



Delaware Research &
Development Center
University of Delaware
Newark, DE 19716

PUBLICATION T06.013.01

August 2006

Relationship between Student Performance on DIBELS Oral Reading Fluency and Third Grade Reading DSTP

In Delaware's Reading First program, fluency instruction is one component of a balanced reading "diet" with the DIBELS assessment providing both an instructional target and a measure of reading growth (DE DOE, 2002.) Yet, the relationship between reading fluency and reading comprehension is often not clear in the eyes of the lay public and historically has been minimized in teacher education programs (Rasinski, 2003.)

This study determines that the DIBELS measure of Oral Reading Fluency (ORF) is a moderately reliable predictor of reading performance on the Delaware Student Testing Program (DSTP) in third grade. Given the three DIBELS ORF predictive levels and considering whether the student did or did not meet the DSTP standard, students were classified into six different groups. In particular, this study examines the characteristics of the students who are in each of these groups. Implications for further research are provided.

For more information or questions regarding this technical report, contact:

Ximena Uribe-Zarain,
University of Delaware Education Research & Development Center
Phone: 302-831-4434

E-mail: ximena@udel.edu

Prepared by XIMENA URIBE-ZARAIN, GRADUATE RESEARCH ASSISTANT

The University of Delaware is committed to assuring equal opportunity to all persons and does not discriminate on the basis of race, color, gender, religion, ancestry, national origin, sexual orientation, veteran status, age, or disability in its educational programs, activities, admissions or employment practices as required by Title IX of the Education Amendments of 1972, Sections 503 and 504 of the Rehabilitation Act of 1973, Titles VI and VII of the Civil Rights Act of 1964, the Americans with Disabilities Act, Executive Orders 11246 and 11375 and other applicable statutes. Inquiries concerning Title IX, Section 503 and 504 compliance, Executive Order 11246 and information regarding campus accessibility and Title VI should be referred to the Affirmative Action Director, 305 Hullihen Hall, 302-831-2835, 302-831-4552 (TDD).

This study was conducted with support from the Delaware Department of Education.

INTRODUCTION:

It has been argued that DIBELS data predicts performance in state reading tests (see Barger, 2003; Buck & Torgesen, 2003; Carlisle, Schilling, Scott, & Zeng, 2004; Elliott, Lee, & Tollefson, 2001; Good, Simmons, and Kame'enui, 2001; Hintze, Ryan, & Stoner, 2003; McGlinchey & Hixson, 2004; Shaw & Shaw, 2002; Vander Meer, Lentz, & Stollar, 2005; Wilson, 2005). Eight studies have sought to explain the relationship between DIBELS measures and high-stake state tests. Most of the studies employed correlation to indicate the strength of the relationship between state test reading scores and the DIBELS Oral Reading Fluency (ORF) scores of third grade students (Barger, 2003; Buck & Torgesen, 2003; Good, Simmons, and Kame'enui, 2001; Hintze, Ryan, & Stoner, 2003; Shaw & Shaw, 2002; Sibley, Biwer, & Hesch, 2001; Vander Meer, Lentz, & Stollar, 2005; Wilson, 2005). These studies use third grade students as their population because state tests are given at the end of the year and it can be related to the three times the ORF is taken during this same year. Overall, the correlations found in the studies range from .60 to .80; but the sample sizes in these studies vary widely.

Besides the correlation of reading state test scores and DIBELS ORF, a common practice among these studies is to examine the accuracy of ORF classification. The researchers analyzed the numbers of students identified in the "at risk" and "low risk" categories (as positioned by the DIBELS ORF) and the students' performance in their reading state test. They looked at the extent to which DIBELS ORF correctly indicated whether third graders would meet grade level expectations in reading as dictated by their state test. For example, Carlisle et al (2004) found that for the case of Michigan Reading First schools students, almost all of the students that were classified as "at risk" on DIBELS ORF in the fall were correctly predicted to be below the fiftieth percentile on the reading state test.

