Effects of Explicit, Multisensory, Structured Language Arts Instruction Compared to Conventional Reading Instruction

Gary Bitter
Mary Aleta White
Arizona State University

Author Note
Dr. Gary Bitter, Technology Based Learning & Research (TBLR)
Correspondence concerning this paper should be addressed to Gary Bitter, bitter@asu.edu

Abstract

The purpose of this four-year study was to determine the effectiveness of explicit, multisensory, structured language arts instruction, compared to conventional reading instruction. Five Arizona experimental and six control schools with an average of 1,000 general education kindergarten through third-grade students participated each year. Over the four-years, experimental group students showed higher mean values on each administration of DIBELS measures demonstrating an average medium effect (0.54) on student achievement. Experimental students in the four-year cohort scored highest (average effect size 0.78). A possible explanation is that the experimental program integrated explicit, multisensory, structured language arts instruction with research-based components of skilled reading.

Effects of Explicit, Multisensory, Structured Language Arts Instruction

Compared to Conventional Reading Instruction

During the last four decades, research focused on how children learn to read and the complex process of reading. Herron (2008) states that early instruction determines how the brain organizes itself for reading. She asserts that children's first experiences with letters and words dictate how the brain establishes neural networks that may become habitual pathways as reading skills develop. She references Shaywitz (2003) and other researchers who have used imaging technology such as functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) to map primarily left hemisphere brain networks that encompass the functions for reading: pronunciation, meaning, and the visual appearance of words. Herron's key premise is that conventional phonics programs teach reading backwards—they go from print to speech rather than speech to print.

Dehaene (2009), in *Reading in the Brain,* provides further insight into the way brain structure determines reading processes:

> Considerable research, both with children and with illiterates, converges in the fact that grapheme-phoneme conversion radically transforms the child's brain and the way in which it processes speech sounds. This process whereby written words are converted into strings of phonemes must be taught explicitly. It does not develop spontaneously and must be acquired (p 219).

Dehaene credits the neurologist Samuel Torrey Orton as “the founding father of the psychology of reading and dyslexia” (p. 207). Although some of Orton’s neurological explanations have not held, his preliminary work on appropriate reading instruction has been validated. Farnham-Diggory (1992) says Orton "foresaw many principles of contemporary neuropsychology that awaited new technology for their verification" (p. 297). Orton's approach to reading instruction was the foundation for multisensory language instruction. The International Dyslexia Association (2000), formerly the Orton Dyslexia Society, elaborated on Orton’s explanation: “Multisensory teaching is simultaneously visual, auditory, and kinesthetic-tactile to enhance memory and learning” (p. 1). Although Orton focused on the difficulty some students had making the connection between print and speech sounds, his approach has been revised and expanded to include additional components of the reading process. Current multisensory language instruction focuses on reading comprehension after decoding skills and vocabulary knowledge are developed. Early research on the effectiveness of these adaptations was published in the *Annals of Dyslexia,* a publication of the International Dyslexia Association (e.g., Cox, 1985; Joshi, Dahlgren, & Boulware-Gooden, 2002; North, 1992; Ogden, Hindman, & Turner, 1989; Vickery, Reynolds, & Cochran, 1987). McIntyre and Pickering (1995) compiled results of eleven multisensory structured language approaches developed over the past 50 years: Alphabetic Phonics, The Association Method, The Herman Approach, The Lindamood-Bell Learning Processes,

The Montessori and Sequential English Education Approach, The Orton-Gillingham Approach, Project Read, The Slingerland Approach, The Spalding Approach, Starting Over, and The Wilson Approach. Although these multisensory approaches vary based on audience, classroom setting, and materials, they agree that instruction of the evidence-based components must be systematic, sequential, explicit, cumulative and use visual, auditory, kinesthetic, and tactile senses for teaching reading (Joshi, Dahlgreen, and Boulware-Gooden, 2002).

An approach that focuses on the progressive use of sounds, syllables, words, sentences, and written text is in stark contrast to the literature-based approach which is based on the premise that reading is a single process of obtaining meaning from print and that phonics should be taught implicitly if at all. This debate over skill-based or literature-based instruction sparked an increase in research on the reading process, essential components of reading, and whether explicit or implicit teaching of these components resulted in improved student learning. The Report of the National Reading Panel (2000) identified phonemic awareness, phonics, vocabulary, text fluency, and comprehension strategies as essential components of reading. Rayner, Foorman, Perfetti, Pesetsky, and Seidenberg (2001) compared literature-based instruction with skills-based instruction. They reported that "mastering the alphabetic principle is essential to becoming proficient in the skill of reading, and … instructional techniques (namely, phonics) that teach this principle directly are more effective than those that do not" (p. 68). The Learning Point

Associates report (2004) supported these previous studies and recommended explicit and structured instruction in the five essential components. Although explicit, structured teaching of phonics was validated in these studies, the contribution of multisensory instruction was not studied.

The past decade, however, brought a shift in research investigating multisensory processes. Calvert, Spence, and Stein (2004) compiled current research from neuroscience and psychology in *The Handbook of Multisensory Processes*. This research focuses on how the brain synthesizes information from the different senses. Previous research investigated sense modality independently, but there is increasing research from different areas in the emerging field of multimodal or multisensory integration. Shams and Seitz (2008) explain "Studies of learning, and in particular perceptual learning, have focused on learning of stimuli consisting of a single sensory modality" (p. 1). However, new findings have overruled the concept that different sensory modalities operate independently. Citing findings by Shimojo and Shams (2001), Shams and Seitz note that multisensory interactions in many different perceptual tasks and settings "are the rule rather than the exception" (p. 1). Unlike the previous "sense by sense" approach, these studies recognize that perception is fundamentally a multisensory experience.

In addition to the work on multisensory processes, there has been an increase in the number of studies examining multisensory instruction. According to Campbell, Helf, and Cooke (2008), a number of studies that included explicit, systematic phonics and

multisensory instruction demonstrated positive effects on literacy: Dev, Doyle, & Valente, 2002; Dickson & Bursuck, 1999; Oakland, Black, Stanford, Nussbaum, & Balise, 1998). Campbell and colleagues conducted their own study to determine whether multisensory activities added to a supplemental reading program increased the ability of six young treatment resisters to master decoding skills. The results indicated that the fluency of decoding nonsense words increased when multisensory components were added to the supplemental reading intervention.

