



# How Do Students Fare after Four Years in Delaware Reading First?

## Longitudinal Study of Student Outcomes

September 30, 2008

Report prepared for:

Delaware Department of Education and Dr. Sandra Meyers, Director  
Delaware Reading First.



**Acknowledgements:** Thank you to Xiaoyu Qian for her assistance with data analysis.

*Recommended citation:* Grusenmeyer, L.H., & Uribe-Zarain, X. (2008). *How do students fare after four years in Delaware's Reading First? Longitudinal study of student outcomes.* [Technical Report No. T2008.07.01] Newark, DE: Delaware Education Research and Development Center.

# Executive Summary

Results are mixed when examining cross sectional data to identify student achievement impact of Delaware’s Reading First program (DERF). Earlier evaluation studies comparing third grade reading performance in DERF schools with that in similar non-DERF schools found little evidence of improvement. (See Table 1.) By identifying students who began and remained in DERF schools for four years and analyzing their progress from kindergarten to third grade, we hoped to gain insight regarding the programs’ student impacts and overcome the limitations inherent in comparing performance of cross sections of students.

Findings indicated that Delaware Student Testing Program (DSTP) reading performance of students with four years of program experience did not significantly differ from that of students with one year of program experience (3rd grade only). Differences were noted in the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) measure of Oral Reading Fluency. When comparing the students who met or exceeded the standard, four year students were more likely to exceed the standard than one year students.

	2003*	2004	2005	2006	2007
School pair #1	↑	↓	↓	↓	↑
School pair #2	↓	↓	↓	↓	↓
School pair #3	↓	↓	↓	↓	↓
School pair #4	↓	↓	↓	↓	↑
School pair #5	↓	↓	↓	N/A	N/A
School pair #6	↓	↔	↑	↔	↑

\* Baseline year

Table 1. Do ↑ higher, ↓ lower, or ↔ similar percents of students in DERF schools meet or exceed third grade reading standard (DSTP) than in similar, non-DERF schools?

## Results in Brief

***Finding 1. When considering students who have participated in DERF for four years, there is little change in those identified to be “at risk” of reading failure.***

- a. The percent of entering kindergarteners described as “at risk” does not change when the same group of children have completed third grade.
- b. Fewer students met third grade Oral Reading Fluency (ORF) benchmarks than met ORF first grade benchmarks.

***Finding 2. When comparing students with four years in DERF schools to those with only one year, there is no significant difference between the percent of students in either group meeting Delaware’s third grade DSTP reading standard; however,***

- a. Significantly more students met DIBELS third grade fluency benchmarks after four years in DERF than after one year in the program.
- b. Of the students who met or exceeded the standard, students enrolled for four years in DERF schools were nearly two times as likely to *exceed* the standard as those with only one year in the program.

## **Recommendations**

For program developers— take a close look at every school’s intervention materials, delivery formats, schedules, and providers.

For teachers and literacy coaches—study your instructional quality regarding vocabulary, comprehension, and critical thinking.

For principals and fourth and fifth grade teachers of the 51— be aware that these students are already at risk. There’s no benefit to a “wait and see” strategy.

# Table of Contents

Executive Summary .....	i
Introduction .....	1
Method .....	1
<i>Sample</i> .....	1
<i>Instruments</i> .....	2
Key Findings.....	3
Limitations .....	5
Discussion.....	6
Recommendations.....	7

# How Do Students Fare after Four Years in Delaware Reading First? Longitudinal Study of Student Outcomes

## Introduction

This study tracks the performance of a cohort of students enrolled in Delaware Reading First (DERF) schools for four years and describes the group's reading growth and achievement. It also compares third grade achievement scores of these four year students with test scores of students with only one year in the program. Data for this study was gathered in the first four years of DERF's five year implementation to answer questions regarding the program's impact on student reading achievement:

1. Have children in DERF classrooms made significant improvement in their reading performance?
2. How does the number of years students spend in a DERF school relate to their achievement?

