those schools implementing the LLI curriculum, the analyses included those students with fall to spring or winter to spring scores.

The measures given included the Phoneme Segmentation Fluency (PSF) test, the Nonsense Word Fluency (NWF) test, and the Oral Reading Fluency (ORF) test. These three measures do not represent all grade-level measures of the DIBELS. In kindergarten and first grade, DIBELS has a total of four different measures.

The time of the year determines which of these measures are administered. In kindergarten, the Read to Achieve Program administers two measures, the PSF and NWF. As designed by DIBELS developers, these measures are first administered in the winter rather than in the fall. The program also administered these same two measures in first grade, but three times a year. The DIBELS for the second grade consist of the PSF in the fall and the ORF three times a year. Third grade students only take the ORF in the fall, winter, and spring. Table 1 shows when each measure was administered.

**Table 1**  
**DIBELS Measures Administered at Which Testing Intervals, by Grade Level**

Grade	PSF	NWF	ORF
Kindergarten	Winter and Spring	Winter and Spring	--
Grade 1	Fall, Winter, and Spring	Fall, Winter, and Spring	Winter and Spring
Grade 2	--	Fall	Fall, Winter, Spring
Grade 3	--	--	Fall, Winter, Spring

Both the DIBELS and the CSAP test were administered by classroom teachers. At some schools, the DIBELS measures were administered by an assessment team rather than the classroom teacher. The DIBELS measures were given in the fall, winter, and spring, while the CSAP was administered to third-grade students once in February 2011. After the administration of the assessments, school staff members entered DIBELS scores into either the online DIBELS database, maintained by the University of Oregon, MClass, or the Colorado Department of Education website data collection tool at the end of the year. Next year, all schools will enter their DIBELS data into the University of Oregon database on an ongoing basis. Education Northwest received a file of all students' scores from the Read to Achieve Program. Each record had the students' identification number, demographic information, and all DIBELS scores and corresponding status levels. For third-grade students, Education Northwest also received scaled scores and proficiency levels for the CSAP reading test.

## Calculation of Risk Levels on Measures and Program Benchmark Goals

The scores on the individual DIBELS measures fall into one of the three levels of risk—"at risk," "some risk," and "low risk." These levels are represented as the "intensive," "strategic," and "at benchmark" levels in this report.

The Read to Achieve Program developed program benchmark goals based on the individual DIBELS measures or a combination of measures. The following table, Table 2, displays the scores needed to achieve the fall and spring benchmark goals for each grade level.



**Table 2**  
**DIBELS Scores Needed to Meet the Fall, Winter and Spring Program Benchmark Goals, by Grade Level**

Grade	Fall Criteria			Winter Criteria*			Spring Criteria		
	PSF	NWF	ORF	PSF	NWF	ORF	PSF	NWF	ORF
Kindergarten	--	--	--	18	--	--	35-	--	--
Grade 1	35	24	--	35	50	--	35	50	--
Grade 2	--	--	44	--	--	68	--	--	90
Grade 3	--	--	77	--	--	92	--	--	110

\* Used with students with only winter and spring scores.

In first grade, students need to meet the criteria on both the PSF and the NWF in order to reach the fall and spring program benchmark goal. The focus of this report is on the individual DIBELS measures and whether students reached the program benchmark goals.

### Matching Students

To conduct the data analyses presented in this report, students were “matched.” “Matching” means that students were only included if they had DIBELS scores for the testing interval of the analysis. Students with both their fall and spring scores were matched as were students who participated in the winter to spring intervention group and had both their winter and spring scores. Students who did not have matched scores were excluded from the analysis. For first grade, students needed to have scores on both the PSF and NWF in the fall/winter and spring to be included in the analysis. Similarly, in the analyses that investigated the movement of students in each status level from fall/winter to spring, only students with matched scores were included. Table 3 shows the number of matched students in each matching group.

**Table 3**  
**Percentage (n) of Students in Each Matching Group, by Grade Level—2010–2011**

Grade	Percentage (n) of Students Matched Fall to Spring	Percentage (n) of Students Matched Winter to Spring	Total Percentage (n) of Students Matched (Fall to Spring and Winter to Spring)
Kindergarten	16% (264)	--	15%(264)
Grade 1	25% (415)	44% (46)	26% (461)
Grade 2	30% (499)	42% (44)	31% (543)
Grade 3	29% (473)	14% (15)	28% (488)
<b>TOTAL</b>	<b>1,651</b>	<b>105</b>	<b>1,756</b>

### Missing Data

The database included a total of 1,901 students. The type of analysis determined the number of missing student scores. When matching on two testing intervals, there will be students with missing data. Some students might have their fall scores but not spring scores, while other students might not have fall scores, but have spring scores. Overall, 8 percent of the students had missing data. First grade had the highest percentage of missing data (12%) because students

needed both the PSF and NWF at each testing period. In the third grade correlation study, students were included only if they had both their DIBELS and CSAP scores. Twelve percent of them had missing data. Table 4 shows the percentage and number of missing cases at each grade level.