A number of studies compared specific multisensory approaches or programs with conventional basal reading instruction. Joshi, Dahlgren, and Boulware-Gooden (2002) examined the effectiveness of teaching reading skills to twenty-four first-grade children in an inner city school using “Language Basics: Elementary,” an Alphabetic Phonics program based on the Orton-Gillingham approach. Students taught with the multisensory teaching approach made statistically significant gains on tests of phonological awareness, decoding, and reading comprehension, but the thirty-two control students using the Houghton-Mifflin Basal Reading Program made gains only in reading comprehension. "Children in the control groups were not taught phonics skills in a systematic and explicit fashion, and they did not show any significant gains in phonological awareness and decoding skills" (p. 238).

Scheffel, Shaw, J., and Shaw (2008) evaluated the efficacy of a supplementary Orton-Gillingham reading program. This study included 224 first-grade students in the

treatment group and 476 first-grade students in the control group. All students received ninety minutes of instruction in the district's core reading program. The treatment group received an additional thirty minutes per day of an Institute of Multi-Sensory Education (IMSE) reading program. "First-grade students in the three treatment group schools made the most dramatic improvement in acquiring alphabetic principle skills" (p. 3).

Despite the extensive use of multisensory approaches, there has been little empirical evidence that it is the multisensory component that provides statistically significent difference. (Campbell, Helf, & Cooke, 2008; Dickson & Bursuck, 1999; Greene, 1996 Joshi et al., 2002; Moats & Farrell, 1999) Given the number of children across the country and internationally who are enrolled in multisensory language instruction courses, Ritchey and Goeke (2006) call for studies that help provide and "establish sufficient evidence" to show that the approach does meet "the requirements of scientifically-based reading instruction" (p. 172). Furthermore, almost all of the previous research on multisensory approaches were conducted in clinical, small-group settings with special populations.

One multisensory approach, commonly referred to as the Spalding method, had previously been investigated in both special (Farnham-Diggory, 1992; Hoerl & Koons, 1995, North, 1995; Aaron, Joshi, Gooden, & Bentum, 2008) and large, general education settings (Aukerman, 1984; North, 1992). Prior data collection, clinical and quasi-experimental studies are summarized in *The Writing Road to Reading* (Spalding, R. B.,

2003). Spalding studied under Orton in the late 1930s. She learned that the method of teaching determines which neural pathways develop in the brain. After observing that Orton's techniques worked successfully with children experiencing reading difficulties, she developed a multisensory method of language arts instruction designed to prevent as well as remediate reading problems. In 1957, Spalding published *The Writing Road to Reading,* which posits that children should be explicitly taught sound-letter correspondences by saying the sound and writing the letter or letters. Children then say sounds and write high-frequency words, then sentences and read aloud their own writing—thus the title of the book. This program integrates explicit, multisensory, interactive, and diagnostic instruction with the five research-based components.

The purpose of this study was to investigate the effectiveness of explicit, multisensory, structured language arts instruction compared to conventional reading instruction. All three programs in the study had been evaluated by the Arizona Department of Education as including the five research-based components identified by the National Reading Panel (NRP). Thus, these components were eliminated as variables. The four-year study design and methods were guided by the following predictions:

1. Children who receive explicit, multisensory, structured language arts instruction will demonstrate significant learning gains in reading skills.

2. The reading skill attainment of children who receive explicit, multisensory, structured, language arts instruction will exceed that of children participating in other, more conventional reading programs.

The specific multisensory strategies in the experimental program and the extent to which teachers accurately and consistently implementated them in their classrooms were also investigated.

Method

This section describes participants in the study, the measures, and the design and procedures.

Participants

This four-year study was conducted in eleven diverse Arizona schools: five experimental and six control schools with an average of forty-seven teachers and an average of 1,000 students each year. The experimental group averaged twenty-three classes per year with an average of twenty-five students in each. The control group averaged twenty-four classes per year with an average of twenty students in each. Table 1 indicates the targeted schools and the number of participating students in the first benchmark test of each year.

Table 1

Schools Included in the Analysis

Group	Name of School	# Total K Students	# Total 1st graders	# Total 2nd graders	# Total 3rd graders
Experimental	Alhambra	101	88	76	85
	Bret Tarver	148	126	128	151
	CTA-Liberty	124	138	141	132
	Gallego	101	92	90	104
	Valley Academy	123	104	105	104
Total	5	**597**	**548**	**540**	**576**
Control	#1 – M	116	92	103	84
	#2 – N	131	104	104	88
	#3 – O	102	86	84	79
	#4 – P	101	70	22	104
	#5 – Q	128	121	111	107
	#6 – R	38	34	32	26
Total	6	**616**	**507**	**456**	**488**

The control sample size between Year 1 and Year 3 slightly decreased, primarily due to a reorganization of one control school. In the first and second years, the school had four classrooms averaging twenty-three students. During the third year of the study, the

school ended with a single second-grade classroom of only twenty-two students. The district's Research Director recommended another, similar school with the following student demographics:

Control School #4 (Original)	Control School #4 (Replacement)
Girls: 53%	Girls: 49%
F/R: 84%	F/R: 93%
ELL: 69%	ELL: 52%
% Minority: 81%	% Minority: 93%

Table 2 presents the third-grade student demographic information for the experimental and control groups at the start of each school's academic year. The school districts provided early data for 1,067 students: 562 in the experimental group and 505 students in the control group.

Table 2

Third-Grade Student Distribution by Student Level Variables

		Treatment	Control	Total
		(n = 562)	(n = 505)	(n = 1,067)
Gender				
	Female	271	253	524
	Male	291	252	543

Ethnicity				
	Asian	52	20	72
	Black	26	26	52
	Hispanic	285	299	584
	Native Am	14	12	26
	White	182	148	330
SES				
	F/R	259	287	546
Language Ability				
	ELL	117	195	312

Table 3 provides the same data but uses percentages to demonstrate the student demographics.

Table 3

Demographics of the Experimental and Control Groups (percentages)

