

# Findings from a Study of Early Reading Interventions with Randomized Assignment to Groups

---

Submitted March 6, 2015

by

Catherine Callow-Heusser, Ph.D.

[cheusser@endvision.net](mailto:cheusser@endvision.net)

Wendy Sanborn, Ph.D.

[wendysanborn@gmail.com](mailto:wendysanborn@gmail.com)



EndVision Research & Evaluation, LLC

41 East 400 North #321

Logan, UT 84321

PH 435.757-2724

FX 435.753-3231

## Table of Contents

<b>I. Study Characteristics .....</b>	<b>1</b>
A. <i>Description of Intervention.....</i>	<i>2</i>
B. <i>Description of Comparison.....</i>	<i>5</i>
C. <i>Setting .....</i>	<i>7</i>
D. <i>Participants .....</i>	<i>7</i>
<b>II. Study Design and Analysis .....</b>	<b>8</b>
A. <i>Sample Formation .....</i>	<i>8</i>
B. <i>Outcome Measures.....</i>	<i>8</i>
C. <i>Analytical Approach .....</i>	<i>9</i>
D. <i>Statistical Adjustments.....</i>	<i>9</i>
E. <i>Missing Data.....</i>	<i>9</i>
<b>III. Study Data.....</b>	<b>10</b>
A. <i>Pre-Intervention Data, Baseline Sample (Table 1).....</i>	<i>10</i>
B. <i>Pre-Intervention Data, Analysis Sample (Table 2).....</i>	<i>10</i>
C. <i>Post-Intervention Data and Findings (Table 3).....</i>	<i>11</i>
<b>IV. Conclusions .....</b>	<b>12</b>
<b>VI. References .....</b>	<b>14</b>

# Findings from a Study of Early Reading Interventions with Randomized Assignment to Groups

---

This report summarizes findings from a study of beginning reading interventions used in early grades: (a) Reading for All Learners (RFAL) distributed by Academic Success for All Learners, the treatment intervention, and (b) Early Reading Intervention (ERI) distributed by Scott Foresman, the comparison intervention. Students in first grade at a Title I school in a Rocky Mountain state who read below grade level at the beginning of school year as measured by the Dynamic Indicators of Early Literacy Skills (DIBELS) were randomly assigned to the RFAL or the ERI groups for Tier 2 reading intervention time (i.e., time outside of core whole group reading time that is set aside for reading intervention for struggling students) during which they received small group instruction in reading.

## I. Study Characteristics

At the beginning of the school year in August, school staff agreed to implement the study through the end of January, at which time they planned to review middle of year DIBELS benchmark data and (a) continue the study as initiated, (b) change all students reading below grade level to the RFAL curricula, or (c) discontinue the RFAL curricula and return to exclusive use of ERI for supplemental Tier 2 instruction. All students in first grade participated in reading instruction with the core language arts program during the regularly scheduled English/language arts time. Students reading below grade level received an additional 45 minutes of reading instruction during Tier 2 intervention time. Students were randomly assigned as successively described to RFAL treatment and ERI comparison groups after beginning of year DIBELS testing. Regrouping of students occurred three to six weeks based on DIBELS progress monitoring and curriculum based assessment scores, though assignment to reading intervention program, RFAL or ERI, was maintained throughout the first half of the school year. Data were collected through the middle of year DIBELS benchmark assessments conducted mid-January. After analyzing DIBELS scores, the school decided to use RFAL with students in almost all reading intervention groups, including students who started the year on grade level but were no longer on grade level at middle of year. Students were assessed at the end of year using the DIBELS benchmark assessments as well as reading subtests of the Woodcock Johnson Tests of Achievement, version 3 (WJ III).

To ameliorate teacher effects, teachers and paraprofessionals taught both RFAL treatment and ERI comparison groups, with an approximately equal distribution of students across treatment and comparison groups for each instructor. Students in both treatment and comparison groups met for Tier 2 intervention time five days per week for approximately 45 minutes per day.

Classrooms were observed and data collected on key indicators of classroom functioning and effective instruction (i.e., lesson activities, percent of students engaged/observably off-task, pacing, student-teacher interactions, group/choral responses, praise/academic feedback, interruptions to instruction, etc.). Observation data indicated that treatment and comparison groups were similar in terms of student behavior, student-teacher interactions, rates of praise and academic feedback, instructional pacing, etc. Additionally, lessons were implemented per

reading curricula instructions with teachers appropriately following lesson plans, activities, teaching strategies, and rates of teacher/student interaction.

