

Changes in Reading Achievement at a Florida Elementary: A Randomized Control Study of Reading Mastery

Technical Report 2011-3



Jean Stockard, Ph.D. Director of Research

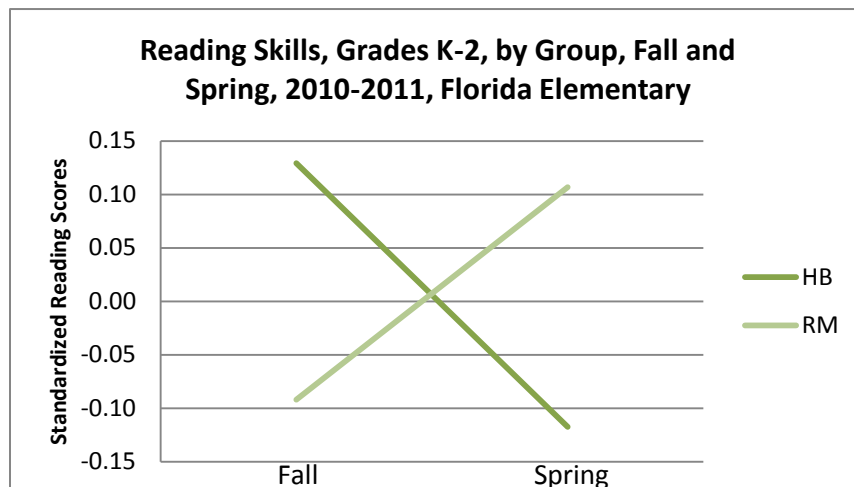
National Institute for Direct Instruction

October 2011

Executive Summary

During the 2010-2011 school year, students in grades K-2 at a Florida elementary school were randomly assigned to receive reading instruction in the Direct Instruction program *Reading Mastery (RM)*, Signature Edition, or in the Harcourt Brace (HB) program currently being used by the school. Training and support for the implementation of the *RM* curriculum were provided by the National Institute for Direct Instruction (NIFDI).

Data routinely collected by the district, including the *Florida Assessments for Instruction in Reading (FAIR)* and the *STAR Reading* assessment (first and second graders only), were used to assess changes over time in reading skills. Comparisons were made between students' scores in the fall, shortly after school began to their scores in the spring near the end of school. Similar findings appeared at all grade levels. Students assigned to RM began the school year with scores that were lower than the students assigned to HB. However, by the end of the school year, their scores were significantly higher. In other words, the gains made over time by the RM students were significantly greater than those made by the HB students. This pattern is shown in the figure below, with data combined across all grade levels. The advantage to the RM students appeared across all groups that are often seen as being at greater risk of having difficulties with reading: special education students, racial-ethnic minorities, students receiving free or reduced lunch, and boys.



Note: The scores are “standardized” scores, centered around an average value of zero for the total group and facilitating the combination of data across grades. Complete details on the analysis and results by grade level are available from the NIFDI research office.

Changes in Reading Achievement at a Florida Elementary School: A Randomized Control Study of Reading Mastery

During the 2010-2011 school year, students in grades K-2 at a Florida elementary school were randomly assigned to receive reading instruction in the Direct Instruction program *Reading Mastery (RM)*, Signature Edition, or in the Harcourt Brace (HB) reading program currently being used by the school. Training and support for the implementation of the RM curriculum was provided by the National Institute for Direct Instruction (NIFDI). Data routinely collected by the district, including the Florida Assessments for Instruction in Reading (FAIR) and the STAR Reading assessment (first and second graders only) were used to assess changes over time in reading skills. Comparisons were made between students' scores in the fall shortly after school began to their scores in the spring near the end of school. This report includes three sections. The first provides summary results for all students in the study (grades K-2). The second and third sections provide more detailed discussions of the measures that were used and the analyses, separating results for the kindergarten students from those for students in grades one and two. Results are illustrated in figures throughout the text and complete statistical results are given in tables at the end of the report.