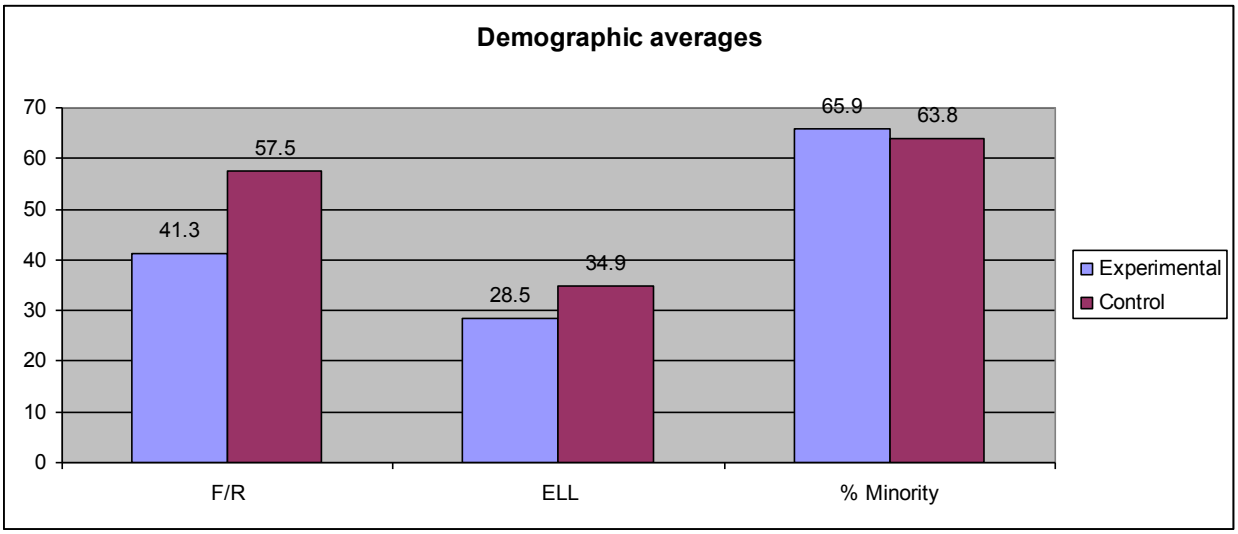
Group	% Girls	% ELL	% Hispanic	% Minorities	% F/RL
Experimental	48.2	20.8	50.7	67.6	46.0
Control	50.0	38.6	59.2	70.6	56.8

Compared to earlier years, there were some shifts in the student population demographics. The percentage of experimental students in the national free/reduced (F/R) lunch program (as a measure of low SES) fluctuated from a low of 39.3% to 46% in the

fourth year. The average low SES measure for the control group averaged 56% over the four years. The overall percentage of minority students in each group also increased over the last three years. Ethnic minority students are primarily Hispanic (55%). Chart 1 provides the average of demographic characteristics for the two groups over time.

Chart 1

Averaged Student Demographics



Over time, the number of students identified as English Language Learners (ELL) declined in both the experimental and the control groups. This reduction in the number of ELL students is the probable result of recent state laws affecting immigration. A law, effective in January 2008, gave the state authority to suspend or revoke the business licenses of employers who intentionally or knowingly employ workers who are unauthorized to work in the United States. For the most part, in Arizona, this means Mexican immigrants. A large number of immigrants left the state withdrewing many

students classified as ELL. In addition, the state's procedure for educating ELL students has undergone a substantial revision since the beginning of this study. As a result of federal court rulings (2000, 2007) that the state was not adequately educating or funding ELL students, the state passed Arizona Revised Statutes §15-756.01. The law requires that a minimum of four hours per day of English language development (ELD) be provided for the first year in which a pupil is classified as an English Language Learner. Both of these changes impacted the number of ELL students enrolled and their educational process within the targeted schools.

Longitudinal Student Analysis

The experimental students can be further divided into two groups: the first group included those treatment students who were in the kindergarten study and the second group included all treatment students in the grade level. Table 4 lists the experimental study schools, the total number of third-grade students and the kindergarten cohort. Bret Tarver experienced the highest rate of student attrition. The school had a high number of ELL students, and many were removed from school after the aforementioned state legislation. In addition, the District restructured and changed the school's boundaries in the last academic year. The average experimental retention rate in the final year of the study was nearly 60%.

Table 4

Third-Grade Students from Year 1, by School, at Spring 2010 Benchmark

Group	Name of School	Total Kindergarten Cohort (06-07)	Total 3rd Graders Spring, 2010	#3rd Graders from Initial Kindergarten Cohort	Retention Rate from Year 1
Experimental	Alhambra	89	85	55	62%
	Bret Tarver	123	151	40	33%
	CTA-Liberty	114	132	77	68%
	Gallego	95	104	65	68%
	Valley Academy	117	104	78	67%
Total	**5**	**538**	**576**	**315**	**59%**

The rate of the control sample isn't available because one district was unable to provide individual student tracking data. Based on a comparative school within the district, we can estimate a high retention rate of at least 60% for those schools. Comparative longitudinal analysis was adjusted for this loss of data by using a separate, matched data file for the remaining students and using a simple random sample of 25%.

Table 5

Retention Rate for Control Group

Group	Name of School	Total Kindergarten Cohort (06-07)	Total 3rd Graders Spring, 2010	#3rd Graders from Initial Kindergarten Cohort	Retention Rate from Year 1
Control	#1 – M	116	84	45	39%
	#2 – N	131	88	N/A*	N/A*
	#3 – O	102	79	N/A*	N/A*
	#4 – P	101	104	5	5%
	#5 – Q	128	107	51	40%
	#6 - R	38	26	1	3%
Total	**6**	**616**	**488**	**102**	**22%**

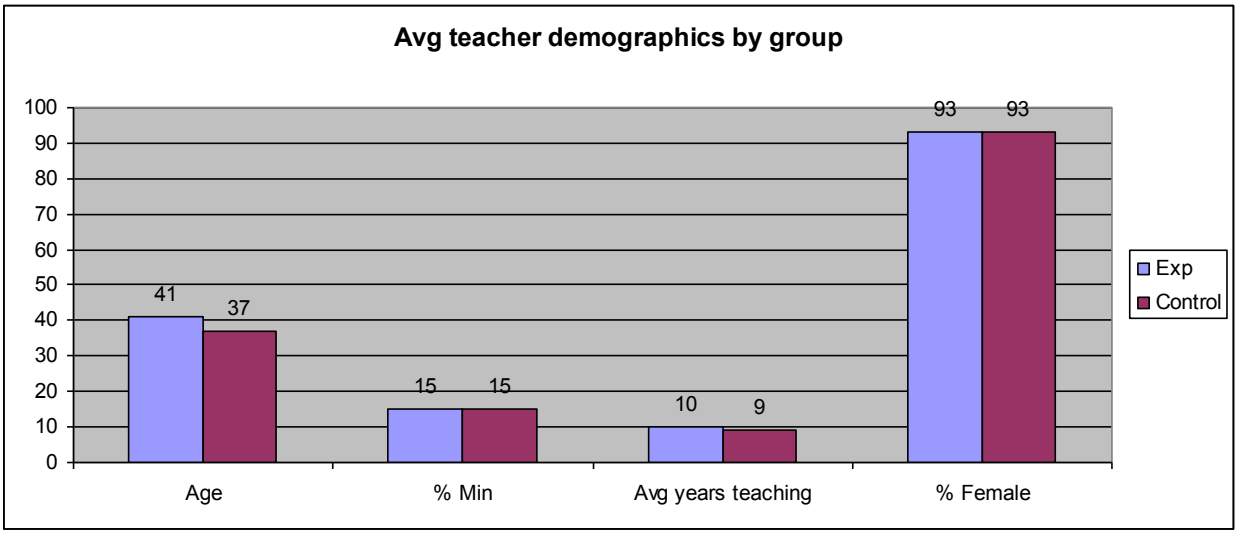
*N/A: Not available

Teacher Demographic Data

Demographic data, obtained from teacher survey questionnaires (see measures below) show that the average teacher age between the two groups is forty one for the experimental group and thirty seven for the control group. Chart 2 displays additional demographics for the participating teachers, averaged over the study. The experimental group is slightly older, but well matched on all variables.