**Table 4**  
**Number of Students Not Matched at Testing Points, by Grade Level—2010–2011**

Grade	Percentage (n) of Students Not Matched	Total N
Kindergarten	4% (12)	276
Grade 1	12% (60)	521
Grade 2	5% (29)	572
Grade 3	8% (44)	532
<b>OVERALL</b>	<b>8% (145)</b>	<b>1,901</b>
Grade 3 (CSAP & DIBELS )	12% (61)	532

## Data Analyses

Data analysis consisted of calculating percentages of students meeting the program benchmark goal and performing at each of the three status levels on the DIBELS measures. Since these data were matched, each set of percentages represents absolute increases or declines for the 2010–2011 cohort of students included in the analysis. The data were disaggregated by grade level and demographics, and the movement of students from the fall/winter 2010 to the spring 2011 was calculated. Due to rounding off, percentages might not always add up to 100 percent. Also, a Chi-square test and correlation analyses were performed on the third-grade DIBELS scores/risk levels and the CSAP scaled scores/performance levels. These analyses explored the existence of a relationship between the two measures, and the strength of that relationship.

## RESULTS

A total of 47 schools enrolling almost 2,000 students in kindergarten through third grade participated in the 2010–2011 Colorado Read to Achieve Program. Of the 47 schools, 28 of them (60%) were part of one of the four consortia. Almost all of the schools implemented the program in second and third grades. Slightly less than half of the schools (47%) had a kindergarten program. Table 5 displays these results.

**Table 5**  
**Number and Percentage of Schools With Program Implemented, by Grade Level**

Grade	Percentage(n) of Schools With Program Implemented (N=47 schools)
Kindergarten	47% (22)
Grade 1	91% (43)
Grade 2	98% (46)
Grade 3	96% (45)

In 2010–2011, 20 schools (43%) implemented the Read to Achieve Program in all four grades—kindergarten through third grade. Almost half of the schools (49%) implemented the program in three grade levels. Except for one school, schools implemented the program in first through third grades. At the other eight percent of schools, the program was implemented in one or two grade levels. Table 6 summarizes this data.

**Table 6**  
**Number and Percentage of Schools Implementing Program in Different Grade Configurations**

Grades Program Implemented	Percentage (n) of Schools (N=47 schools)
All grades	43% (20)
3 grade levels	49% (23)
2 grade levels	6% (3)
1 grade level	2% (1)

A total of 1,901 students participated in the Read to Achieve Program in 2010–2011. Across grade levels, 92 percent of the students participated in the program for the entire year with kindergarten students participating at the highest rate of 96 percent. Table 7 shows student participation rates by grade level.

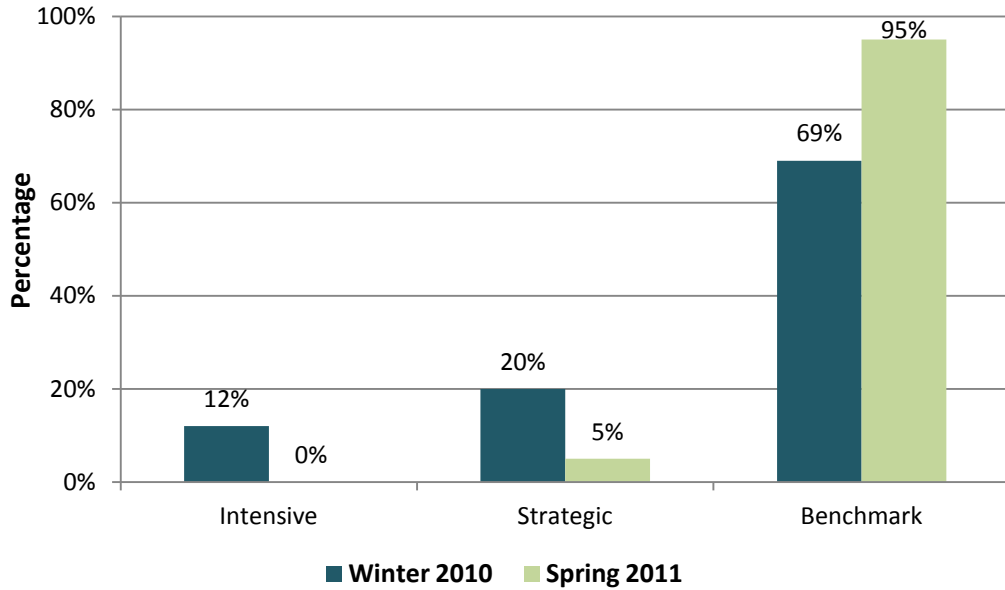
**Table 7**  
**Percentage of Students in Project for Full Cycle and for Less Than Full Cycle, by Grade Level (N=1,901)**

