ASU ARIZONA STATE UNIVERSITY

FINAL SUMMARY REPORT: EVALUATION STUDY OF *THE WRITING ROAD TO READING*

Gary Bitter, Mary Aleta White
Arizona State University

Arizona State University researchers conducted a four-year quasi-experimental study that spanned from 2006-2007 through 2009-2010. This study involved 5 experimental and 6 control schools with an average of 47 teachers and an average of 1,000 students each year. The following research questions guided the study design and methods:

1. Do children who participate in *The Writing Road to Reading* program demonstrate significant learning gains in reading skills?
2. How does the reading skill attainment of children participating in Spalding's *Writing Road to Reading* compare to that of children participating in other, more traditional reading programs?
3. How well do teachers implement *The Writing Road to Reading* in their varied classrooms?

METHOD

To study teacher implementation, researchers utilized a uniform quantitative instrument to measure how *The Writing Road to Reading* was being implemented in the experimental classrooms. To measure program implementation, researchers collected data through classroom observations using observation protocols designed to measure constructs such as classroom management, adherence to program philosophy, and strategies for spelling, writing, and reading content. Both experimental and control teachers also completed a survey questionnaire that provided a variety of background information including degrees, certifications, endorsements, professional development activities over the past ten years, length of time implementing reading programs, materials used, assessment practices, and the number of years teaching at the current level.

For student measures, researchers employed the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) as the primary measure to assess changes in students' reading skills during the study. Researchers selected DIBELS as the assessment because it has broad visibility and acceptance in the field; it demonstrates technical merit; and it was adopted by the Arizona Department of Education. DIBELS tests were administered at the beginning, middle, and end of each school year. In the third year (grade 2), TerraNova standardized reading test results were also compared because Arizona law required testing all second-graders.

MARY LOU FULTON COLLEGE OF EDUCATION
TECHNOLOGY BASED LEARNING AND RESEARCH
ASU SkySong
1475 N. Scottsdale Rd., Suite 200 Scottsdale, AZ 85257-3538
Phone # (480) 884-1694 Fax # (480) 884-1888
http://tblr.ed.asu.edu

Participants

This study was conducted in 11 diverse Arizona schools: 5 experimental and 6 control schools with an average of 47 teachers and an average of 1,000 students each year. Complete descriptions and information are provided with Tables 1-5 in the complete report available on <www.spalding.org/research>. Classes in the experimental group (average of 23 classrooms per year) used the Spalding curriculum an average of 90 minutes each day, while control classes (average of 24 classrooms per year) used either Houghton Mifflin or Harcourt an average of 80 minutes each day. Class size averaged 25 students in the experimental group and 20 in the control group. Experimental and control students were matched by gender, ethnicity, socioeconomic status, and language ability. The experimental students can be further separated into two groups: those who were in the kindergarten study and *all* treatment students in the grade level. Experimental and control groups were compared for attrition; experimental and control schools that had a high number of English Language Learner students experienced the highest rate of student attrition over the four years, mostly do to recent Arizona state legislation impacting Mexican-American immigration.

Student Performance Results

Table 6 displays the comparative performance of the experimental and the control students on the DIBELS measures administrated in each benchmark assessment over the course of the study. In each administration, Spalding students had consistently higher *mean* values on all DIBELS areas which provides evidence that Spalding has been more effective than the methods used in the control schools to teach those reading skills.

*Table 6: Descriptive statistics, all students**

			Experimental		Control	
		Measure[+]	Mean	SD	Mean	SD
Year 1, Kindergarten	Fall, 2006	ISF	10.80**	10.35	7.32	8.052
		LNF	15.81**	16.17	10.79	13.92
		WUF	12.01**	15.59	3.23	7.51
	Winter, 2007	ISF	18.99	12.61	17.37	13.847
		LNF	32.75**	19.02	28.44	19.89
		PSF	27.77**	17.96	20.21	16.85
		NWF	26.61**	18.26	20.28	22.03
		WUF	22.89**	18.11	9.51	13.32
	Spring, 2007	LNF	47.97**	18.117	44.39	20.812
		PSF	47.68**	16.51	39.62	19.52
		NWF	46.17**	25.77	35.36	25.25
		WUF	39.91	18.12	26.92	17.69

* p=<.05, ** p =<.01

[+] Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Word Use Fluency (WUF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Oral Reading Fluency (ORF), Retell Fluency (RF).