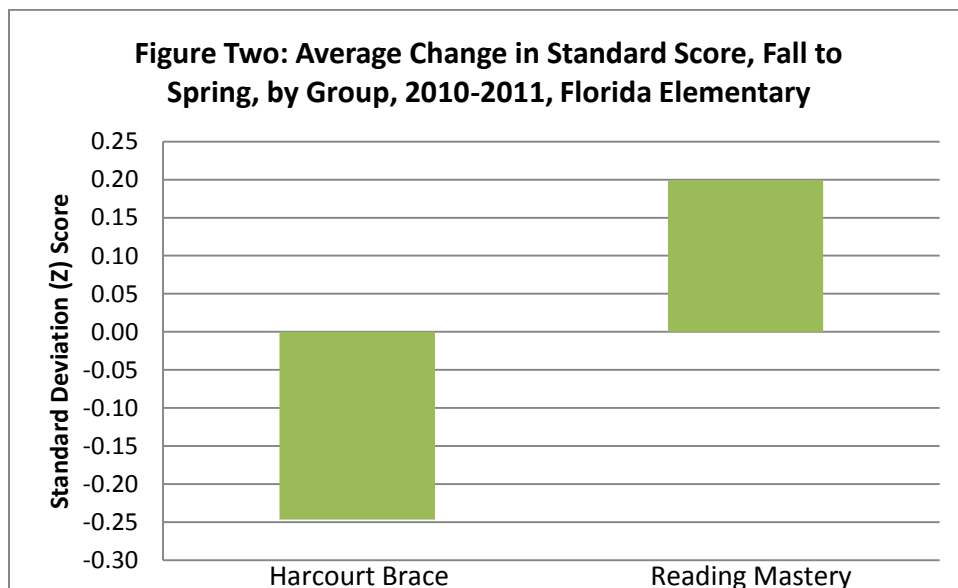
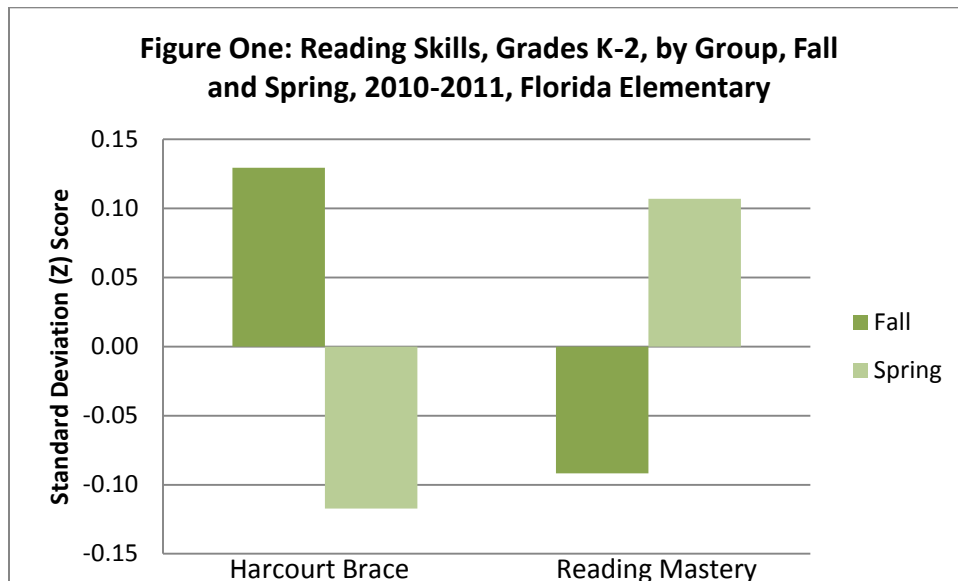
Similar findings appeared at all grade levels. Students using the Reading Mastery curriculum had significantly stronger growth in reading skills over time than those using Harcourt Brace. The results appeared across students with different individual characteristics (e.g. free or reduced lunch status, gender, race-ethnicity) and replicate those found in other studies of the Reading Mastery curriculum.