Grade	Percentage		N
	Full Cycle	Less Than Full Cycle	
Kindergarten	96%	4%	276
Grade 1	91%	9%	520
Grade 2	92%	8%	569
Grade 3	90%	10%	531
<b>TOTAL</b>	<b>92%</b>	<b>8%</b>	<b>1896</b>

## Overall Student Performance by Grade Level

The trend for kindergarten students on the PSF from winter 2010 to spring 2011 represents the trend for a successful program—the percentage of intensive and strategic students declined as the percentage of benchmark students increased. The percentage of intensive students declined from 12 percent to 0 percent! In the strategic group, only 5 percent of the kindergarten students remained. The vast majority of students (95%) reached benchmark on the PSF, the program benchmark goal, by spring 2011. Figure 1 compares the performance of kindergarten students on the PSF from winter 2010 to spring 2011.

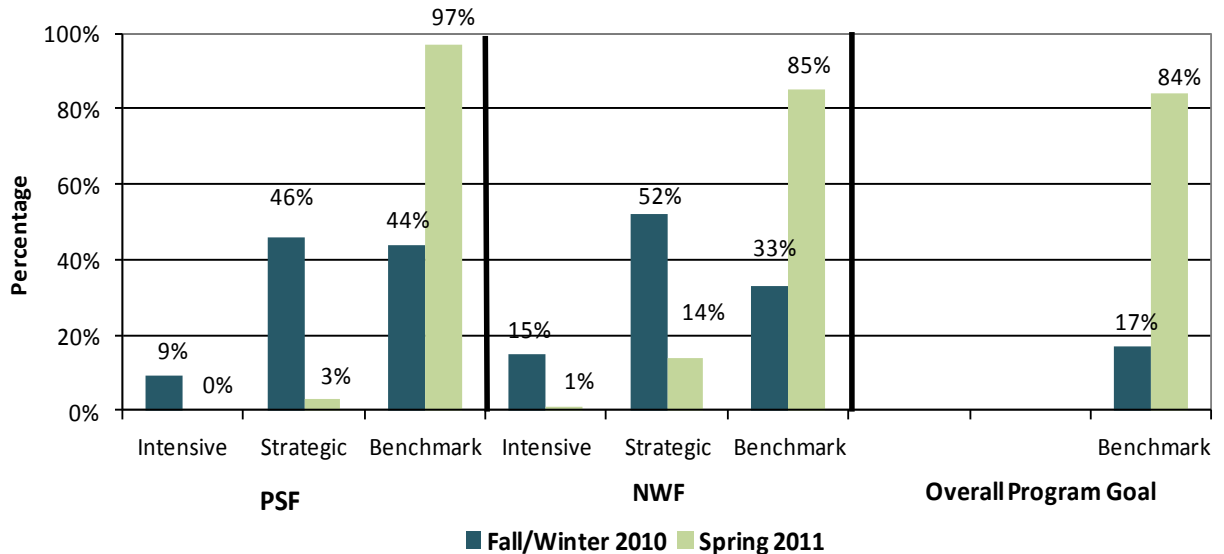
**Figure 1**



**Kindergarten—Percentage of Students at Each Level on the PSF (N=264)**

Similar to kindergarten, first grade results also dramatically demonstrated the desired trends for a successful program. By the end of the year, 97 percent and 85 percent of the students mastered benchmark on the PSF and NWF respectively, and 84 percent of first-grade students reached the program benchmark goal. Figure 2 displays the performance of first-grade students on the PSF, NWF, and overall program goal.