Chart 2

Average Teacher Demographics



Research Design

Arizona State University researchers used a quasi-experimental design to assign schools as control or experimental. The independent variable was the type of reading instruction provided in the experimental and control classrooms. Classes in the experimental group (average of twenty-three classrooms per year) used *The Writing Road to Reading* curricula an average of ninety minutes each day, while control classes (average of twenty-four classrooms per year) used either Harcourt or Houghton Mifflin, their standard basal reading programs, for an average of eighty minutes each day. The dependent variable was student scores on DIBELS subtests that assess reading skills identified in the *Report of the National Reading Panel* (2000). To control for ecological

factors, researchers matched schools on socioeconomic status of students, class size, race/ethnicity/gender of students, and geography.

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 the DIBELS assessment because it has broad visibility, acceptance in the field, and it demonstrates high technical merit. In addition, the Arizona State Department of Education adopted DIBELS as the assessment for its Reading First program, and DIBELS was already used by schools in the study. DIBELS was designed to assess the five major components in early reading identified by the National Reading Panel: phonemic awareness, phonics/decoding, vocabulary, text comprehension, and fluency (Good & Kaminski, 2003). Administration periods for DIBELS occur at the beginning, middle, and end of each school year. The DIBELS subtests used were Initial Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Word Use Fluency, and Oral Reading Fluency and Retell Fluency were added for first grade and above (Kaminski & Good, 1996, 1998). Initial Sounds Fluency, Letter Naming Fluency, and Phoneme Segmentation Fluency are standardized, individually administered measures of phonological awareness. Nonsense Word Fluency, Oral Reading Fluency, and Retell Fluency are standardized, individually administered measures of the alphabetic principle: speed of decoding and remembering what was read.

Experimental Teacher Implementation of *The Writing Road to Reading*

Researchers used a uniform quantitative instrument to measure how experimental teachers were implementing *The Writing Road to Reading* in the classrooms. Researchers designed and used observation protocols to collect data during classroom observations. To measure program implementation, the third-grade scoring guide protocol (see Appendix A) measures constructs such as classroom management, adherence to program philosophy, and strategies for implementing spelling, writing, and reading content.

Procedure

Arizona State University researchers conducted this four-year study from 2006-2007 through 2009-2010. As an incentive for participation, experimental teachers received materials and training in *The Writing Road to Reading* program without charge. Control teachers each received $200 gift certificates to a bookstore for classroom materials.

To investigate whether teachers' prior experience and training in reading affected student performance, experimental and control teachers completed survey questionnaires at the beginning of each year. Copies of the control and experimental questionnaires are provided in Appendix B. These questionnaires revealed that control teachers had an average of eight hours on the selected reading program plus an additional average of four hours in training on a variety of reading methods such as Reading A-Z, Read Naturally,

EdMark reading, and Accelerated Reading. The majority of experimental teachers received forty-five hours of preservice training on *The Writing Road to Reading* including explicit instruction of the five research-based components of reading.

All students were tested within the first three weeks of the school year as required by the Arizona State Department of Education, again in December or January, and finally in May or early June of each year. Four researchers, in teams of two or as a whole group, visited the classrooms three times per year and observed individual teachers to ensure inter-observer agreement and reliability. Classroom observations lasted approximately forty-five minutes to one hour and focused on whole-group instruction.

Results

This section presents the results of the student assessment scores and an analysis of teacher characteristics as they relate to student achievement.

Student Performance Results

An alpha level of .05 was adopted. Table 6 displays the comparative performance of experimental and control students on the DIBELS measures administered in each benchmark assessment over the course of the study. In each administration over the four years, experimental students had consistently higher *mean* values on all DIBELS measures which supported the hypothesis.

Table 6

Comparative Four-Year Mean Scores of Experimental and Control Students on the DIBELS*

			Experimental		Control	
		Measures+	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
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	34.70	25.8
		WUF	35.84****	16.8	18.80	15.4

	Winter, 2008	PSF	48.20****	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

Table 6

Comparative Four-Year Mean Scores of Experimental Students on the DIBELS*

cont'd

			Experimental		Control	
		Measure+	Mean	SD	Mean	SD
Year 2 First Grade	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	44.50*	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,	Fall, 2009	ORF	96.99**	37.11	72.03	37.54
Third Grade	Winter, 2010	ORF	110.68**	38.23	88.14	41.14
	Spring, 2010	ORF	120.55**	36.28	105.65	39.18

Note: * p=<.05, ** p =<.01, ***p=<.005, ****p=<.001

[+] 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)

Kindergarten experimental students had consistently higher *mean* values on all DIBELS areas in spite of the fact that the experimental group included a higher percentage of ELL students. By the Spring testing, all scores were statistically significant (p = .001), except Initial Sound Frequency (no significance) and Letter Name Fluency (p = <.005).

First-grade experimental students again had consistently higher *mean* values on all DIBELS subtests. Two treatment schools and two control schools did not administer the Word Use Fluency measure in Fall of 2007. Oral Reading Fluency (ORF) and Retell Fluency (RF) were measured in Winter, 2008. During the spring testing, all schools used the same subtests as they did in the Winter. To examine threats to the study's internal

validity (e.g., the Hawthorne effect), an additional analysis was conducted comparing 2007-2008 experimental kindergarten student achievement with that of the 2006-2007 kindergarten cohort. During the 2007-2008 year, kindergarten students, teachers, and administrators were not participants in the study; however, data was collected on an identically matched sample of 2007-2008 kindergarten students. Although kindergarten teachers and classrooms were not the grade level under study, student performance slightly exceeded the kindergarten cohort of 2006-2007, indicating that the difference was not due to the Hawthorne effect. Table 7 on the next page provides mean scores on the final testing categories.