Summary Results, Grades K-2 Combined

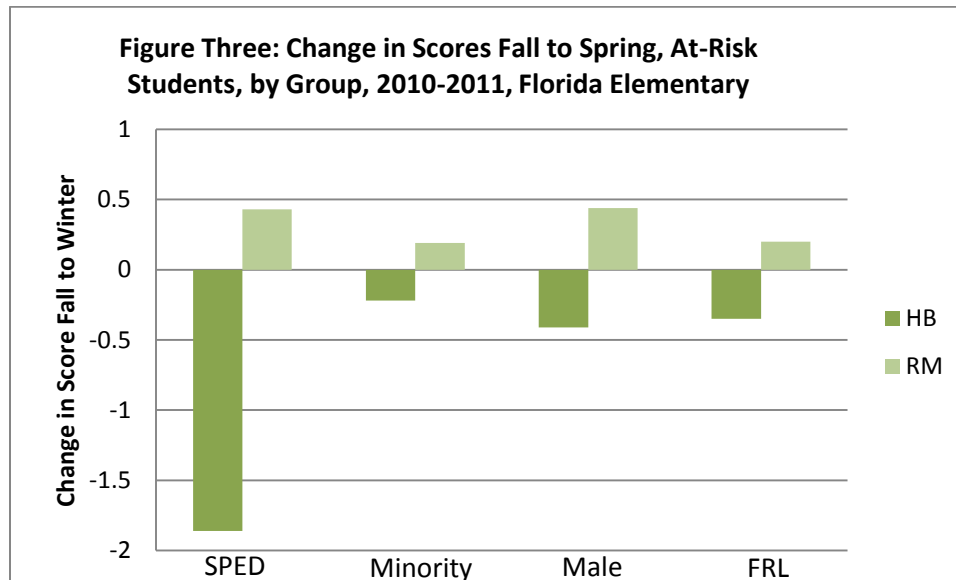
To provide comparable results across grade levels students' scores were transformed, using standard statistical methods, into standard deviation scores (also called z-scores), which are centered around zero. Students scoring higher than the average have positive standard scores, while those scoring lower than average have negative scores. The fall scores were based on the probability of success reported by the FAIR assessment and the spring scores were based on the probability of success from the FAIR (for kindergarten students) or the STAR reading assessment (for first and second graders). Data were available at both time points for 72 HB students and 79 RM students.

Students assigned to RM began the school year with scores that were slightly lower than the students assigned to HB. However, by the end of the school year their scores were significantly higher. In other words, the gains made over time by the RM students were significantly greater than those made by the HB students. These results are illustrated in Figures 1 and 2 and summarized in Table 1. As shown in Figure 1, in the fall (the blue columns) the HB students scored higher than the average and the RM students scored lower than average. By spring (the red columns), this had reversed, with the HB students scoring below the average and the RM students scoring above the average. Figure 2

summarizes the average change in scores over time (the difference in fall and spring scores). Compared to other students, the average HB student had declining reading skills, while the average RM student had increasing skills. Over the course of the school year, the average HB student's skills declined by almost one-fourth of a standard deviation unit, while the average RM student's skills increased by about one-fifth of a unit. Educational researchers often represent these differences as effect sizes, with effects greater than .25 seen as educationally important. The differences shown in Figures 1 and 2 represent an effect size of .45 and would thus be seen as educationally important. As shown in Table 1, the differences in change over time are also statistically significant. That is, they would not be expected to occur by chance.



It is also important to note that the advantage to the RM students appeared across all groups that are often seen as at greater risk of having difficulties with reading: special education students, racial-ethnic minorities, students receiving free or reduced lunch, and boys. Figure 3 shows the average change in scores over time for students in each of these categories. In all instances, the changes over time were stronger for the RM students than for the HB students. The associated effect sizes (shown in Table 2 at the end of this report) were substantial, ranging from .41 for the comparison for minority students to over 2.0 for special education students.