## Methods

### *Sample*

The study examined test scores of third graders enrolled in 11 DERF schools in spring 2007. One group included all 350 students continuously enrolled in DERF schools since kindergarten in 2003. The other was composed of the 216 third graders who enrolled in DERF schools for the first time in 2007. The two groups were similar to each other in demographic make-up, but they were not similar to total statewide third grade enrollment regarding income, special education status, or racial make-up. (See Table 2.) **Note:** Demographic information and DSTP test scores were missing for 18 students in the 1 Year group; however, because we determined that the 18 students were distributed across all schools and evenly distributed across all three DIBELS benchmark levels, we included them in the DIBELS analysis.

		4 Years in RF %	1 Year in RF %	*All DE 3 <sup>rd</sup> graders %
Race	American Indian	.6	.5	.4
	African American	49.3	64.1	35.3
	Asian	.8	1.5	3.1
	Hispanic	10.2	9.1	7.6
	White	38.8	24.7	53.6
Special Education	Yes	15.3	16.7	3.8
	No	84.7	83.3	96.5
Low Income	Yes	62	68.2	41.3
	No	38	31.2	58.7
Limited English	Yes	4.5	3	3.2
	No	95.5	97	96.8
Total N		350	198	7874

\*Data taken from Delaware DOE public DSTP reporting website

Table 2. Demographic characteristics of Delaware third graders, spring 2007 DSTP (%)

## ***Instruments***

Two different measures were used to determine an individual student’s reading achievement in DERF schools, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and total reading scores from the state mandated Delaware Student Testing Program (DSTP).

DIBELS subtests were administered to students at least three times each year in grades K to 3. DIBELS protocol recommends testing incoming kindergarten students on Letter Naming Fluency (LNF) and Initial Sounds Fluency (ISF). Typically, measures of LNF are repeated in spring of kindergarten, but ISF is not. Each first to third grader is screened for progress toward an Oral Reading Fluency (ORF) benchmark. Student scores are used to predict the likelihood that he or she will achieve a favorable reading outcome. Scores are reported in terms relative to predictions of future reading success such as “At risk,” “Some risk,” or “Low risk.”

The DSTP assesses student achievement relative to Delaware’s educational standards and is reported on a 5-point scale called performance levels (PL). Students who score PL3 and above are considered to have “Met or exceeded the standard.” Although students take a form of the DSTP in second grade, this study only examines aggregable, third grade DSTP “total reading” scores.

## Key Findings

**Finding 1. When considering students who have participated in DERF for four years, the number of entering kindergarteners described as “at risk” changes little when compared to the same group of children in third grade.**

The cohort of four-year students improved their kindergarten LNF performance from fall 2003 to spring 2004. However the number of students meeting ORF benchmarks did not grow relative to the rising fluency benchmarks from first to third grade. That is, more students were categorized “at risk” in third grade than were “at risk” in first grade. (See Figure 1.)