Table 7

Comparative Mean Scores of Selected Experimental Kindergarten Students on the Year-End DIBELS from Two-Year Study (Spring, 2007 = Year 1; Spring, 2008 = Year2)

		Year 1 ('07) Kindergarten	**Year 2 ('08) Kindergarten**	**Difference**
Year End Testing Results	Letter Naming	48.43	49.35	0.92
	Phoneme Segmentation	50.09	51.49	1.4
	Nonsense Word	45.92	50.41*	4.49
	Word Use	37.51	41.46*	3.95

*p<.05

Similar to Year 2, second-grade experimental students had consistently higher *mean* values on all DIBELS subtests than control school students using other reading methods. None of the experimental schools and only one control school administered the Phoneme Segmentation Fluency test in second grade. Therefore, PSF results were deleted from the Fall, 2008 analysis, and WUF results were deleted from the Winter and Spring analysis. All schools administered Nonsense Word Fluency and Oral Reading Fluency. By Year 4, data from all schools was available only on Oral Reading Fluency. Experimental students had a statistically significant advantage ($p = < .01$) which supported the hypotheses.

Table 8 provides the results of the longitudinal student analysis. In this set, the longitudinal experimental students have higher scores than the overall experimental students as well as the longitudinal control group. Students in each group improved in reading skills by the end of each year; however, at each benchmark experimental group students had higher mean scores than control group students by an average of eleven points. The longitudinal data represents a random sample of the students for whom all years of data was available (three control and three experimental schools).

Table 8

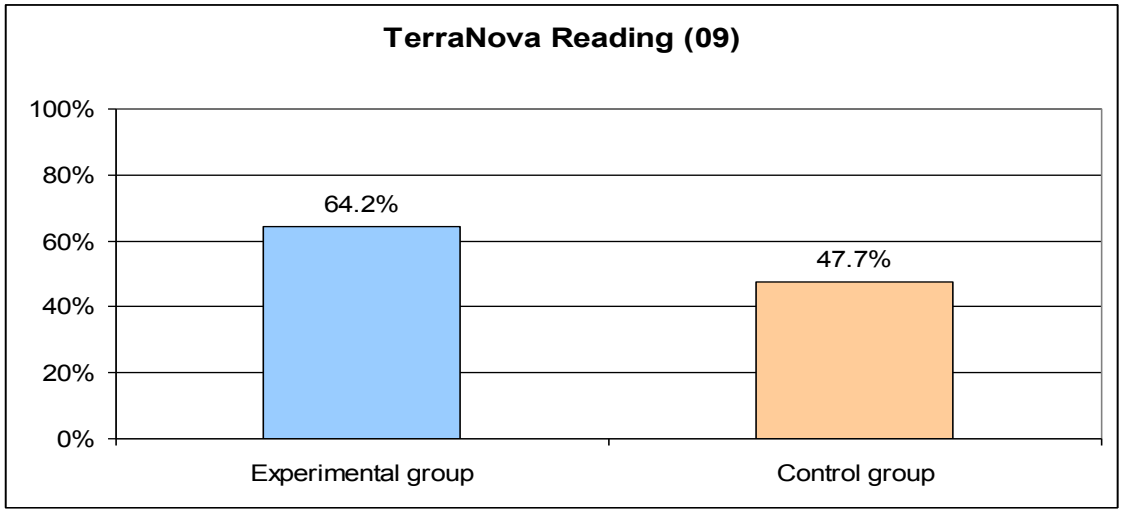
Comparative 2010 Mean Scores of Longitudinal Experimental and Control Students on the DIBELS

			Longitudinal Experimental		Longitudinal Control	
		Measure	Mean	SD	Mean	SD
	Spring, 2010	ORF	121.0	28.75	88.6	34.66

Another analysis of reading achievement was available in the third-year's study because all second-grade students in Arizona were required to complete the state's norm-referenced achievement test, the TerraNova. This assessment is designed to assess student achievement in reading, language arts, mathematics, science, social studies, vocabulary, spelling, and other areas. The chart on the next page represents a sample of the study students (three control and three experimental schools) and their average Normal Curve Equivalence (NCE) score on the TerraNova reading portion. As would be expected from reviewing the DIBELS scores, the experimental students' NCE scores were significantly higher than the control students on the state test ($p < .01$) supporting the hypotheses.

Chart 3

Student NCE Reading Scores from Spring, 09 AZ TerraNova Reading Portion



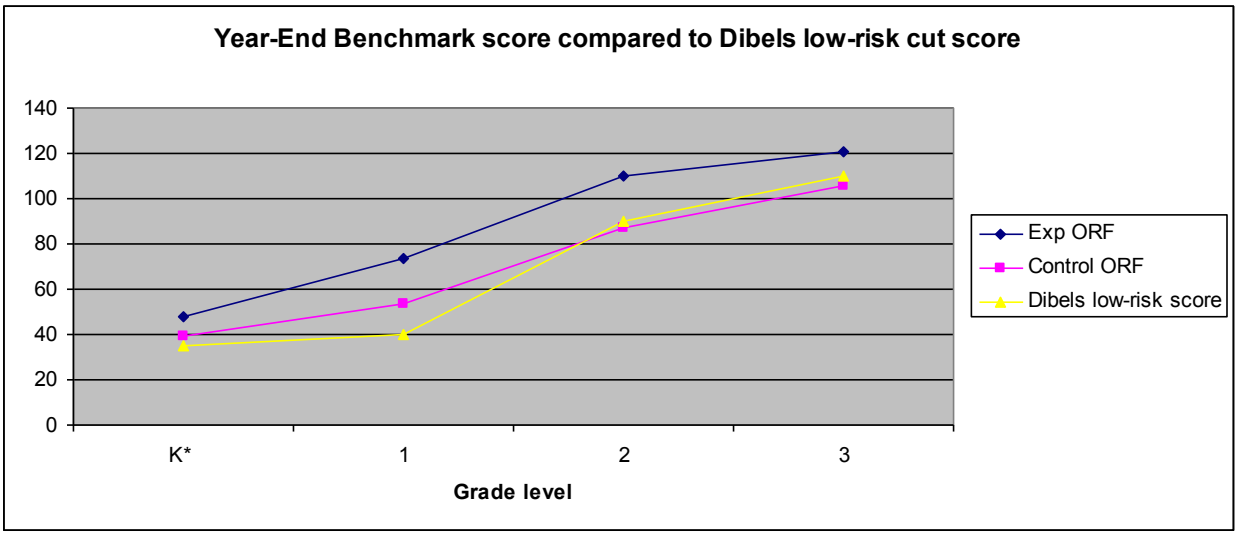
p = <.01

As shown in Chart 4 on the next page, 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). The average control student did not meet the DIBELS assessment for low-risk scores as of the middle of second-grade. Control students' achievement level average score was approximately five points below the DIBELS low risk threshold.