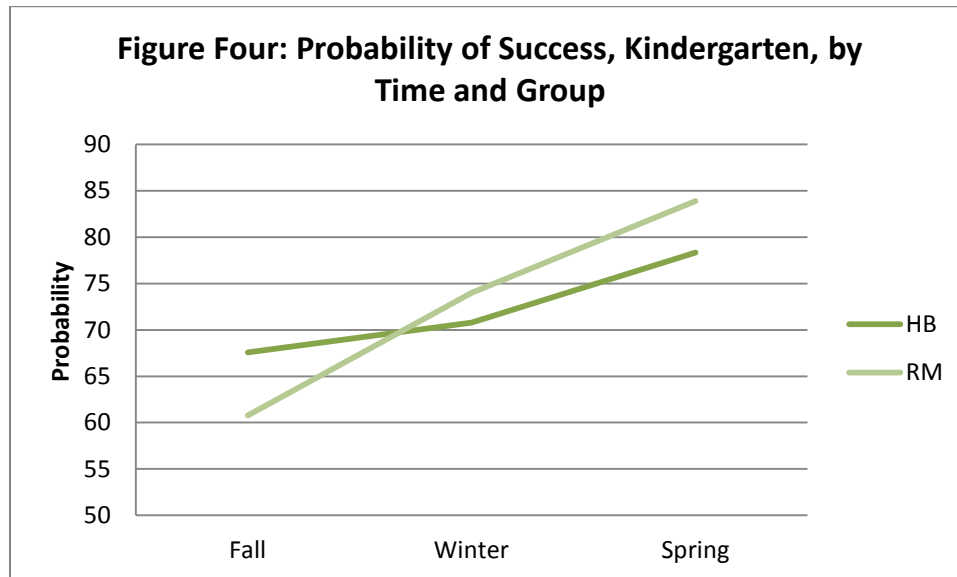


Kindergarten Students

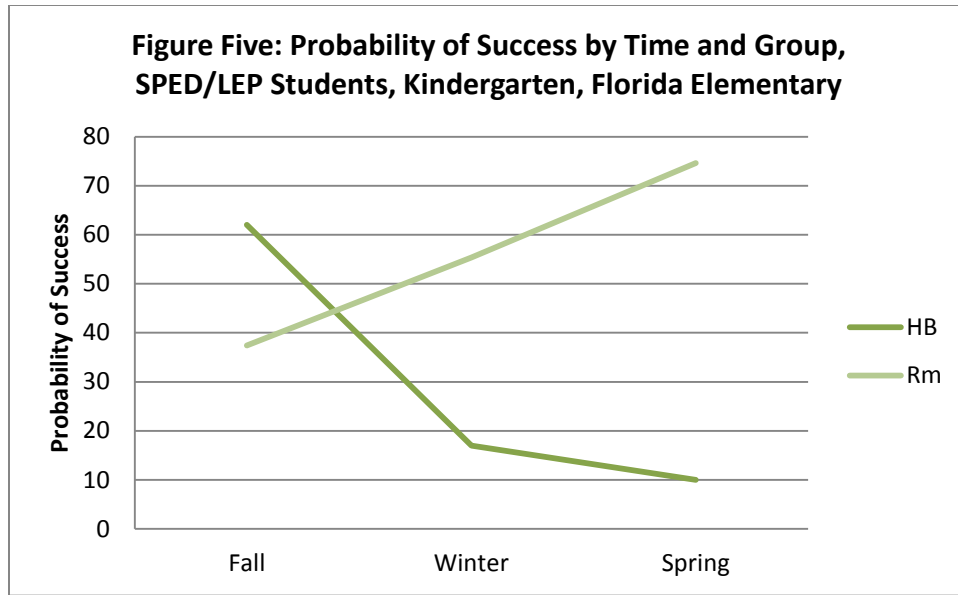
Data were available for both the fall and spring testing for 76 kindergarten students, 43 of whom (57%) were assigned to RM and 33 of whom (43%) were assigned to HB. The Florida Assessments for Instruction in Reading Test (FAIR) is administered individually to kindergarten students in Florida three times a year, and summary scores on this assessment were provided to the researcher. The FAIR is designed to monitor students' progress toward learning goals and identify students in need of additional help. The key summary score used is a derived measure of the students' probability of future success, defined as the probability that a student will "perform at the 40th percentile or better on the end-of-year gold standard outcome" (FAIR Technical Manual, 2009, p. 13). For kindergarten students the score is based on tests of letter naming and phonemic awareness in the fall, letter sounds and phonemic awareness in the winter, and phonemic awareness and word reading in the spring (FAIR, 2009, p. 4).

On average, all of the kindergarten students at the school had higher probabilities of success in the spring than in the fall of the year. However, the gain was significantly greater for the RM students than for the HB students. Figure 4 and Table 3 show these results. In the fall the RM students were less likely than the HB students to be predicted to succeed, but by winter and continuing through the spring they were more likely to be

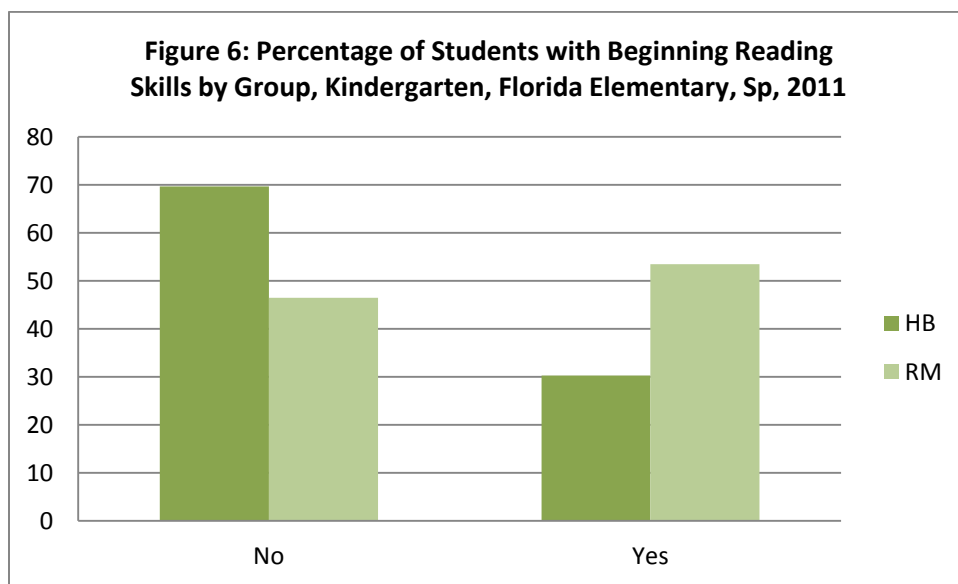
expected to succeed. The difference in the rate of gain was statistically significant (the significant interaction effect in the analysis of variance and the t-test associated with rate of change) and the effect size associated with the change (using the standard scores) was substantial ($d = .46$). In other words, at the start of school, the results favored the HB students, but the pattern had reversed by the spring, when results favored the RM students.