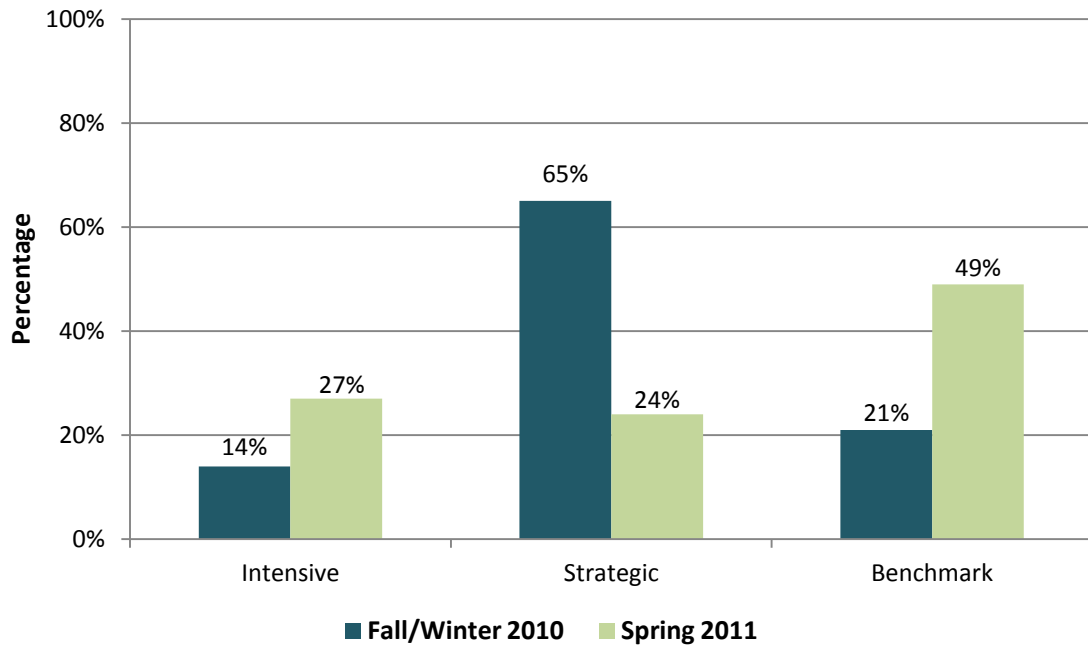
**Figure 2**



**Grade 1—Percentage of Students at Each Level, by DIBELS Measure and Overall Program Goal (N=461)**

Second grade results showed a different trend. The percentage of intensive students increased from fall/winter to spring while the percentage of strategic students declined substantially from 65 percent to 24 percent, and the percentage of benchmark students substantially increased by 28 percentage points. By spring 2011, almost one half of the students (49%) reached the ORF benchmark or the program benchmark goal. Figure 3 shows the trends of second-grade students on the ORF in each of the status levels.

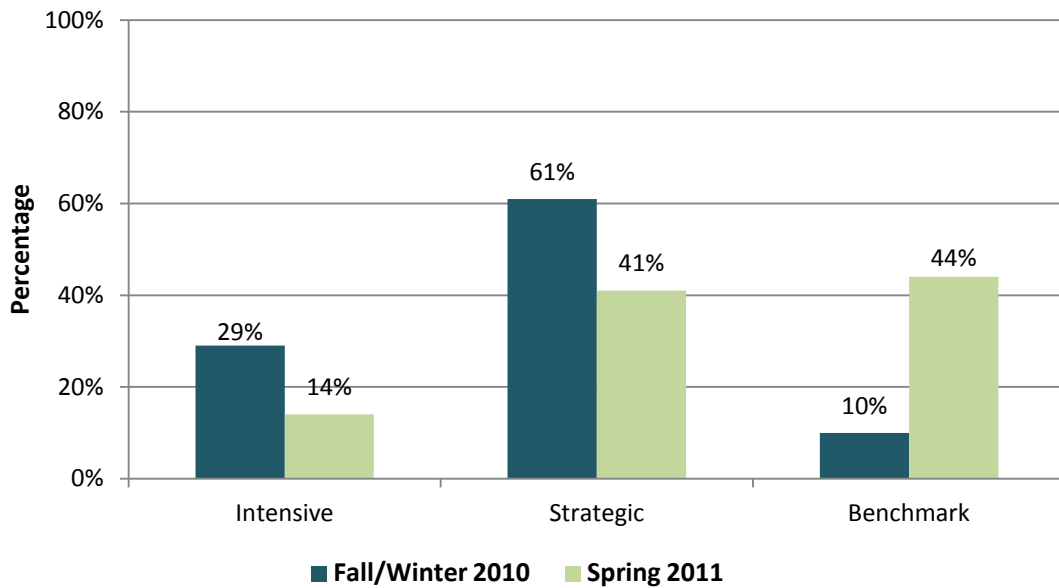
**Figure 3**



**Grade 2—Percentage of Students at Each Level on the ORF (N=543)**

The trend for third grade results was quite similar to those trends found in kindergarten and first grade—the percentage of intensive and strategic students declined over the year while the percentage of benchmark students increased from fall/winter 2010 to spring 2011. The percentage of benchmark students increased substantially from 10 percent to 44 percent—a noteworthy increase of 34 percentage points. Figure 4 displays trends in the three status groups on the ORF for third-grade students.