## **A. Description of Intervention**

### ***Reading for All Learners (RFAL)***

In 1990, Dr. Alan Hofmeister at the Center for Information Technology (CIT) at Utah State University initiated a long-term, programmatic research and development effort in beginning reading instruction. After surveying a wide range of different approaches to reading instruction, the CIT selected the Beginning Reading Program developed by the Southwest Regional Laboratory (SWRL) for Educational Research and Development and initially released in 1972. The original “Little Books” Beginning Reading Program combined the needs and interests of children with research on structured, systematic phonics-based approaches to reading instruction. The SWRL Little Books included 60 inexpensive softback black and white line drawing books that combined both engaging stories and systematic instruction and that were field-tested and revised over the course of a decade to ensure students made adequate progress.

Given theoretical and research base upon which the SWRL books were developed, the available longitudinal research with at-risk learners, and the comparative cost-effectiveness data, CIT researchers decided to build upon SWRL products and findings rather than try to develop a beginning reading program from scratch based on a similar approach. During the next 17 years, the CIT systematically and progressively revised the SWRL reading program based on field-testing data to incorporate research-based practices well-aligned with recommendations in the National Reading Panel’s (2000) summary of evidence-based practices.

The revised reading program was extensively adapted to further incorporate evidence-based strategies and field-tested to ensure reading success. Embedded teaching strategies included systematic and explicit phonics instruction using controlled introduction of sounds and words in connected text, cumulative review of sounds and words, comprehension questions asking students to infer or predict, repeated readings, explicit teaching of morphological rules for silent “e” words and prefixes and suffixes, regular reminders to praise the child, teacher/parent training materials, and other teaching/learning activities. Additionally, original books that introduced too many new sounds or words were split into multiple stories to further control sequencing and introduction of sounds and words. New stories were written to ensure consistent timing for new sound and word introduction and to add sufficient cumulative review, and curriculum-based assessments were developed. The revised program was released as Reading for All Learners with 141 different books containing more than 300 lessons in eight sets. The sets take students from beginning kindergarten reading levels into third grade.

For each lesson, teachers only need to locate the next book (or books if more than one lesson is covered) in the set, with enough copies for each student in the group. The lessons are self-contained within the book, including all sound and word practice, stories, progress monitoring/formative assessments, comprehension questions, instructions for implementation, etc., contained in each book.

The revised RFAL, copyrighted by Dr. Hofmeister, is distributed by Academic Success for All Learners (Utah entity #5642708-0142, established May 2004; DUNS 159364368) to schools and

districts in over 40 states, with extensive use in Native American schools throughout the country, inner-city boroughs of New York City, schools in Puerto Rico, Guyana, South America, Kenya, Uganda, and English reading instruction programs in other countries throughout Africa, Europe, Asia and Russia. Additionally, as the result of a National Science Foundation (NSF) Small Business Innovation Research (SBIR) grant (#1330901), RFAL apps are available for mobile devices including iPads, iPhones, Android tablets, Android phones, and Kindle Fire tablets, along with a web-based Student Assessment and Monitoring (SAM) data and reporting system.

There are several major differences between RFAL and other popular reading curricula. One substantive difference includes the sequencing of sounds and words. RFAL incorporates high frequency sounds and words in the initial instruction to capitalize on student familiarity with commonly used words. As a result, this curriculum differs from others because it introduces letters with long vowel sounds (e.g., I, me, see) before teaching all short vowel sounds. Additionally, letters with dissimilar shapes and sounds are initially taught (i.e., s, a, and m), with letters that look or sound similar taught separated in time. Furthermore, RFAL varies sentence length, patterns, and rhythms so students must pay attention in order to read fluently and accurately. Teachers, in particular, note this difference from other reading curricula they have implemented. Finally, in the first 3 sets (of 8), each story is self-contained in one book to help children feel successful in “closing the book” and completing a story. Particularly for children reading below grade level, completing a book with mastery of sounds and words, as well as comprehension of the story, provides momentum for continued reading and engagement and appears to increase students’ feelings of success.