They found very few false positives or cases where students were predicted to fall below the fiftieth percentile, but actually scored above it. Studies on students' scores from North Carolina, Colorado, Arizona, and Illinois found that DIBELS ORF correctly predicted the performance in the reading state test for students classified as "low risk" (Barger, 2003; Buck & Torgesen, 2003; Shaw & Shaw, 2002; Sibley, Biwer, & Hesch, 2001; Wilson, 2005). Conversely, the Michigan study showed the lowest accuracy for students classified as "low risk" with more than 50% of false positive cases.

PURPOSE OF THE STUDY

This study was conducted to determine if the measure of reading fluency in the DIBELS is a reliable predictor of reading performance on the Delaware Student Testing Program (DSTP) in third grade. Given the three DIBELS ORF outcome levels and considering whether the student met or not the DSTP standard, students were classified into different groups. In particular, this study examines the characteristics of the students who are in these groups.

METHOD

The nine schools participating in the Reading First program in the state of Delaware in the 2004-2005 year provided the data for this analysis. From the 652 third grade students from Reading First schools in **2004-2005**, only **630** were included in the study. Only students with DIBELS and DSTP scores were considered. Fifty percent of the students were female, 15% were classified as special education students, less than 3% were considered limited English proficient, and 59%

received free or reduced lunch (an indication of low income status). The racial composition of the sample is illustrated in Table 1.

Table 1. Racial composition of the sample

	frequency	percent
American Indian	3	0.5
African American	308	47.2
Asian	9	1.4
Hispanic	43	6.6
White	289	44.3
Total	652	100.0

The measures employed were the oral reading fluency (ORF) subtest of the DIBELS and the reading portion of the DSTP. The ORF is taken three times a year, but because the reading DSTP is taken in March, the winter ORF was selected. The ORF is a one-minute test and its score yields a risk level: “at risk,” “some risk,” and “low risk.” Reading DSTP scale scores were used in this analysis. The DSTP scores are classified into five performance levels: (1) well below the standard, (2) below the standard, (3) meeting the standard, (4) above the standard, and (5) distinguished. The performance levels were used to see whether the student was meeting the standard or not.

RESULTS

There was significant correlation between ORF scores and reading DSTP scores ($r = .61$, $p < .01$). Figure 1 shows the relationship between ORF scores and reading DSTP scores of third grade students in 2005¹. Students scoring above the horizontal line met or exceeded the Delaware standards in reading for third grade (i.e., their DSTP performance level were 3, 4, or 5). The vertical lines represent the benchmarks for the DIBELS ORF subtest according to its creators.² Third grade students with winter ORF scores below 67 are considered at “at risk” of achieving a below grade level score on reading comprehension measures. Students scoring between 67 and 91 correct words per minute are considered to be at “some risk” of showing a lower grade level score in comprehension. Finally, students with ORF score of 92 or more are considered to be at “low risk.”