**Figure 4**



**Grade 3—Percentage of Students at Each Level on the ORF  
(N=488)**

### **Overall Student Performance in Meeting Benchmarks in Different Demographic Categories**

By the spring 2011, 64 percent of all students in Colorado’s Read to Achieve Program reached the program benchmark goal. A greater percentage of American Indian/Native American (75%) and Hispanics (65%) exceeded the percentage for all students. Compared to the overall percentage of 64 percent, the percentage of white (63%) fell slightly lower. On the other hand, only 59 percent of Asian/Pacific Islander and 57 percent of the black students met their program goals. Because there were few students in the American Indian/Native American group, its percentage should be interpreted with caution.

Female students and English language learners (ELL) performed almost the same as their counterparts. Special education students did not perform as well as non-special education students—59% compared to 64% respectively. Table 8 presents these results.

**Table 8**  
**Overall Percentage of Matched Students at Program Benchmark Goal in Spring 2011,**  
**by Demographics, Grades K-3 Combined**

Demographic Characteristics	Percentage(n) of Students Meeting Program Benchmark Goal in Spring 2011	Total Matched N
<b>Colorado Read to Achieve—Overall</b>	<b>64%</b>	<b>1,756</b>
<b>Ethnicity—using the federal reporting categories</b>		
Hispanic	65%	1,279
White, non-Hispanic	62%	183
Black, non-Hispanic	57%	200
Am Indian or Native American	75%	12
Asian or Pacific Islander	59%	39
<b>Gender</b>		
Female	65%	897
Male	62%	858
<b>English Language Learners (ELL)</b>		
ELL	63%	970
Non-ELL	64%	784
<b>Special Education (SPED)</b>		
SPED	59%	81
Non-SPED	64%	1,673

## Movement of Students Across Time

Two indicators of program success are to see how well the program helped intensive and strategic students to progress in their reading, and how well the program kept benchmark students at benchmark. Examining the movement of students in the intensive, strategic, and benchmark groups to other categories from fall/winter 2010 to spring 2011 provides this information. This section examines the percentage of students that changed their status on the DIBELS measures from the fall/winter 2010 to spring 2011, by grade level. At kindergarten, the movement examined was on the PSF measure. For first grade, the movement of students on both the PSF and NWF measures was examined. In second and third grades, student movement on the ORF was investigated.

### Movement Between Measure/ISR Status Levels From Fall 2010 to Spring 2011

**Kindergarten.** On the PSF measure of the DIBELS, schools with kindergarten students were highly successful in keeping their benchmark students at benchmark from fall to spring. Almost all benchmark students (99%) stayed at benchmark. The vast majority of strategic students (88%), and almost four out of five intensive students (77%), moved up to benchmark by spring 2011. None of the intensive students and few strategic students remained at their fall level. Table 9 summarizes these findings.

**Table 9**  
**Kindergarten—Changes in ISR Status\* From Winter 2010 to Spring 2011**

PSF Risk Group in Winter 2010	Spring 2011 Percentage (n)
<b>Phoneme Segmentation Fluency (PSF)</b>	
<b>Intensive (N=31)</b>	
Remained in Intensive	--
Moved to Strategic	23% (7)
Moved to Benchmark	77% (24)
<b>Strategic (N=52)</b>	
Moved to Intensive	--
Remained in Strategic	12% (6)
Moved to Benchmark	88% (46)
<b>Benchmark (N=181)</b>	
Moved to Intensive	--
Moved to Strategic	1% (1)
Remained in Benchmark	99% (180)

\* Scores matched winter to spring.

**Grade 1.** The vast majority of first-grade students either remained at benchmark on the PSF (99%) or moved up to benchmark from the intensive or strategic status in fall/winter 2010—88 percent and 97 percent respectively. None of the fall/winter intensive students remained in the intensive group. Student performance on the NWF generally mirrored their performance on the PSF except that a smaller percentage of intensive (71%) and strategic (84%) moved up to the benchmark level. Overall, schools were quite successful in helping their first-grade students to become successful readers. Table 10 shows these results.