For this study, teachers implemented RFAL lessons during Tier 2 intervention time five days per week for approximately 45 minutes per day (which included transit time to and from classrooms). Students were assigned to homogenous groups based on Dynamic Indicators of Beginning Early Literacy Skills (DIBELS) beginning of year benchmark scores and curriculum based placement assessments, which also helped determine the set and book used as a starting place for reading intervention.

After students had been randomly assigned to treatment or comparison groups, teachers, paraprofessionals, and literacy coaches participated in a four-hour afternoon training conducted by Academic Success for All Learners staff. Approximately once per month through January, a RFAL trainer visited the school to answer questions specific to RFAL, observe Tier 2 reading instruction, and offer suggestions to improve quality of instruction. Most of these suggestions applied equally to both RFAL and ERI reading groups, as they involved pacing, error correction, academic feedback, frequency of group choral versus individual responses, and other effective teaching practices.

### *Prior Evidence of Effectiveness*

Some evidence supports the effectiveness of RFAL and its precursor, the SWRL little books, as shown in Table 1. For example, Hanson and Farrell (1995) reported on the long-term effects of learning to read using the SWRL books, based on data from 3,959 high school seniors in the U.S. national follow-up study. The resoundingly positive results for the group who used the SWRL books beginning in kindergarten include the following: both disadvantaged and advantaged SWRL students outperformed advantaged students in the comparison group; students with ethnic

minority backgrounds outperformed students in the comparison group; and there were approximately one-third fewer functionally illiterate high school seniors who received early grades instruction with SWRL books when compared to students who did not, despite a higher percentage of disadvantaged students in the SWRL group. Differences between representative groups who used the SWRL books beginning in kindergarten and those who did not were statistically significant at the end of high school, including differences in reading comprehension, vocabulary, reading stage, and illiteracy rates.

In 2002, Cibecue Community School, located on the White Mountain Apache Reservation in Arizona, was one of the lowest performing Bureau of Indian Education (BIE) schools, with most students performing below the 10<sup>th</sup> percentile on standardized tests of achievement. During the 2002-2003 school year, Cibecue participated in a pilot program in anticipation of the BIE's application for a Reading First grant (Skenandore, Hofmeister & Willis, 2003). Cibecue implemented RFAL with all students in kindergarten through third grade as well as older students who were reading below grade level. By the end of the 2002-2003 school year, students at Cibecue performed, on average, at the 35<sup>th</sup> percentile in reading on the Stanford Achievement Test, version 9 (SAT-9). Additionally, the percentage of students proficient in math changed from 27% in 2001-2002 to 43% by the end of the 2002-2003 school year, with changes in math scores attributed to reading skills and confidence.

The READ Alliance based in New York City (<http://www.readalliance.org/>) provides additional data demonstrating the effectiveness of RFAL. The READ Alliance is a non-profit organization using a peer-tutoring model with RFAL focusing primarily with first-grade learners identified as struggling readers. RFAL is the only reading curriculum used by the READ Alliance during an afterschool program implemented at schools in the five boroughs of New York City. During the 2014-2015 school year, READ served students identified as struggling readers at 41 schools. Participating students made an average gain of one full grade level after completing 40, 45-minute after school peer tutoring sessions using RFAL.

Cache County School District ranks at the top in the state of Utah in reading performance, second only to the affluent Park City School District. Over 91% of students were proficient on the state assessment in 2013-2014, with 82% of Spanish-speaking students proficient in reading. In contrast, the state average was 78% for all students and 60% for Spanish-speaking students. Since the millennium, Cache county School District has exclusively used RFAL as their reading intervention program (e.g., Tier 2 in a 3-Tier model of instruction, Vaughn, Wanzek, & Fletcher, 2007) with struggling elementary age readers, and with all Spanish-speaking children and English learners regardless of reading level on beginning of year DIBELS benchmark scores (S. Krebs, Cache School District Literacy Specialist, personal communication). Their strong district-wide implementation of a 3-tier model has resulted in very few students with intensive needs in reading in elementary and middle school grades, and the highest percentage of students reading on grade level in the state. Literacy specialists in the district credit RFAL as a key piece of the district's reading instruction.