The lower performance of the RM group in the fall reflected the fact that this group had significantly more students with SPED or LEP classifications (an outcome that can occur by chance, even with random assignment). Eleven of the RM students (26% of the group) but only three of the HB students (9% of the group) had a special education and/or an LEP classification ($\chi^2 = 3.38, p = .07$). The differences in growth over time also appeared with these students who might be considered at risk given their special education or LEP status, as summarized in Table 4 and Figure 5. In the fall the SPED/LEP students in the HB group had, on average, a 62 percent probability of success, while the average for the RM group was 37 percent. By the winter testing this had changed dramatically, with sharp increases for the RM group and declines for the HB group. By the spring term the average RM SPED/LEP student had a 75 percent probability of success while the average HB SPED/LEP student had only a 10 percent probability. The analysis of variance results again indicated that the difference in rates of change was statistically significant.



The FAIR Assessment includes a reading test on instructional-level passages for kindergarten students who perform well on a screening tool, described as a “Word Placement List that has been linked empirically to 90% accuracy in the passages” (FAIR technical manual, 2009, p. 3). Thus, the only students who are given the reading passages are those who can demonstrate reading skills. The RM students were significantly more likely than the HB students to demonstrate these skills and be administered the reading passages (the red bars). As shown in Figure 6, slightly less than one-third of the HB students but more than two-thirds of the RM students passed this level. Among those who took the test, the RM students had slightly higher fluency scores and significantly higher accuracy scores. (See Table 5.)



To summarize, the results indicate that, on average, all students in the kindergarten program increased their probability of success over time. Yet, the kindergarten students who were assigned to the RM program had significantly greater growth in their reading skills, were twice as likely to have beginning reading skills by the end of the year, and were significantly more accurate in their reading. Special Education and LEP students in the RM program had significant gains over time, while the SPED and LEP students in HB had declining scores.

The analysis presented here may well understate the extent of the differences. We were unable to obtain the raw data that were used to determine the measure of probability of success. As noted above, this measure is a relatively low bar, indicating the probability of scoring at only the 40th percentile and appears to be relatively insensitive to variations among students with more advanced skills. For instance, among the students reported to have a probability of success of .88 or higher (78% of the total group), only half read well enough to take the advanced reading test. Among those with the highest ratings (probability = .91), 40 percent were non-readers. Among the HB students a majority of students at each of the two top probabilities (.89 and .91) did not pass the reading screen. Table 6 gives more details. The NIFDI Research Office hopes that a more complete data set will be made available and, when that occurs, additional analyses will be reported.