**Table 10**  
**Grade 1—Changes in ISR Status\* From Fall/Winter 2010 to Spring 2011, by Measure**

Risk Group in Fall/Winter 2010	Spring 2011 Percentage (n)
<b>Phoneme Segmentation Fluency (PSF)</b>	
<b>Intensive (N=41)</b>	
Remained in Intensive	--
Moved to Strategic	12% (5)
Moved to Benchmark	88% (36)
<b>Strategic (N=211)</b>	
Moved to Intensive	--
Remained in Strategic	3% (7)
Moved to Benchmark	97% (204)
<b>Benchmark (N=209)</b>	
Moved to Intensive	--
Moved to Strategic	1% (2)
Remained in Benchmark	99% (207)
<b>Nonsense Word Fluency (NWF)</b>	
<b>Intensive (N=69)</b>	
Remained in Intensive	4% (3)
Moved to Strategic	25% (17)
Moved to Benchmark	71% (49)
<b>Strategic (N=241)</b>	
Moved to Intensive	2% (5)
Remained in Strategic	14% (33)
Moved to Benchmark	84% (203)
<b>Benchmark (N=151)</b>	
Moved to Intensive	--
Moved to Strategic	7% (11)
Remained in Benchmark	93% (140)

\* PSF and NWF scores were matched fall to spring.

**Grade 2 and Grade 3.** Second- and third-grade school programs were not as successful as the programs in the other grade levels. In both second and third grades, the majority of students remained at benchmark on the ORF from the fall/winter to spring—81 percent and 88 percent respectively. Only about 45 percent of strategic students, in both of these grades, succeeded in attaining benchmark level by spring 2011. Also, a large percentage of second-grade, intensive students (62%) remained in intensive the entire year while about a third of the third-grade, intensive students remained in the same group. Table 11 portrays this data.

**Table 11**  
**Grade 2 and Grade 3—Changes in ISR Status\* from Fall/Winter 2010 to Spring 2011**

ORF Risk Group in Fall/Winter 2010	Spring 2011 Percentage (n)
<b>Grade 2</b>	
<b>Intensive (N=77)</b>	
Remained in Intensive	62% (48)
Moved to Strategic	21% (16)
Moved to Benchmark	17% (13)
<b>Strategic (N=351)</b>	
Moved to Intensive	27% (95)
Remained in Strategic	28% (98)
Moved to Benchmark	45% (158)
<b>Benchmark (N=115)</b>	
Moved to Intensive	3% (4)
Moved to Strategic	16% (18)
Remained in Benchmark	81% (93)
<b>Grade 3</b>	
<b>Intensive (N=142)</b>	
Remained in Intensive	36% (51)
Moved to Strategic	38% (54)
Moved to Benchmark	26% (37)
<b>Strategic (N=296)</b>	
Moved to Intensive	6% (19)
Remained in Strategic	48% (141)
Moved to Benchmark	46% (136)
<b>Benchmark (N=50)</b>	
Moved to Intensive	--
Moved to Strategic	12% (6)
Remained in Benchmark	88% (44)

\* Scores matched fall to spring.

### **Correlation Between the Grade 3 Oral Reading Fluency (ORF) Measure and the Reading Test of the Colorado Student Assessment Program (CSAP)**

Two different analyses were performed to look at the relationship between the third-grade ORF on the DIBELS and the CSAP reading test in spring 2011. Both analyses included “all” students enrolled in the Colorado Read to Achieve Program with both their ORF scores/levels and CSAP scaled scores/reading test proficiency levels.

## First Analysis

In the first analysis, a Chi-square test was performed to determine if a relationship existed, using the performance descriptors for both measures. A significant relationship between the two measures did exist (Chi-square=74.718; significance at  $p=0.000$ ). The strength of the relationship as measured by the Spearman rho was 0.318 with  $p=0.000$ , meaning that only 10 percent of the variance was accounted for by the scores. This is a very low correlation.

Only half of the benchmark students on the ORF were proficient on the CSAP reading test. If the two assessments were more closely related, we would expect a greater percentage of benchmark students to be proficient or advanced on the CSAP reading test. About one-third of the strategic students (34%) scored proficient on the CSAP while the vast majority of the intensive students (91%) were classified as not proficient on the CSAP—not a surprising result. Table 12 shows the percentage distribution between the ORF and the CSAP levels.