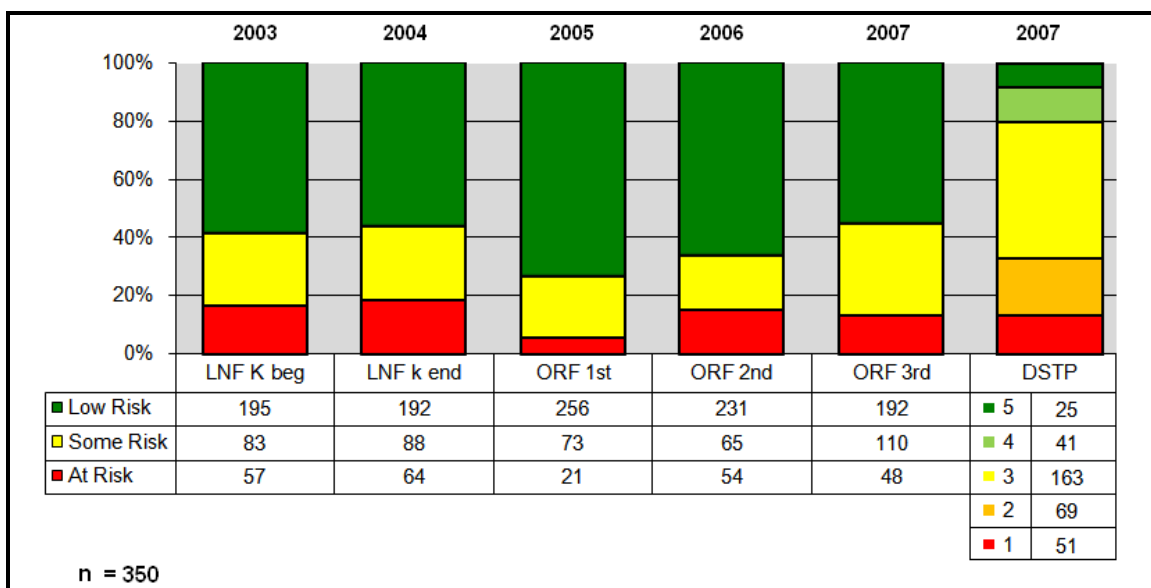


Figure 1. Cohort performance on DIBELS and DSTP1 from entry at kindergarten through third grade (Ns)

**Finding 2a. Significantly more students met third grade fluency benchmarks after four years in DERF than after one year.**

A chi-square test of independence was calculated comparing the frequency of third grade students classified as “Low risk,” “Some Risk,” and “At Risk” by the ORF test for students that were in the program for 4 years and for only one year. A significant interaction was found ( $\chi^2(2) = 8.67, p < .05$ ). Students who were in the program for four years were more likely to be classified as “Low Risk” by the spring DIBELS ORF than students who were in the program for only one year. Also, students who were in the program for one year were more likely to be classified as “At Risk” on measures of ORF than students who were four years in the program. (See Figure 2.)



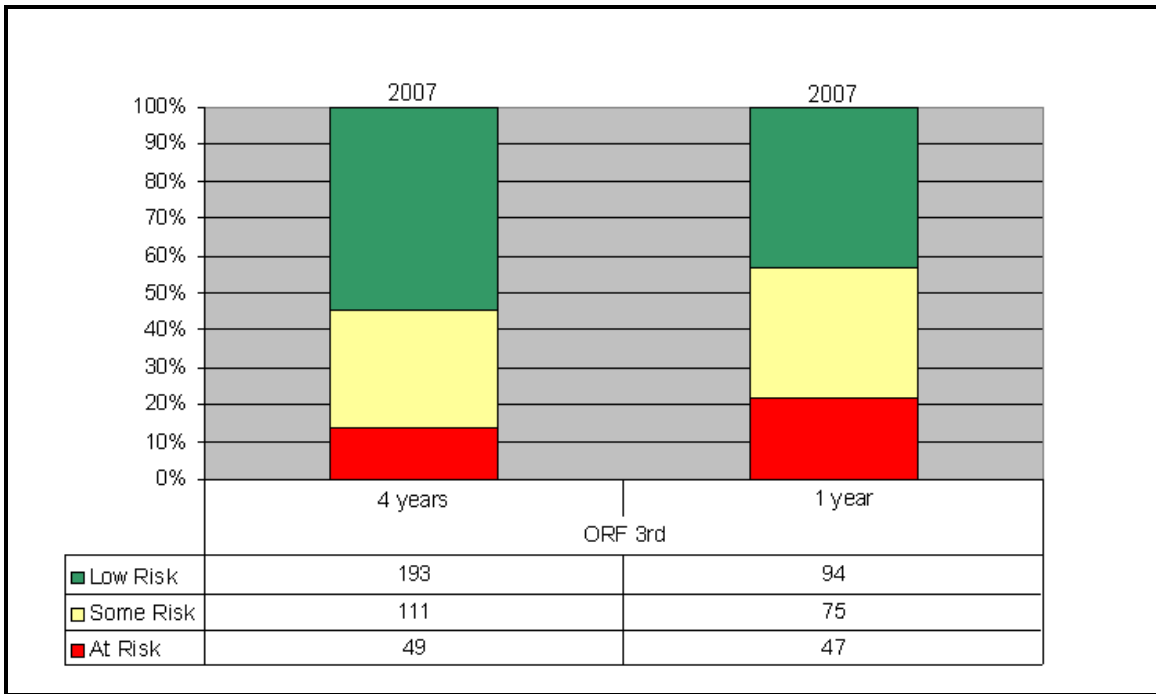


Figure 2. Third grade students 2007 DIBELS ORF performance by years in program (N)

**Finding 2b. Whether a student is in a DERF school for one year or for four years, there is no significant difference in the likelihood that he/she will meet Delaware’s third grade reading standard.**

While fewer students who participated in DERF programs for one year met or exceeded the reading standard as measured by the DSTP than students who participated for four years, this difference was not statistically significant. In the same way, the difference between the numbers of children in each group scoring at each performance level (PL) is not significant. (See Figure 3.)