Table 6: Descriptive statistics, all students cont'd*

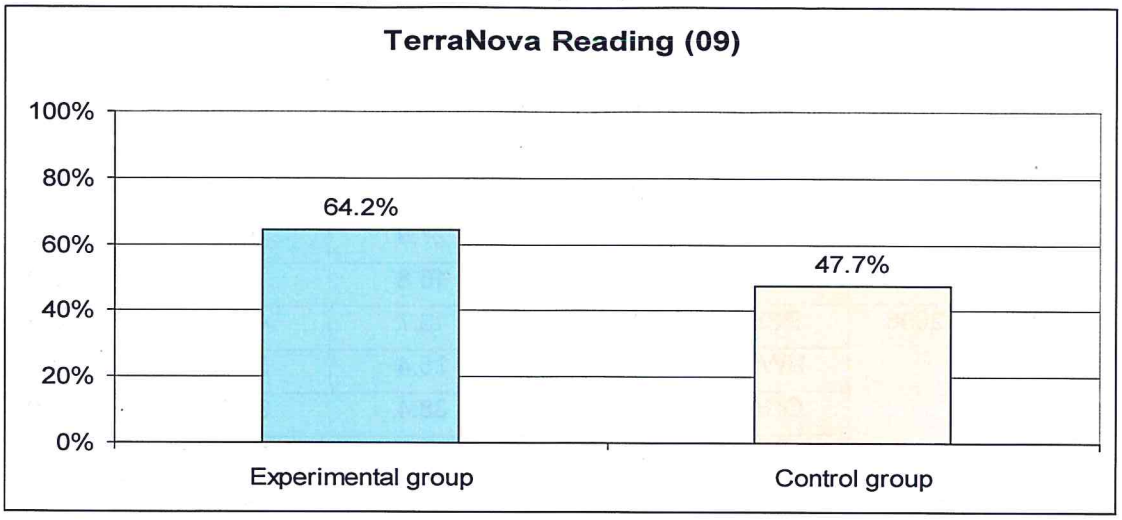
			Experimental		Control	
		Measure[+]	Mean	SD	Mean	SD
Year 2, First Grade	Fall, 2007	LNF	41.54**	18.7	42.18	18.7
		PSF	45.38**	16	35.55	16.8
		NWF	45.75**	27.9	34.7	25.8
		WUF	35.84**	16.8	18.8	15.4
	Winter, 2008	PSF	48.2**	13.7	45.31	15.5
		NWF	57.41**	35.4	56.73	30
		ORF	49.26**	38.4	31.36	32.2
		RF	17.98**	15.8	11.82	15.4
	Spring, 2008	PSF	51.44**	13.6	48.95	14.2
		NWF	71.02*	39.8	72.69	33.2
		ORF	64.57**	39.9	53.51	39.3
		RF	23.34**	16.7	19.76	13.4
Year 3, Second Grade	Fall, 2008	NWF	84.88*	34.36	67.27	33.59
		WUF	44.5*	15.76	32.59	15.74
		ORF	73.63*	37.93	49.71	34.78
		RTF	28.59*	15.76	15.42	12.23
	Winter, 2009	ORF	98.59*	38.01	66.33	38.03
		RTF	38.72	17.36	25.39	16.71
	Spring, 2009	ORF	109.96**	37.58	87.48	42.73
		RTF	44.83**	17.47	33.47	21.03
Year 4, Third Grade	Fall, 2009	ORF	96.99**	37.11	72.03	37.54
	Winter, 2010	ORF	110.68**	38.23	88.14	41.14
	Spring, 2010	ORF	120.55**	36.28	105.65	39.18

* p=<.05, ** p =<.01

[+] Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Word Use Fluency (WUF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Oral Reading Fluency (ORF), Retell Fluency (RF).