**Table 12**  
**Relationship Between Grade 3 ORF and CSAP Performance Levels, Spring 2011 (N=487)**

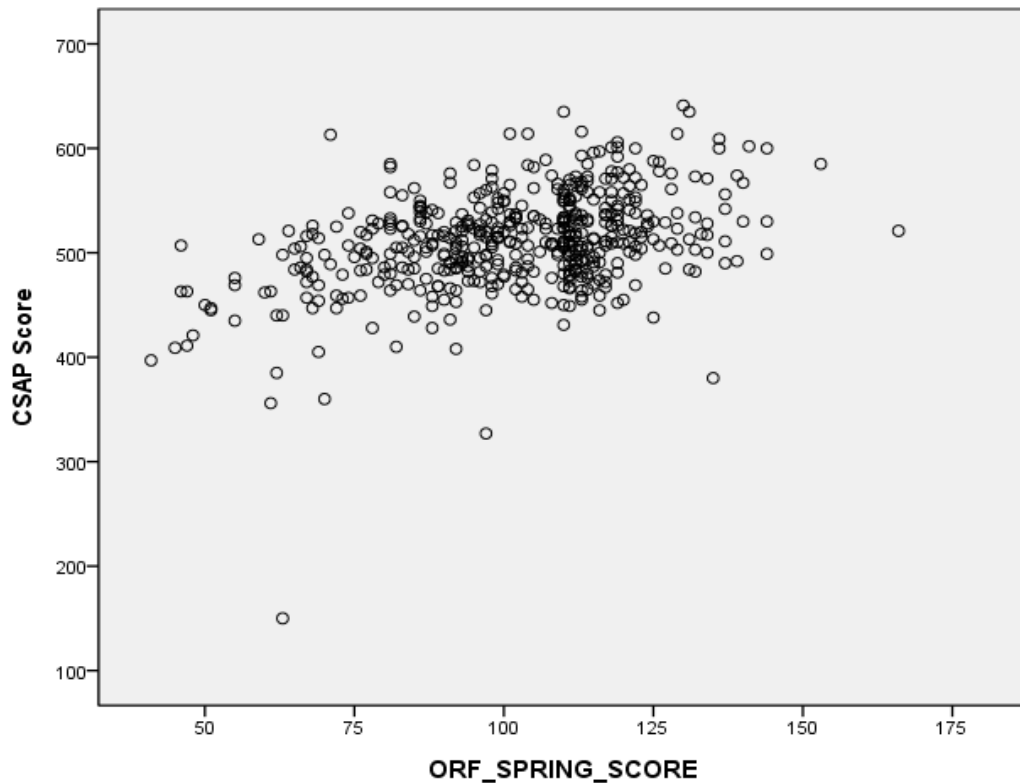
ORF Risk Levels	Percentage of Students—CSAP				N
	Unsatisfactory	Partially proficient	Proficient	Advanced	
Intensive	41%	50%	9%	--	68
Strategic	10%	56%	34%	--	202
Benchmark	7%	43%	50%	--	217

\*\* Chi-square=74.718; significance at  $p=0.000$

## Second Analysis

In the second analysis, the Oral Reading Fluency scores were used to correlate with the CSAP scaled scores using the Pearson R correlation. The correlation was 0.420 with  $p=0.000$  which means that approximately 18 percent of the variance was accounted for by the two scores. Other factors would contribute to the remaining variations in the scores. Figure 5 shows a scatterplot of the scores and the modest correlation between the scores, i.e., if a line was drawn through the dots, so all dots were as close as possible to the line, the line's slope would not be very steep.

**Figure 5**



**Scatterplot of Spring 2011 CSAP Scaled Scores and DIBELS ORF Scores**

While the correlation found was a very modest one, it might have been higher if there had not been a restriction in the range of scores—only poor readers had been selected to participate in the program. When there is a restriction of range, the Pearson R can shrink because of less variability in the scores. For example, in a 2002 study of 58 students from all ability levels using the same type of scores, the Pearson correlation was found to be 0.80<sup>1</sup>—quite a difference from what was found with the Read to Achieve scores. For this reason, these findings should be interpreted cautiously.

In both of these analyses, there was a very low/modest correlation between the Oral Reading Fluency measure and the CSAP reading test in the third grade for this population of students. Scoring at benchmark on the ORF was not a particularly good predictor of meeting proficiency on the CSAP, at least not in this group of students.