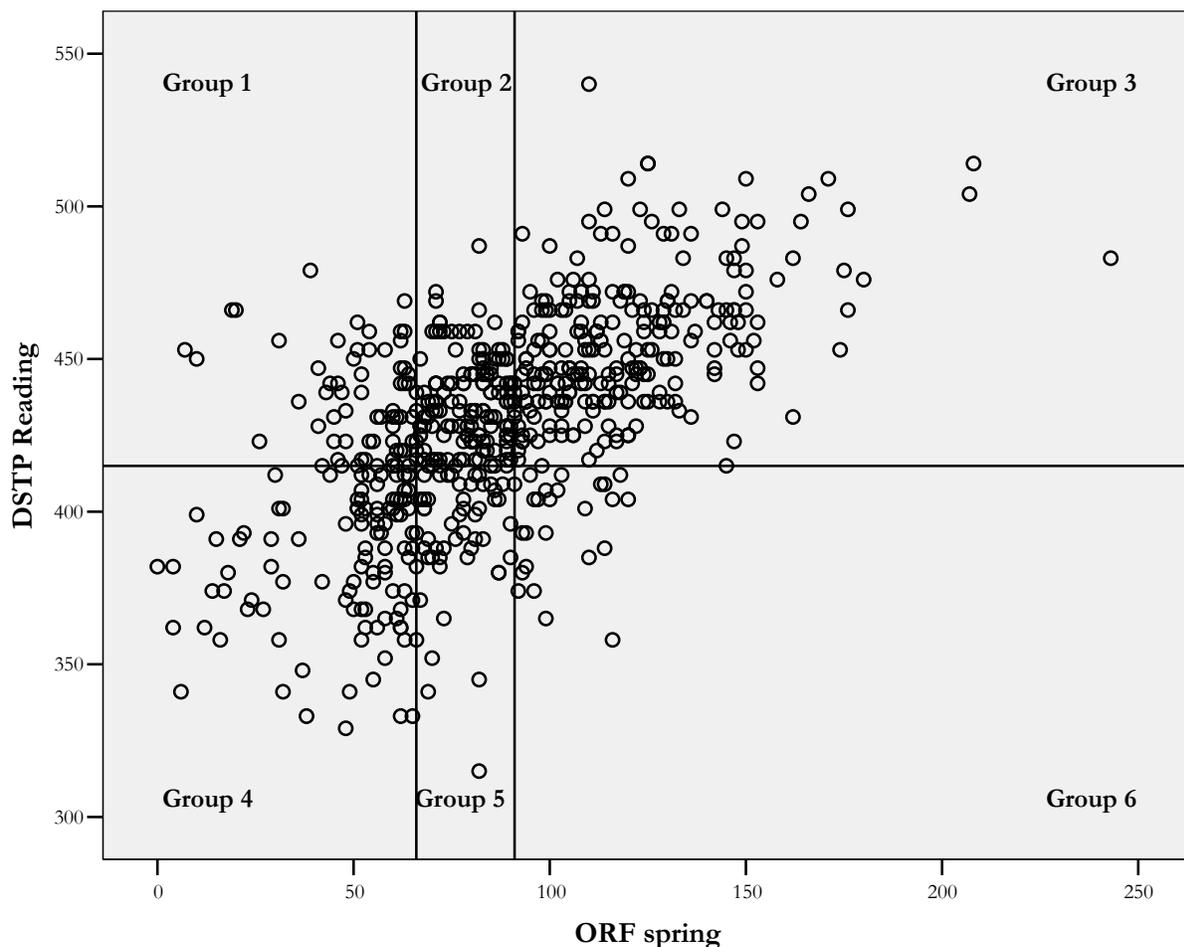


Figure 1. Relationship between ORF scores and reading DSTP scores of third grade students in 2005.

Students have been classified into six groups according to their scores in both tests. Group 1 includes 10.8% of the students; they are meeting or exceeding the reading DSTP standard but are classified as “at risk” by the ORF measure. Group 2 includes 22.9% of the students; they are meeting or exceeding the reading DSTP standard but are classified as “some risk” by the ORF measure. Group 3 includes 38.3% of the sample; they are meeting or exceeding the reading DSTP standard and are classified as “low risk” by the ORF measure. Group 4 includes 16.3% of the sample; they are below the reading DSTP standard and are classified as “at risk” by the ORF measure. Group 5 includes 7.9% of the students; they are below the reading DSTP standard and are classified as “some risk” by the ORF measure. Group 6 includes 3.8% of the students; they are below the reading DSTP standard but are classified as “low risk” by the ORF measure.

Students in group 3 and group 4 perform as we would expect if the DIBELS ORF score is considered a good predictor of the reading DSTP score. In other words, students identified as “at risk” in fact performed below the standards on the DSTP (this is called sensitivity). Now, students identified as “low risk” in fact met or exceeded the standards on the DSTP (this is called specificity).