Chart 4

Overall Year-End Assessment Scores Compared

to DIBELS Decision Rules Benchmarks



*Note: *K, Spring 2007 is Phoneme Segmentation Fluency, all others are Oral Reading Fluency.*

In addition to measures of statistical significance, researchers used Cohen's d to measure the effect size. Cohen's d using a pooled standard deviation was computed for DIBELS ORF 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 was .4. This means that the intervention had a positive, medium effect (average of 0.54) on student achievement. Using scores from the longitudinal group, the effect size for Fall 2009 was .9; for Winter 2010 it was .8; and in Spring 2010 the effect size was .6. This means that the intervention had a large effect size (average of 0.78) on student reading achievement.

Implementation by Experimental Teachers

Classroom observations were the primary measure for program implementation. The goal for observations was to see accurate and consistent instruction across grade levels and schools. Using the observation protocals (see Appendix A), observers noticed an increase in accuracy and consistency within and across the five schools by the fourth-quarter observations each year. The year-end researchers' overall observation protocol results showed that at least 76% of program practices were satisfactorily implemented by experimental teachers with 15% of experimental teachers' behaviors needing further refinement. The final year observation ratings were consistent with prior years; therefore, researchers concluded that the program was satisfactorily implemented.

Discussion

The results of this study supported and expanded findings of previous investigations of explicit, multisensory, structured instruction as taught in *The Writing Road to Reading* (Aaron, Joshi, Gooden, & Bentum, 2008; Aukerman, 1984; Farnham-Diggory, 1992; Hoerl & Koons, 1995; North, 1995, 2001). The outcomes of this study can be summarized as follows:

1. Children who received explicit, multisensory, structured language arts instruction demonstrated significant learning gains in reading skills.

2. Reading skill attainment of children who received explicit, multisensory, structured language arts instruction exceeded that of children instructed in other, more conventional reading programs.
3. The majority of the experimental teachers satisfactorily implemented *The Writing Road to Reading* program in their classrooms.

According to four years of results, students who received explicit, multisensory, structured language arts instruction demonstrated higher and statistically significant scores as measured by DIBELS compared to students in the control classrooms using more conventional, basal reading programs. 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. One possible reason is that control teachers may have received mimimal training in the five research-based components identified by the NRP report, but they may not have received training in how to implement them in the classroom. Another possible reason is the way in which the components are taught. Integration of the components with explicit, multisensory, structured instruction in the experimental program differed from instruction in the control classrooms. Experimental school instruction began by connecting speech to print rather than print to speech. Experimental kindergartners were taught all common sounds (e.g., long and short vowel sounds) represented by a single letter or letter combination from the beginning of kindergarten, enabling them to sound out high-frequency words such as *me*

and *do.* The predominate multisensory strategy observed was teaching the sound (e.g., /*b*/) or sounds of the letters and letter combinations (e.g., *ay* or *ai*) with manuscript handwriting from the beginning of instruction. The experimental school advantage became apparent in the kindergarten Fall, 2006 DIBELS results in which experimental children had statistically significant differences (p = .001) on Initial Sound Fluency and Letter Naming Fluency. By the Winter testing, there was no significant difference of Initial Sound Fluency, possibly because by midyear the control school programs had taught initial sounds of the alphabet letters. Additional multisensory strategies observed were pronouncing the sounds of letters in high-frequency words and writing them in spelling notebooks. In addition, a marking system helped connect speech to print. Students underlined multiletter combinations representing one sound, and placed a number above letter or letter combinations representing a sound other than the first one. They reread the words in the notebooks sequentially by sound and in normal speech. This multisensory activity may explain the positive difference demonstrated by kindergarten and first grade students on reading nonsense words (p=<.001). Students at all four grade levels continued multisensory activities in the writing lesson where the same high-frequency words were used to teach vocabulary, grammar, and syllable division. The final distinctive strategy observed in the experimental classrooms was the explicit, multisensory teaching of five mental actions (comprehension strategies). A 2010 What Works Clearinghouse (WWC) report, *Improving Reading Comprehension in Kindergarten through Third Grade*, provides a set of evidence-based practices that

improve reading comprehension of young readers. The first recommendation was to "teach students how to use reading comprehension strategies" (Shanahan, Callison, Carriere, Duke, Pearson, Schatschneider, & Torgeson, 2010, p. 1) They defined comprehension strategy as "intentional mental actions during reading that improve reading comprehension" (p. 11).

Limitation of the Present Study

One of the limitations of the study was that the individual control and experimental school administrations determined which DIBELS subtests were administered during each testing period in each year. If test data was incomplete when submitted, researchers reported data on matched control and experimental schools. Researchers could not impose additional tests beyond the state and district required assessments (e.g., DIBELS in kindergarten through grade three and Terra Nova in second grade). Therefore, future studies should include more discriminating assessments. Another confounding factor was the change in Arizona law and the federal court rulings governing immigration and instruction for ELL students. Student populations most affected by these changes were the two experimental and two control schools that had minority populations exceeding 80% at the beginning of the four years. Also, school and district administrations did not identify mainstreamed special education students in the control or experimental classrooms. Future investigations should compare reading gains

of mainstreamed special education students in experimental and control classrooms to determine the effectiveness of the treatment with each group.

Conclusions

This four-year study compared reading achievement of general education students trained in an explicit, multisensory, structured progam with general education students trained in conventional reading programs. Both control and experimental students received instruction in the five NRP research-based components. The experimental students, however, received explicit, multisensory, structured instruction. Experimental students demonstrated statistically significant gains at every grading period. The medium to large effect sizes demonstrated in this study suggest that the combination of concerted, simultaneous integration of research-based components and explicit, multisensory, structured instruction contributed to the experimental groups' significantly greater gains in reading. Studies by Calvert et al., Shams and Seitz, and Shimojo and Shams point out that multisensory interactions occur in many different perceptual tasks and settings. Herron (2008), Dehaene (2009), and others, emphasize that children's first experiences with letters and words dictate how neural networks are established and habitual pathways are formed in the brain as reading skills develop. With forty percent of fourth graders reading at the basic level, it is important that additional research be focused on the integration of explicit, multisensory, structured instruction with research-based components of skilled reading.

References

Aaron, P., Joshi, R., Gooden, R., & Bentum, K. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of LD. In *Journal of Learning Disabilities*, *41*, 67-84.

Aukerman, R. (1984). *Approaches to beginning reading* (2nd ed). New York: Wiley.

Calvert, G., Spence, C., & Stein, B. (2004). *The handbook of multisensory processes*. Boston, MA: The Massachusetts Institute of Technology.

Campbell, M., Helf, S., & Cooke, N. (2008). Effects of adding multisensory components to a supplemental reading program on the decoding skills of treatment resisters. *Education and Treatment of Children*, *31*:3, 267-295. Morgantown, WV: West Virginia University.

Cohen's D. Retrieved online, September 23, 2009. University of Colorado at Colorado Springs website, link: http://www.uccs.edu/-faculty/lbecker/#means and standard deviations.