**Finding 2c. Of students at or above the standard, four year students were significantly more likely to “exceed” the standard (PL 4-5) than one year students.**

While students with four years in the program were no more likely to meet the standard than students with one year, they were almost twice as likely to *exceed* the standard ( $\chi^2 (1) = 5.70, p = .02$ ). For students below the standard, more four year students scored PL2 than PL1; however this was not a statistically significant difference.

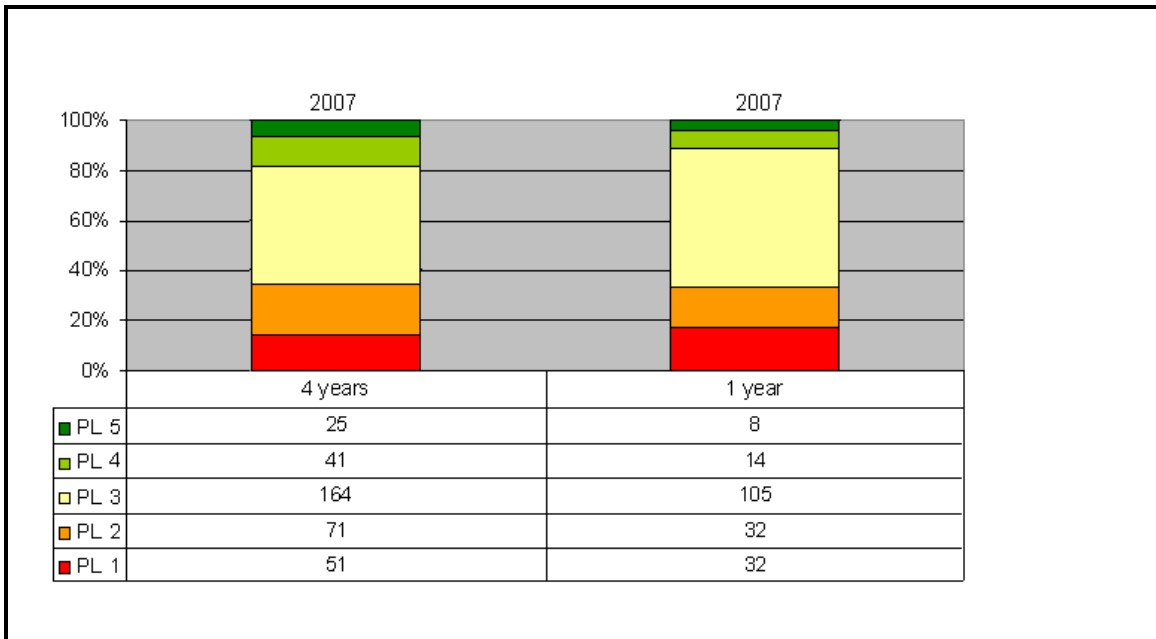


Figure 3. Third grade students 2007 DSTP Reading Performance Levels by years in program (N)

## Limitations

As noted earlier, DERF students differ in several ways from students in other schools and districts. Their rates of disability and low income bear this out. Because progress monitoring, curriculum standards, and achievement tests are all based on the relationship between one student’s development and *time*, these instruments may not fully or meaningfully capture growth in this group of students.

In its plan to provide for greater levels of student needs, DERF offered high levels of support for teacher professional development and high levels of funding for targeted student invention. However, we don’t know how schools were impacted by teacher and administrator turnover or to what degree other site-based or teacher- level differences might have effected achievement.

Finally, we don’t know what student mobility means in these low performing, high poverty schools. Are these 350 children (and their families) substantially different from the hundreds of students who moved away or transferred to other, non-DERF schools sometime between kindergarten in 2004 and third grade in 2007?