Conversely, students in groups 1 and 6 perform as we would not expect. Students in group 1 are considered “false positives.” False positives are individuals that a screening mechanism (in this case the DIBELS ORF) classifies as “at risk,” but later perform satisfactorily on the criterion outcome (in this case the reading DSTP). Similarly, students in group 6 are considered “false

negatives.” False negatives are individuals that a screening mechanism classifies as “low risk,” but later perform unsatisfactorily on the criterion outcome.

A good screening measure must be accurate at distinguishing students that require intervention from those who do not. A screening measure is characterized by its degree of sensitivity and specificity. We would like to find high sensitivity and high specificity in the relation between the ORF subtest and the reading DSTP. Since sensitivity increases as false negative cases decreases and specificity increases as false positive cases decreases, the fewer false negative and false positive cases the better.

Table 2 is a summary of how the DIBELS ORF scores predict the reading DSTP scores. From the table, one can see that students in the “some risk” category were more likely to meet the reading standard. Also, students classified as “low risk” were very likely to meet the standards. In contrast, the prediction of students “at risk” yielded a large number of false positive cases.

Table 2. ORF predicting Reading DSTP

Reading DSTP	ORF Classification			Total
	At risk	Some risk	Low risk	
Meeting or exceeding the standards	68	144	241	453 72%
Below the standards	103	50	24	177 28%
Total	171 27%	194 31%	265 42%	630

Sensitivity= .81
Specificity= .78

Note: Sensitivity and specificity were calculated without considering the students at "some risk."

Notice that the sensitivity and specificity were calculated without considering the students at "some risk." If we included the students in this category, the sensitivity would be .58 and the specificity would be .85. This means that if we consider the students in both the “some risk” and “at risk” categories, we create more false negative cases. Since “some risk” is a gray area in the field of decision making, it is better to calculate sensitivity and specificity without them. However, we do not want to disregard the importance of the students in this indeterminate state. It is important to know who they are. In the next section, a description for each of the groups is given.

DESCRIPTION OF THE GROUPS

The way the groups are shown in the following figures mimic the position of the groups in Figure 1. The top three groups refer to students who met or exceeded the standard and groups on the left hand are the ones classified as being “at risk” by the DIBELS ORF.

✚ Racial composition

Since African American, Hispanic and white students are more than 98% of the student in the sample, we will focus our attention in these racial groups. Figure 2 shows the difference in racial make up of the groups.

	ORF: “at risk”	ORF: “some risk”	ORF: “low risk”
Meeting the DSTP standard	Group 1: 48% African American 6% Hispanic 46% White	Group 2: 44% African American 6% Hispanic 48% White	Group 3: 36% African American 7% Hispanic 54% White
Below the DSTP standard	Group 4: 62% African American 8% Hispanic 30% White	Group 5: 56% African American 12% Hispanic 32% White	Group 6: 79% African American 0% Hispanic 21% White

Figure 2. Racial composition by group

By looking at the ORF classification, White students have higher concentration in the groups where students met or exceeded the state reading standards. For example, in the “at risk” groups, 46% of the students meeting the standard are White, where only 30% of the students below the standard are. On the contrary, African American students have higher concentration in the groups where students were below the DSTP standards. For example, in the “low risk” groups, 36% of the students meeting the standard are African American, where 79% of the students below the standard are. It is interesting to see that Hispanic students have larger concentration in the groups below the DSTP standard with the exception of groups 3 and 6. There are no false negative cases of Hispanic students.