Cox, A. (1985). Alphabetic Phonics: An organization and expansion of the Orton-Gillingham method. *Annals of Dyslexia, 35*, 187-198.

Dehaene, S. (2009). *Reading in the brain*. New York: Viking Penguin.

Dev, P., Doyle, B., & Valente, B. (2002). Labels needn't stick: "At-risk" first graders rescued with appropriate intervention. *Journal of Education for Students Placed At Risk,* 7(3), 327-332.

Dickson, S., & Bursuck, W. (1999). Implementing a model for preventing reading failure; A report from the field. *Learning Disabilities Research & Practice, 14,* 191-202.

Farnham-Diggory, S. (1990). *Schooling,* 124-131. Cambridge, MA: Harvard University Press.

_____, S. (1992). *Cognitive processes in education*. New York: HarperCollins.

Good, R. H., & Kaminski, R. A. (2003). Dynamic indicators of basic early literacy skills (6th Ed.). Longmont, CO: Sopris West.

Greene, J. F. (1996). *LANGUAGE!:* The effects of an individualized structured language curriculum for middle and high school students. *Annals of Dyslexia, 46,* 97-121.

Herron, J. (2008). Why phonics teaching must change. Association for Supervision and Curriculum Development, 77-81. Alexandria, VA: Association for Supervision and Curriculum Development.

Hoerl, M., & Koons, D. (1995). Effect of Spalding multisensory phonics instruction on the literacy skills of high school special education students. In C.W. McIntyre and J.P. Pickering, eds., *Clinical studies of multisensory structured language education*. Salem, OR: International Multisensory Structured Language Education Council.

International Dyslexia Association. (2000). *Multisensory Teaching*. Baltimore, MD: The International Dyslexia Association.

Joshi, R., Dahlgren, M., & Boulware-Gooden, R. (2002). Teaching reading in an inner city school through a multisensory teaching approach. *Annals of Dyslexia, 52,* 229-242.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*, 215-227.

Kaminski, R., & Good, R. (1998). Assessing early literacy skills in a problem-solving model: Dynamic Indicators of Basic Early Literacy Skills. In M.R. Shinn (Ed.) *Advanced applications of curriculum-based measurement* (pp. 113-142). New York, NY: Guilford Press.

Learning Point Associates. (2004). A closer look at the five essential components of effective reading instruction: A review of scientifically based reading research for teachers. Naperville, IL: Learning Point Associates.

McIntyre, C. W., & Pickering, J. P. eds. (1995). *Clinical studies of multisensory structured language education*. Dallas, TX: International Multisensory Structured Language Education Council.

Moats, L. C., & Farrell, M. L. (1999). Multisensory instruction. In J. R. Birsh (Ed.), *Multisensory teaching of basic language skills* (pp. 1–17). Baltimore, MD: Paul H. Brookes Publishing Co.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

North, M. E. (1992). The writing road to reading: From theory to practice. *Annals of Dyslexia 42*: 110-123.

_____ M. E. (1995). The effects of Spalding instruction on special education students. In C. McIntyre and J. Pickering, eds. *Clinical studies of multisensory structured language education*. Salem, OR: International Multisensory Structured Language Education Council.

Oakland, T., Black, J., Stanford, G., Nussbaum, N., & Balise, R. (1998). An evaluation of the dyslexia training program: A multisensory method for promoting reading in students with reading disabilities. *Journal of Learning Disabilities, 31*, 140-147.

Ogden, S., Hindman, S., & Turner, S. (1989). Multisensory programs in the public schools: A brighter future for LD children. *Annals of Dyslexia, 39*, 247-267.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D. & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31-74.

Ritchey, K. D., & Goeke, J. L. (2006). Orton-Gillingham-Based reading instruction: A review of literature. *Journal of Special Education. 40*(3), 171-183.

Scheffel, D., Shaw, J., & Shaw, R. (2008). The efficacy of a supplementary multisensory reading program for first-grade students. *Journal of Reading Improvement, 45*, 3.

Shams, L., & Seitz, A. (2008). Benefits of multisensory learning. In *Trends in Cognitive Science*, xxx:x, 1-7 retrieved from web.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgeson, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4-38). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from whatworks.ed.gov/publications/practiceguides.

Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. New York: Knopf.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology, 11*, 505-509.

Spalding, Romalda B., (2003). *The Writing Road to Reading*, North, Mary E. Ed. Harper Collins.

Vickery, K., Reynolds, V., & Cochran, S. (1987). Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. *Annals of Dyslexia, 37*, 189-200.

Appendix A

Final Third-Grade Teacher

Observation Checklist

SPALDING COMPONENT	Observation			
	S	*NR*	*NO*	*Comments*
Philosophy				
Does the teacher …				
1. Make children's physical and mental well-being a primary concern?				
• Ensure that children's feet rest comfortably on the floor.				
• Ensure that children's arms rest comfortably on the desk.				
• Have all children face the front for whole group instruction.				
• Maintain a well-organized classroom.				
• Praise children for good performance.				
• Demonstrate a positive feeling tone.				
2. Have high expectations for children of differing ability levels?				
3. Consistently encourage higher-level thinking by all children?				
• "Why is another /p/ added to the word *drop* before adding *ed*? (spelling example)				
• "By adding the suffix *ent* to *excel*, how is the part of speech changed?" (writing example)				
• "How does reformatting help you to discern an implied main idea?" (reading example)				
4. Provide direct, sequential instruction?				
• Use the Collins Model to model, check understanding, coach, scaffold/fade.				
• Follow the sequence in the Third-Grade Teachers' Guide.				
5. Consistently provide multisensory instruction?				
• Have children see, say, write, and read phonograms, words, etc. (spelling example)				
• Have children compose oral/written sentences and read their own. (writing example)				
6. Encourage active participation by all children in each lesson?				
• Have children say phonogram sounds in unison? (spelling example)				
• Have children read literature, science, social studies books aloud? (reading example)				
7. Demonstrate diagnostic teaching?				
• During OPR, stop to correct mispronunciations. (spelling example)				
• Identify incorrect word usage and reteach. (writing example)				
• Identify incorrect mental action responses and reteach. (reading example)				
8. Demonstrate integrated language arts instruction?				
• Use spelling words to teach pronunciation, rules, grammar, and reading.				
• Demonstrate the connection between spelling, writing, and reading lesson objectives. "We learned how to spell these words; let's learn their meanings."				