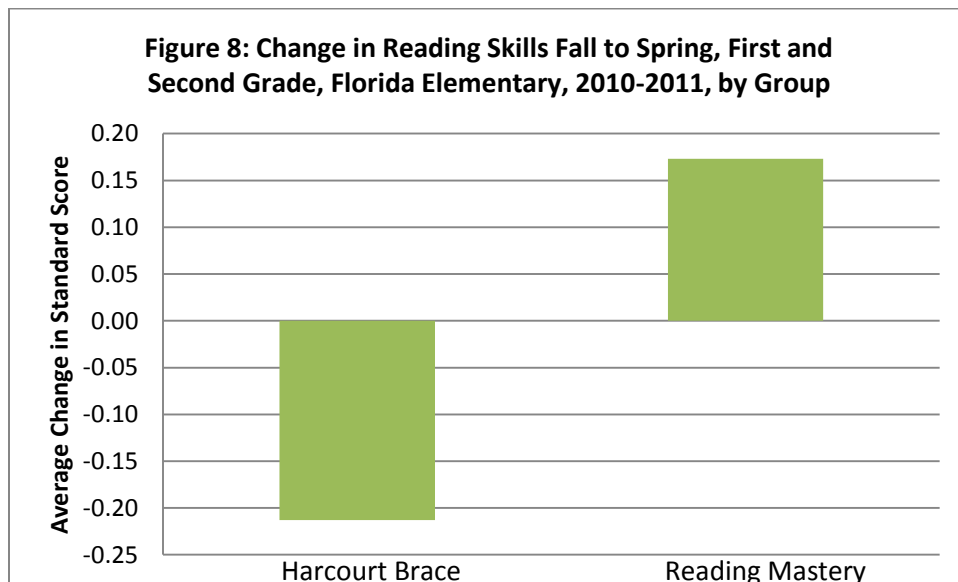
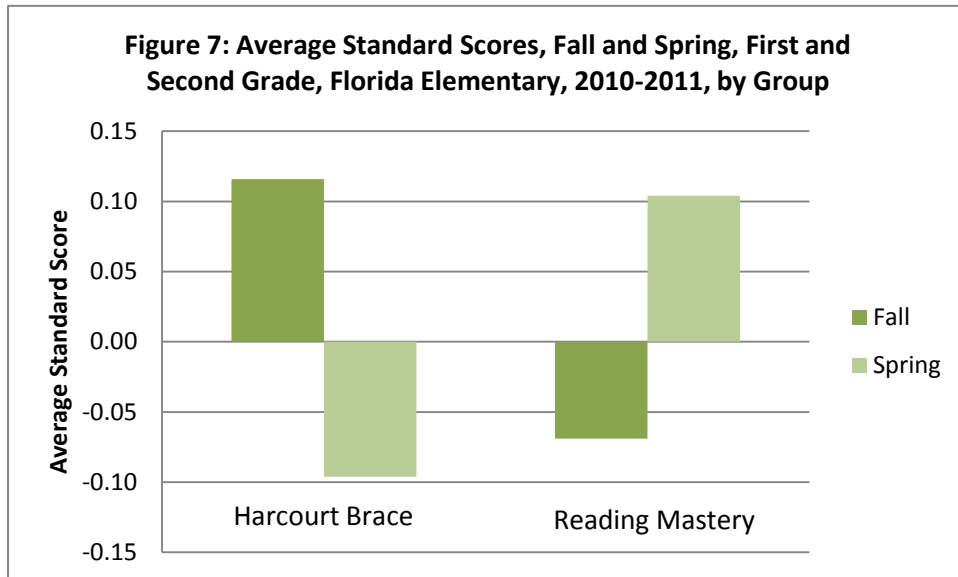
First and Second Grade Students

In the spring of 2010 first and second grade students at the Florida elementary school were randomly assigned to receive either *Reading Mastery* (RM) or continue in the Harcourt Brace (HB) curriculum. Students new to the school in the fall of 2011 were also randomly assigned to one group or the other. Because RM includes the option of cross-grade placement results and because these students had comparable reading measures, results for the two grades are combined.

Students' probability of success on the FAIR test was used as a pre-test measure to indicate reading skills at the start of school. As described above, this is a derived measure defined as the probability that a student will perform at the 40th percentile at the end of the school year. For first and second graders it is based on a short word reading task. Our measure of reading for the spring came from the STAR Reading test, a criterion referenced test, administered simultaneously to a whole classroom. We chose to use these test results rather than the FAIR because the STAR results are more detailed and precise. As recommended by statisticians, Normal Curve Equivalent (NCE) scores, which are adjusted for students' age and grade, were used. To provide comparability between the two tests (the FAIR and STAR) we transformed scores for each student to standard scores (also called z-scores), using the overall mean and standard deviation as the basis for these transformations. The resulting scores indicate the extent to which a given student's scores exceeded or fell below the average score for all students.

Data were available on both tests for 39 HB and 36 RM students. Results with the first and second graders are similar to those found with the kindergarteners. In the fall, the HB students had slightly higher scores than the RM students, but by the spring this pattern had reversed, with the RM students having higher scores than the HB students. The difference in the rates of change over time was statistically significant (as shown by the significant

interaction effect). On average, the HB students declined relative to their peers (an average change in z-scores of $-.21$), while the average RM student had increasing scores (an average change of $.17$ standard deviation units). The effect size associated with the different rates of change is $.46$, well beyond the level seen as necessary for educationally important results. These results are illustrated in Figures 7 and 8 and summarized in Table 7.



To summarize, results with first and second graders are similar to those obtained with the total group and with the kindergarten students. The students assigned to Reading Mastery had lower scores than the Harcourt Brace students at the start of the school year but higher scores at the end of the school year. The differences in rate of change over time were statistically significant and large enough to be considered educationally important.