## Discussion

This study is an attempt to address limitations inherent in other types of analysis. Evidence of non-program schools adopting the program's practices existed from its beginning and limited the usefulness of comparisons between DERF and non-DERF schools. Because No Child Left Behind and accompanying federal regulations (such as the requirement that all students have Highly Qualified Teachers) came into effect alongside Reading First, comparison of cross-sectional achievement within any one school is also problematic. Here, we focused narrowly on students with the longest exposure to the program in order to investigate the long term cumulative effects of Reading First on student achievement. Our study of longitudinal performance indicates uneven results.

Upon entering kindergarten, 57 of these 350 children were predicted to be "at risk" of not reaching reading success. At the end of second grade, preparing to enter third grade, 54 continued to score "at risk." Indeed, DIBELS' projection seemed to hold true; that is, one year later, 51 of these same students scored at the lowest performance level on the state third grade reading standards assessment. Ultimately DERF schools' targeted interventions were not effective for these "at risk" students even though, at the end of both kindergarten and first grades, they appeared to have made measurable progress toward fluency benchmarks.

At the program level, if Delaware Reading First's primary task is to intervene for those identified most "at risk," it seems to have failed. In fact, Delaware's third grade PL1 label has proven to be a fairly stable predictor; i.e., in one long term study, the *majority* (66%) of third graders who scored PL1 failed to meet the state's standards for 8<sup>th</sup> grade reading five years later.<sup>1</sup> The future is not promising for the 51 children in our study who began kindergarten in 2003.

Although our state test was not designed to detect instructional influences (Popham, 2007)<sup>2</sup>, small differences in student performance hint that, in fact, DERF's instructional program may have positive student impacts. Students with four years have higher rates of oral reading fluency and they exceeded the state's reading standard more frequently than those with one year in the program.

We don't know if these subtle differences will positively impact future school success. However, if the DSTP performance levels indicate qualitatively different types of reading, then perhaps it is meaningful for the children.

## Recommendations

**For program developers— take a close look at every school’s intervention materials, delivery formats, schedules, and providers.**

“At risk” students appeared to approach the benchmark in greater numbers after kindergarten and first grades. Did these students continue to receive targeted intervention in second grade? If so, why was it less effective? How and when are schools making intervention decisions?

If you find that one or two sites appear to succeed with their interventions, share the results widely. Develop consensus among the other school leaders that the status quo isn’t working. Target professional development and technical support to improve intervention strategies state-wide.

**For teachers and literacy coaches—study your instructional quality regarding vocabulary, comprehension, and critical thinking.**

We know from earlier research, that fluency accounts for only one part of a third grader’s DSTP reading performance.<sup>3</sup> While the 1 Year DERF students may be less fluent readers, other factors account for their similar success rates on the DSTP. By looking closely at indicators of your own instructional quality, you may uncover rich opportunities for professional improvement that yield measurable student impacts.

**For principals and fourth and fifth grade teachers of the 51— be aware that these students are already at risk. There’s no benefit to a “wait and see” strategy.**

While an earlier study showed that 66% of PL1 third graders did not meet the 8<sup>th</sup> grade standard, it also demonstrated that 34% DID. It can be done. For the children in this study, every effort can be made to make connections with their parents and with school support personnel now in order to provide them with a better chance of catching up and meeting the standard by 5<sup>th</sup> or 8<sup>th</sup> grade.

## **Endnotes**

---

<sup>1</sup> Uribe-Zarain, X. and Noble, A. J. (2007). How do Performance Level 1 and 2 students in Delaware perform over time? [Technical Report No. T2006.16.02]. Newark, DE: Delaware Education Research and Development Center.

<sup>2</sup> Popham, W. J. (2007). Instructional insensitivity of tests: Accountability's dire drawbacks. *Phi Delta Kappan*, 146-151.

<sup>3</sup> Uribe-Zarain, X. (2007). Relationship between Performance on DIBELS Oral Reading Fluency and Performance on the Reading DSTP 2006 [Technical Report No. T2007.06.02 ]. Newark, DE: Delaware Education Research and Development Center.