✚ Socioeconomic status

Socioeconomic status is characterized by whether the students received free or reduced lunch. Remember, 59% of the students in the sample received free or reduced lunch. Figure 3 shows the difference in socioeconomic status of the groups. By looking at the ORF classification, low income students have a higher concentration in the groups where students were classified as “at risk” by the DIBELS ORF. For example, in the “at risk” groups, more than 70% of the students are low income compared to 59% in the whole sample. Also notice that in group 3 where students met or exceeded the DSTP standard and were classified as “low risk” by the ORF, we found the higher concentration of not low income students: 58% versus 41% in the whole sample.

	ORF: “at risk”	ORF: “some risk”	ORF: “low risk”
Meeting the DSTP standard	Group 1: 29% Not low income 71% Low income	Group 2: 32% Not low income 68% Low income	Group 3: 58% Not low income 42% Low income
Below the DSTP standard	Group 4: 25% Not low income 75% Low income	Group 5: 38% Not low income 62% Low income	Group 6: 37% Not low income 63% Low income

Figure 3. Income status by group

✚ Special education

Also of interest is the make up of special education in each group. Remember, 15% of the students in the sample are in special education.

	ORF: “at risk”	ORF: “some risk”	ORF: “low risk”
Meeting the DSTP standard	Group 1: 59% Not special ed. 41% Special education	Group 2: 88% Not special ed. 12% Special education	Group 3: 95% Not special ed. 5% Special education
Below the DSTP standard	Group 4: 70% Not special ed. 30% Special education	Group 5: 90% Not special ed. 10% Special education	Group 6: 83% Not special ed. 17% Special education

Figure 4. Special education by group

From Figure 4 we can see the difference in proportions of special education students in the different groups. The highest concentration of students with special education needs is in the groups where ORF classified them as “at risk” as one would expect. For example, in the “at risk” groups, 41% and 30% of the students have special education needs, where only 12% and 10% of the students at “some risk” have. However it is important to notice two things. First, in the “at risk” and “some risk” classifications, there were higher percentages of special education students meeting or exceeding the standard than scoring below the standard, which is an unanticipated result. Second, the proportion of special education students in groups that denote false positive and false negative cases --groups 1 and 6-- is higher than the proportion in their counterpart groups --groups 4 and 3. More specifically the largest proportion of special education students is in group 1, which means they have been classified as “at risk” by the DIBELS ORF, but met the reading DSTP standards.

The disaggregation of the data within the six groups gives us a better picture of who are the students in each group. Likewise, it is of interest to see how the correlation coefficients of the measures ORF and reading DSTP change by disaggregating the data. In the next section, these issues will be explored.

✚ Correlations by groups

To restate, a significant correlation between ORF scores and reading DSTP scores ($r=.61$, $p<.01$) for all the students was found. The relation is graphically displayed in Figure 1. Below, Table 3 shows the correlations for students of different races, income status and special education placement.

Although all the correlations are significant, the correlation coefficients vary widely. For example the differences across races are very clear. The relationship between ORF scores and reading DSTP scores of third grade students in 2005 is stronger for Hispanic students than for everybody else. Also, the relation between these two scores is stronger in higher socioeconomic status level. Finally, the largest difference in correlation coefficients is present in the special education students.

Table 3. Correlation of disaggregated data

		n	Pearson Correlation
Race	African American	294	.536*
	Hispanic	43	.723*
	White	282	.624*
SES	Not low income	259	.643*
	Low income	371	.537*
Special education	Not special ed.	529	.579*
	Special ed.	96	.347*

*. Correlation is significant at the 0.01 level (2-tailed).

✚ What does “some risk” mean?

Students in the middle groups 2 and 5 were classified as being at “some risk” by the DIBELS ORF. But the meaning of “some risk” is not clear. Apparently they are not “at risk” per se, but they are not in the clear either. According to the results in the reading DSTP, if a student is classified as being at “some risk,” he or she is more likely to meet or exceed the DSTP standards in reading than scoring below. Almost 75% of the students classified as being at “some risk” met or exceeded the reading standard. The question is, can we bring the other 25% left behind to meet the standards?