Results for the Total Group are in Tables 1 and 2.

Table 1
Average Standard Scores, Fall and Spring, and Average Change Score, By Group, All Students, K-2

| | <u>Fall</u> | <u>Spring</u> | <u>Change</u> |
|-----------------|-------------|---------------|---------------|
| Harcourt Brace | 0.13 | -0.12 | -0.25 |
| Reading Mastery | -0.09 | 0.11 | 0.20 |
| t-test | -1.37 | 1.39 | 2.83 |
| prob. | 0.91 | 0.08 | 0.003 |
| Effect size | -0.22 | 0.22 | 0.45 |

Note: Probabilities are one-tail, reflecting the research hypothesis that the RM group would have higher scores than the HB group. N = 72 for HB, 79 for RM. Repeated Measures analysis of variance results were F (group) = 0.00, p = .99; F (time) = .09, p = .76; F (interaction) = 8.03, p = .005; df = 1, 149 for each F ratio.

Table 2
Average Change Scores by At Risk Status and Curriculum Group, Total Sample (K-2), Florida Elementary, 2010-2011

| | | <u>HB</u> | <u>RM</u> | <u>Effect Size</u> |
|----------|-----|-----------|-----------|--------------------|
| SPED | no | -0.1 | 0.16 | 0.26 |
| | yes | -1.86 | 0.43 | 2.29 |
| Minority | no | -0.28 | 0.2 | 0.48 |
| | yes | -0.22 | 0.19 | 0.41 |
| Male | no | -0.04 | 0.02 | 0.06 |
| | yes | -0.41 | 0.44 | 0.85 |
| LEP | no | -0.2 | 0.16 | 0.36 |

| | | | | |
|-----|-----|-------|------|------|
| | yes | -3.4 | 0.75 | 4.15 |
| FRL | no | 0.11 | 0.2 | 0.09 |
| | yes | -0.35 | 0.2 | 0.55 |

Scores are the average change in standard score from fall to spring.

Results for kindergarten students are in Tables 3 through 6.

Table 3

Probability of Success by Group, Fall, Winter, and Spring, Kindergarten Students, Florida Elementary

Descriptive Statistics by Term and Comparisons Across Groups

| | Fall | | Winter | | Spring | |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | <u>HB</u> | <u>RM</u> | <u>HB</u> | <u>RM</u> | <u>HB</u> | <u>RM</u> |
| Mean | 67.6 | 60.8 | 70.8 | 74.0 | 78.3 | 83.9 |
| s.d. | 24.3 | 28.1 | 27.7 | 25.8 | 26.0 | 18.6 |
| t-value | 1.11 | | 0.52 | | 1.09 | |
| prob. (2-tail) | 0.27 | | 0.60 | | 0.28 | |
| common s.d. | 26.5 | | 26.5 | | 22.09 | |
| Cohen's d | -0.26 | | 0.12 | | 0.25 | |

Repeated Measures Analysis of Variance

| | F | df | prob |
|---------------|-------|--------|--------|
| Group | 0.02 | 1, 148 | 0.89 |
| Time | 20.21 | 2, 148 | <.0001 |
| Group by Time | 3.03 | 2, 148 | 0.05 |

Note: The t-tests examine the null hypothesis that the average for the two groups in each time period is the same. The analysis of variance tests the null hypothesis that the change over time for the two groups was the same.

Table 4

Probability of Success by Group, Fall, Winter, and Spring, SPED/LEP Kindergarten Students, Florida Elementary

| <i>Descriptive Statistics by Term and Comparisons Across Groups</i> | | | | | | |
|---|-----------|-----------|-----------|-----------|-----------|-----------|
| | Fall | | Winter | | Spring | |
| | <u>HB</u> | <u>RM</u> | <u>HB</u> | <u>RM</u> | <u>HB</u> | <u>RM</u> |
| Mean | 62.0 | 37.4 | 17.0 | 55.4 | 10.0 | 74.64 |
| s.d. | 0.0 | 22.7 | 6.9 | 31.6 | 5.0 | 26.21 |
| t-value | 3.6 | | 3.71 | | 7.68 | |
| prob. (2-tail) | 0.005 | | 0.003 | | <.0001 | |
| common s.d. | 22.5 | | 32.3 | | 35.91 | |
| Cohen's d | -1.09 | | 1.19 | | 1.80 | |
| <i>Repeated Measures Analysis of Variance</i> | | | | | | |
| | F | df | prob | | | |
| Group | 4.45 | 1, 24 | 0.06 | | | |
| Time | 1.13 | 2, 24 | 0.34 | | | |
| Group by Time | 13.04 | 2, 24 | 0.0001 | | | |
| <p>Note: The t-tests examine the null hypothesis that the average for the two groups in each time period is the same. A t-test that adjusted for unequal variances for the two groups was used. The analysis of variance tests the null hypothesis that the change over time for the two groups was the same.</p> | | | | | | |