<sup>1</sup> Shaw, R. and Shaw, D. (2002) DIBELS Oral Reading Fluency-based Indicators of Third Grade reading for Colorado State Assessment Program (CSAP) found at <http://www.docstoc.com/docs/37042757/Technical-Report-DIBELS-Oral-Reading-Fluency-Based-Indicators-of-Third>

## Thoughts for Reflection

This year was the first year that the fourth cohort of schools implemented the Read to Achieve Program. The first year of implementing any program usually is more challenging for schools and does not produce as positive results as when a program has been implemented for more years. It was interesting that when this year's results were compared to the first year for the third cohort of schools (2007–2010), this year's students demonstrated substantially better results. In the first year of the third cohort, 77 percent of kindergarten students, 47 percent of first-grade students, and 43 percent of students in the second and third grades met the program benchmark goals. This year, about 95 percent of kindergarten students and 84 percent of first-grade students met the program benchmark goals. In fact, this year's results for kindergarten and first-grade students were comparable to results found for students in the same grade levels in schools in their third year of implementing the Read to Achieve Program. Also, 49 percent and 44 percent, of second- and third-grade students, respectively, reached the program benchmark goals this year. Unfortunately, there was no data to explain possible reasons for these differences between cohorts.

Across years, kindergarten and first-grade students continued to show better results than those found for students in the second and third grades. Regardless of the year, only about one-half or fewer of second- and third-grade students met their program goals. Also, a larger percentage of intensive and strategic students remained in their risk groups for the entire year compared to kindergarten and first-grade students. This is not really an acceptable outcome for improving student reading. Why were so few second- and third-grade students meeting the program benchmark goal? Are these grade-level programs different than the ones implemented in kindergarten and first grade? If so, how do they differ? Should the second- and third-grade programs be changed? If so, how should they be modified? These questions are hard to answer without information about how schools were implementing Read to Achieve at their schools or whether schools were consistently implementing their programs within their schools and across schools.

Improving the reading skills of poor readers can be challenging, especially when a program is not tightly defined and more closely aligned with research. Without clearly articulated program guidelines, it is difficult to say whether or not a program impacted reading. For example, it might be that specific strategies such as an additional 30 minutes every day at one school improved students reading while pulling out students during reading instruction at another school did not. When the results from the two schools are aggregated, they will dilute the impact of the state program—Colorado Read to Achieve, in this case. Implementing a variety of school programs also makes it extremely difficult to interpret findings and to make program improvement decisions at the state level. What do you change if all the school programs are doing something different? And how do you know what is working?

Once the state program is clearly defined—how it should be implemented, the intensity of intervention, appropriate curricula to use, teacher training, etc.—and expectations are articulated to school programs, systematic data collection about fidelity of implementation will help to clarify school program needs and needed changes, leading to better alignment with state

program guidelines. Greater alignment will eventually lead to more informative interpretations of findings and better future program decisions. More importantly, if guidelines are based in research, then results should be more positive, and the students benefit.

How can fidelity of implementation data be collected? Multiple methods exist which might be used in combination or singularly, such as:

- *Survey schools.* The survey might ask questions about curriculum, implementation, and intensity at each grade level and teacher training and be completed by a primary school contact person and/or a sample of grade-level teachers and principals.
- *Conduct site visits to a random sample of schools.* The site visits might consist of interviews with the principal and a sample of teachers implementing the intervention and regular classroom teachers, in addition to observations of the intervention.
- *Conduct phone interviews.* A random sample of schools might be selected for conducting phone interviews with principals and teachers.
- *Develop an implementation checklist.* The checklist might be used during site visits by the state consultants to identify needed changes and to provide evidence of level of fidelity.

Besides the lack of tight program guidelines based in research, another concern is the use of the DIBELS assessment as both the screening assessment and the fall benchmark assessment. When the same instrument is used for screening and pretest (i.e., fall benchmark), and students are selected based on low scores, as in Colorado Read to Achieve, regression to the mean can account for some or all of positive changes found from pretest to posttest. A different instrument than the DIBELS should be used to select the students and the DIBELS used as the pretest.

Simply put, this is what happens: Each score is made up of a “true” score and a certain amount of measurement error caused by factors such as room conditions, student’s physical well being, and test-taking skills. Scores at the extreme ends of the score distribution have more error. When students are selected based on their low scores and the scores are also used as the pretest, their scores will have more error than those in the middle of the score distribution. At a second testing or posttest, the average score for this group of low performing students will positively change because the measurement error for the group will decrease. The decrease in measurement error causes the average score to improve without any real change in achievement. This increase in the average score is due to what is called “regression to the mean.” Because Read to Achieve selects students and pretests them using the same fall DIBELS benchmark assessment, any improvement we might see at posttest or spring DIBELS benchmark can be, in part, due to regression to the mean.

Finally, and as mentioned in the last section of this report, the selection of poor readers or Read to Achieve students for correlating DIBELS and CSAP scores calls into question the appropriateness of this correlation because of a restriction in the range of scores. Because there are no middle or high scorers, there will be less variation among the scores. As a result, the correlation between these two sets of scores can shrink due to the less variability in their scores.