FINAL REMARKS

The results of relationship between the performance on DIBELS ORF and the performance on the reading high-stakes test for the state of Delaware resemble the results of other states. However the purpose of this analysis was to go beyond the correlation coefficient by looking to the characteristics of students in different positions in relation to the two tests. We found that the total sample of 652 third grade Reading First students in the 2004-2005 year, approximately 72% met or exceeded the reading standards. The other 28% needs to be addressed in order to comply with the *No Child Left Behind Act*. We know that by the time the DSTP was taken, 16% out of the total sample (group 4) was still lacking basic skills such as reading fluency as measured by the DIBELS ORF. But we also know that 125 (groups 5 and 6) could have done better on the reading DSTP. Although this 12% seems to be a small number, we need to concentrate on them. It should be easier to pull their DSTP reading scores up than the scores of students classified “at risk.” Is targeted instruction the solution?

Other questions need answers as well. Why are the scores of certain students more correlated than others? Why does it seem that the DIBELS ORF is a better predictor of the reading DSTP for Hispanic students than everybody else? Why is it that the vast majority of false negative cases are African American students? Furthermore, why is it that the vast majority of false positive cases are low income students? Do testing accommodations moderate the correlational effects for students receiving special education services? How will changes in cut scores affect these correlations? How will these students perform on future state reading assessments? This paper’s aim was to describe the relationship between the performance on DIBELS ORF and the performance on the reading DSTP. It is from this description that new questions have arisen and the need for further investigation is revealed.

NOTES

1. In 2006, the third grade DSTP performance level cut scores were revisited and revised. All tables, figures, and textual comparisons reported here use the numbers of students who met or exceeded performance level scores as they existed in spring 2005. **Caution must be used** when considering any comparisons of this data with future years' DSTP results. Current and previous cut scores are reported by Delaware DOE at <http://www.doe.state.de.us/AAB/Cut%20Points%202006%20Marked%20Changes.pdf>.
2. DIBELS Oral Reading Fluency benchmarks can be viewed online at <http://dibels.uoregon.edu/benchmarkgoals.pdf>.

REFERENCES

- Barger, J. (2003). *Comparing the DIBELS oral reading fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Asheville, NC: North Carolina Teacher Academy.
- Buck, J., & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (FCRR Technical Report #1) Tallahassee, FL: Florida Center for Reading Research.
- Carlisle, J.F., Schilling, S.G., Scott, S.E., & Zeng, J. (2004). Do fluency measures predict reading achievement? Results from the 2002–2003 school year in Michigan's Reading First schools. Ann Arbor: University of Michigan.
- Delaware Department of Education. (2002). *Delaware's Reading First Grant Application*, (p.64). Available from <http://www.doe.k12.de.us/reading/rfol/RF%20Grant/READING%20FIRST%20APPLICATION.pdf>
- Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills-Modified. *School Psychology Review*, 30, 33-49.
- Good, R. H., Simmons, D., & Kame'enui, E. (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, 257-288.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review*, 32, 541-556.
- McGlinchey, M. T. & Hixson, M. D., (2004). Using Curriculum-Based Measurement to Predict Performance on State Assessments in Reading. *School Psychology Review*, 33, 193-203.
- Rasinski, T. (2003). *The fluent reader: Oral reading strategies for building word recognition, fluency & comprehension*. New York, NY: Scholastic Professional Books.

Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP)*. (Technical Report).

Sibley, D., Biwer, D., & Hesch, A. (2001). Establishing curriculum-based measurement oral reading fluency performance standards to predict success on local and state tests of reading achievement. Unpublished data. Arlington Heights, IL: Arlington Heights School District 25.

Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). *The relationship between oral reading fluency and Ohio proficiency testing in reading* (Technical Report). Eugene, OR: University of Oregon.

Wilson, J. (2005). The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency to performance on Arizona Instrument to Measure Standards (AIMS). Tempe, AZ: Tempe School District No. 3.