Table 5

Reading Fluency Tests by Group, Kindergarten Students, Spring, Florida Elementary

Passed Screening Test to Take Fluency Assessment

| | <u>HB</u> | <u>RM</u> | <u>chi-square</u> | <u>prob.</u> |
|-------|-----------|-----------|-------------------|--------------|
| No | 69.7 | 46.5 | 4.08 | 0.04 |
| Yes | 30.3 | 53.5 | | |
| Total | 100 | 100 | | |

Fluency Score

| | <u>HB</u> | <u>RM</u> | <u>t-value</u> | <u>prob.</u> | <u>d</u> |
|------|-----------|-----------|----------------|--------------|----------|
| Mean | 29.3 | 30 | 0.14 | 0.89 | 0.06 |
| s.d. | 9.5 | 13.2 | | | |

Accuracy Score

| | <u>HB</u> | <u>RM</u> | <u>t-value</u> | <u>prob.</u> | <u>d</u> |
|------|-----------|-----------|----------------|--------------|----------|
| Mean | 92.7 | 94.7 | 2.25 | 0.03 | 0.80 |
| s.d. | 2.2 | 2.4 | | | |

Note: Probabilities associated with the t-values are two-tail. The effect size, Cohen's d, is the difference between the means divided by the common standard deviation. For the analyses of fluency and accuracy there were 10 HB students and 23 RM students.

Table 6

Probability of Success Scores and Reading Skills, Kindergarten Students, Florida Elementary, 2010-2011

| <i>Total Group</i> | | | | | |
|-------------------------------|-----------------|-------------|-------------|-------------|--------------|
| <u>Probability of Success</u> | | | | | |
| <u>Passed Reading Screen</u> | <u>< .87</u> | <u>0.88</u> | <u>0.89</u> | <u>0.91</u> | <u>Total</u> |
| No | 100.0 | 57.1 | 60.0 | 40.4 | 56.6 |
| Yes | 0.0 | 42.9 | 40.0 | 59.6 | 43.4 |
| total | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| n | 17 | 7 | 5 | 47 | 76 |
| <i>HB Students</i> | | | | | |
| <u>Probability of Success</u> | | | | | |
| <u>Passed Reading Screen</u> | <u>< .87</u> | <u>0.88</u> | <u>0.89</u> | <u>0.91</u> | <u>Total</u> |
| No | 100.0 | 0.0 | 66.7 | 60.0 | 69.7 |
| Yes | 0.0 | 100.0 | 33.3 | 40.0 | 30.3 |
| total | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| n | 9 | 1 | 3 | 20 | 33 |
| <i>RM Students</i> | | | | | |
| <u>Probability of Success</u> | | | | | |
| <u>Passed Reading Screen</u> | <u>< .87</u> | <u>0.88</u> | <u>0.89</u> | <u>0.91</u> | <u>Total</u> |
| No | 100.0 | 66.7 | 50.0 | 25.9 | 46.5 |
| Yes | 0.0 | 33.3 | 50.0 | 74.1 | 53.5 |
| total | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| n | 8 | 6 | 2 | 27 | 43 |

Results for first and second graders are in Table 7.

Table 7

Average Standard Scores, Fall and Spring, and Average Change Score, By Group, First and Second Grade, Florida Elementary, 2010-11

| | <u>Fall</u> | <u>Spring</u> | <u>Change</u> |
|-----------------|-------------|---------------|---------------|
| Harcourt Brace | 0.12 | -0.10 | -0.21 |
| Reading Mastery | -0.07 | 0.10 | 0.17 |
| t-test | 0.81 | 0.87 | 2.05 |
| prob. | 0.79 | 0.19 | 0.02 |
| Effect size | -0.19 | 0.20 | 0.46 |

Note: Probabilities are one-tail, reflecting the research hypothesis that the RM group would have higher scores than the HB group. N = 39 for HB, 36 for RM. Repeated Measures analysis of variance results were F (group) = 0.00, p = .97; F (time) = .04, p = .84; F (interaction) = 4.20, p = .04; df = 1, 73 for each F ratio.