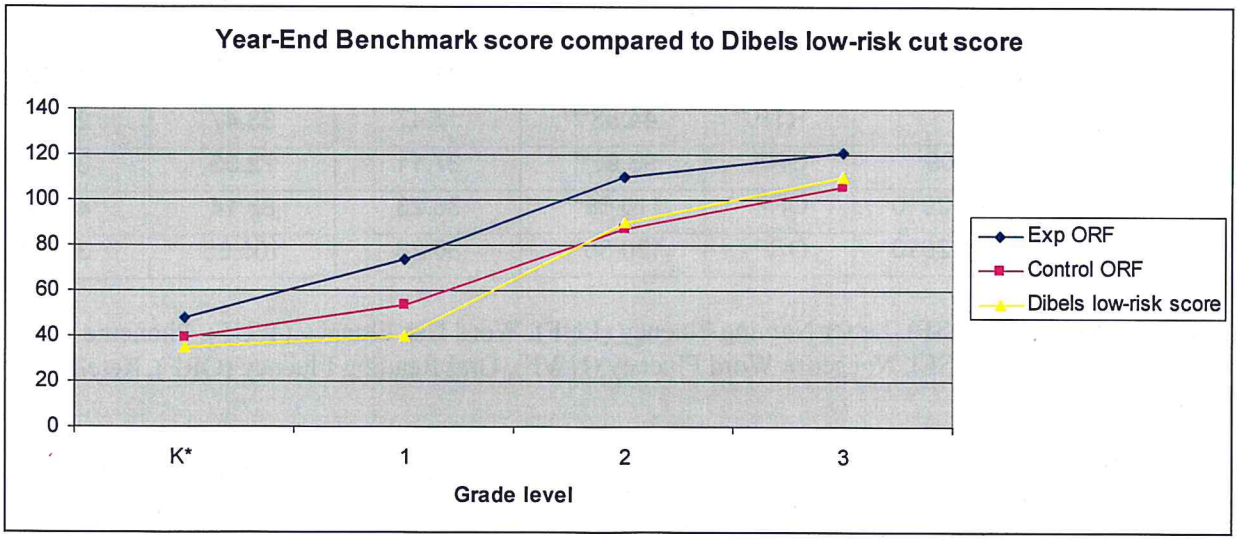
Another analysis of reading achievement was available in the third year's study because all second grade students are required to complete the state's norm-referenced achievement test, TerraNova. Chart 3 below represents a sample of the study students (three control and three experimental schools) and their average NCE score on the TerraNova reading portion. As would be expected from reviewing the DIBELS scores, the Spalding students' NCE scores were significantly higher than the control students on the state test ($p < .01$).

Chart 3: *Student NCE reading scores from Spring, 09 AZ TerraNova exam*



As shown in Chart 4 below, additional analyses of the extent to which experimental students experienced learning gains by the end of each grade level show that they exceeded the DIBELS decision rules benchmarks for achievement each year (as well as at each testing period). Unfortunately, since the middle of their second grade year, the average control student is not meeting the DIBELS assessment for low-risk scores. Their achievement level averages a score approximately 5 points below the low risk threshold.

Chart 4: Overall, year-end assessment scores compared to DIBELS decision rules benchmarks



**K, Spring 2007 is PSF, all others are ORF*

In addition to measures of statistical significance, researchers frequently calculate and report measures of practical significance, known as the effect size, to help educators decide whether a statistically significant difference between programs translates into a meaningful difference—one that would justify a program adoption for instance. One commonly used measure of effect size is called Cohen's d. Researchers computed Cohen's d using a pooled standard deviation for DIBELS Oral Reading Fluency scores at the end of each benchmark assessment. Using the entire

treatment group, the effect size for Fall 2009 was .7; for Winter 2010 it was .6; and in Spring 2010 the effect size is .4. This means that the intervention has a positive, medium effect (average of 0.54) on student achievement. Converting the effect size to percentiles would mean the average student in the Spalding sample, at the end of the year, would score higher than 69% of the control sample even though some students entered the program in the final year.

Using scores from the longitudinal group, the effect size for Fall 2009 was .9; for Winter 2010 was .8; and for Spring 2010 was .6. This means that the intervention has a large effect size (average of 0.78) on student reading achievement. In converting effect size to percentiles, the average student in the longitudinal Spalding group would score higher than 79% of the control student sample.

SUMMARY

According to the year four results, students who used *The Writing Road to Reading* demonstrate higher and statistically significant learning as measured by DIBELS. Since both the control groups and the experimental groups used detailed teacher guides evaluated by Arizona Department of Education for research-based reading components, theoretically, they should have produced similar results. This was not the case. The four-year findings strongly suggest that use of *The Writing Road to Reading* program is an effective method for enhancing performance on critical early literacy skills.