S=Satisfactory; NR=Needs Refinement; NO=Not Observed

Final Third-Grade Teacher

Observation Checklist

SPALDING COMPONENT	Observation			
	S	*NR*	*NO*	*Comments*
Spelling: *Phonemic Awareness, Phonograms with Handwriting, Dictation*				
Does the teacher …				
1. Provide daily *Oral Phonogram Reviews*?				
Step 1 (Delivering 10-11, 13, WRTR 39):				
• Use maximum 30 cards at one time.				
• Pass cards back to front, showing only one card at a time.				
• Listen for precise pronunciation of sound(s) *only*, e.g., /dge/, not cue words.				
• Cover card if mispronounced, then pronounce precisely, e.g., /j/ (not /juh/).				
• Have children repeat sound(s) precisely.				
Step 2 (Delivering 10-11, 13, WRTR 39):				
• Use maximum 30 cards at one time.				
• Show phonograms a second time.				
• Listen for precise pronunciation of *only* sounds in unison.				
• Ask questions that clarify which pronunciation or phonogram to use?				
• Have children respond and explain.				
2. Provide daily *Written Phonogram Reviews*? *(Delivering 14-17, WRTR 41-42)*				
• Coach children's accurate and legible handwriting. Set handwriting focus for each WPR, e.g., "Today focus on beginning clock letters at 2 on the clock."				
• Use maximum 20 cards at one time.				
• Pronounce phonograms precisely.				
• Require children to say sounds in unison. (no teacher voice overs)				
• Give cues as needed.				
• Require children to say sounds softly as they write. (no teacher voice overs)				
• Give delayed feedback, e.g., write each phonogram from children's dictation.				
• Have children evaluate their handwriting based on the day's focus.				

S=Satisfactory; NR=Needs Refinement; NO=Not Observed

SPALDING COMPONENT	Observation			
	S	*NR*	*NO*	*Comments*
3. Follow daily spelling dictation procedure? *(Delivering 25, WRTR 49)*				
• Say word in normal speech.				
• Give sentence when needed for meaning.				
• Use fingers for *all* sounds (two hands for multisyllable words).				
• Have children say sounds in unison, then *sound* and write the word.				
• Have children dictate word to teacher. (no teacher voice overs)				
• Write word on board.				
• Have children dictate markings, then read word. (no teacher voice overs)				
• Have children give rule when appropriate. (no teacher voice overs)				
• Have children *read for spelling*. (no teacher voice overs)				
• Have children *read for reading*. (no teacher voice overs)				
Writing: *Sentence Construction*				
Does the teacher …				
1. Model composing oral/written sentences that demonstrate usage and meaning of week's *unfamiliar* spelling/vocabulary words? (Writing Objective #1)				
2. Coach as children compose oral/written sentences that demonstrate usage and meaning of week's spelling/vocabulary words?				
3. Model one additional writing objective in the *Teachers' Guide?*				
4. Coach as children perform the task?				
Reading: *Literary Appreciation, Mental Actions*				
Does the teacher …				
1. Coach as children use all five mental actions in a McCall-Crabbs passage with underlining and/or taking notes of essential words (reformat) to discern an implied main idea (mentally summarize).				
2. Coach as children read literature/science/social studies books aloud to demonstrate fluent and expressive reading.				

S=Satisfactory; NR=Needs Refinement; NO=Not Observed

Appendix B

Teacher Questionnaire for Experimental Schools

Name:

School:

School Schedule:

Fall Semester From: to:

Spring Semester From: to:

Summer School From: to:

Please list the K-3 standardized tests currently used at your school that are relevant to reading:

Test Title	Approx. Date of Administration

Total number of years that you have been teaching: ____

Number of years teaching at this school: ____ Grade you currently teach: ____

Number of years teaching:

Kindergarten: ____ 1st grade: ____ 2nd grade: ____ 3rd grade: ____ 4th grade: ____

Degrees Earned:	Year Received:

AZ Certification:	Year Received:	Endorsement(s):	Year Received:

Your age group (please check appropriate category):

___Under 30 ___30 -39 ___40-49 ___50-59 ___60 or above

Gender (please check): ___F ___M

Ethnicity (please check): ___African American ___Asian ___Caucasian

___Hispanic ___Native American ___Other

Do you have access to the Internet in your classroom? ___Yes ___No

Please specify the basic instructional reading program/method that you personally use on a regular basis:

Program Title	# Hours/Day

Please list all commercial readers and supplemental reading programs and/or materials that you personally use in your classroom:

Program/Material Title	# Hours/Week

Number of years using the Spalding Method: ____
Spalding Courses Completed: Year(s) When Trained:
Spalding 1 (45 hours, between 1996 and 2003):
Spalding 2 (45 hours, between 1998 and 2004):
Integrated Language Arts 1 (45 hours, since 2003):
Integrated Language Arts 2 (45 hours, since 2004):
"Spalding Certified" (please check appropriate answer): Instructor ___Yes ___No Teacher ___Yes ___No
List other reading (or reading-relevant) programs for which you have received training: Program Title Target Grade Year(s) When Trained # Hours of Training
List all reading (or reading-relevant) programs that you have used so far (with or without training): Program Title Target Grade Used From Year… To Year…
List other professional development courses you have completed (e.g. Essential Elements of Instruction): Program Title Target Grade Year(s) When Trained

Thank you for responding to this questionnaire.

Teacher Questionnaire for Control Schools

Name:

School:

School Schedule:

Fall Semester From: to:

Spring Semester From: to:

Summer School From: to:

Please list the K-3 standardized tests currently used at your school that are relevant to reading:

Test Title	Approx. Date of Administration

Total number of years that you have been teaching: ____

Number of years teaching at this school: ____ Grade you currently teach: ____

Number of years teaching:

Kindergarten: ____ 1st grade: ____ 2nd grade: ____ 3rd grade: ____ 4th grade: ____

Degrees Earned:	Year Received:

AZ Certification:	Year Received:	Endorsement(s):	Year Received:

Your age group (please check appropriate category):

___Under 30 ___30 -39 ___40-49 ___50-59 ___60 or above

Gender (please check): ___F ___M

Ethnicity (please check): ___African American ___Asian ___Caucasian

___Hispanic ___Native American ___Other

Do you have access to the Internet in your classroom? ___Yes ___No

Please specify the basic instructional reading program/method that you personally use on a regular basis:

Program Title	# Hours/Day

Please list all commercial readers and supplemental reading programs and/or materials that you personally use in your classroom:

Program/Material Title	# Hours/Week

List reading (or reading-relevant) programs for which you have received training:			
Program Title	Target Grade	Year(s) When Trained	# Hours of Training
List all reading (or reading-relevant) programs that you have used so far (with or without training):			
Program Title	Target Grade	Used From Year…	To Year…
List other professional development courses you have completed (e.g. Essential Elements of Instruction):			
Program Title	Target Grade	Year(s) When Trained	

Thank you for responding to this questionnaire.