Finally, Lignugaris/Kraft, et al., (2001) reported that when parents used RFAL at home, approximately twice as many students in the treatment group scored in the top quartile on a standardized assessment of reading as students in the comparison group. Additionally, the

average score of children in the home reading treatment group was higher than 87% of scores for children in the comparison group.

Prior to the recent study described in this report, no other study using RFAL with random assignment to groups had been conducted.

Table 1. Prior Reading For All Learners Evidence of Effectiveness

Study	# Students	Study Duration	Outcomes
Hansen & Farrell (1995); National study	3,959	12 years	Statistically significant positive differences found in comprehension, vocabulary, drop out rates, illiteracy, and attitudes towards reading.
Skenandore, Hofmeister & Willis (2003); Cibecue Community School	230	1 year	On average, students' scores on the SAT-9 reading subtests changed from below the 10 <sup>th</sup> percentile to the 35 <sup>th</sup> percentile in 1 year.
Utah School Ranking, Utah State Office of Education Website ( <a href="http://schools.utah.gov">http://schools.utah.gov</a> )	Utah elementary schools	End of school reading performance rankings	RFAL is used by top performing schools in Utah when schools are ranked based on average end-of-level test scores in reading and math, and by the 2 <sup>nd</sup> highest performing school district (Cache School District, 57% proficient in English/Language Arts) with an average per pupil expenditure (\$6,972) about 59% that of the top-performing district (Summit, , 59% proficient in English/Language Arts, \$11,815 per pupil, per Utah Taxpayers Association, 2015).
Lignugaris/ Kraft, et al., (2001); Cache County School District in Utah	45	16 Weeks	Twice as many children in intervention group scored in the top quartile of the WRMT-R than the comparison group. The intervention group scored 87% higher than comparison group on curriculum-based measures.
READ Alliance, New York City (2015)	1,055	50 sessions, 45 mins	74% of participants made gains of greater than one full grade level on state assessment. 94% made gains of one half to a full grade level.

## B. Description of Comparison

Scott Foresman Early Reading Intervention (ERI) was used for instruction with comparison group students in kindergarten and first grades. ERI is a widely used program that is generally accepted as an effective reading intervention program. Though research to demonstrate the effectiveness of ERI cannot be located on the What Works Clearinghouse or Best Evidence Encyclopedia websites, it was widely used throughout the country in Reading First schools and is widely used throughout Utah school districts as both a core reading program in kindergarten and/or a Tier 2 intervention program in kindergarten and first grades.

On the Scott Foresman website, the publishers make the following claims about ERI (<http://www.pearsonschool.com/index.cfm?locator=PSZu68&PMDbProgramID=13301>):

Scott Foresman Early Reading Intervention is based on Project Optimize, a five-year longitudinal research study by Dr. Edward J. Kame'enui and Dr. Deborah C. Simmons. It identifies at-risk children in kindergarten and grade 1 and provides

intervention to improve reading achievement. Research shows 97% of kindergarten children who were taught with our program experienced faster achievement rates and were able to sustain that level of achievement into second grade.

In 2006, Deborah Simmons, one of the authors of ERI, was awarded a nearly \$2.9 million, 5-year grant from the Institute for Education Science to investigate the efficacy of ERI when compared to other school-based reading interventions. At that time, Simmons claimed that ERI was used in “4,000 school districts in all 50 states.” The research was based in Texas, Florida and Connecticut. In 2011, Simmons and other authors (Hagen-Burke, et al., 2011) published findings from the study indicating that students’ entry level abilities as well as teacher experiences and knowledge and group size explained student outcomes to a greater extent than curricula or instructional materials. Additionally, findings indicated that interactions between learner characteristics and intervention features resulted in no differences between ERI and school-based interventions. Subsequent publications (Coyne, et al., 2013; Hagen-Burke, et al., 2011; Simmons, et al., 2014) also describe findings from the research funded by IES. For example, Hagen-Burke with Simmons and others reported that rapid automatized naming, student problem behavior, and phonological awareness were the strongest predictors of reading performance in the research, far outweighing treatment or control group membership, which was not statistically significant as a predictor. Coyne with Simmons and others reported results from the initial study (n = 206 in Texas and Connecticut) and a replication study (n = 162 in Florida). The initial study showed statistically significant findings with moderate effect sizes (i.e., 0.40-0.51) favoring ERI on several tests of reading skills (letter sounds, phonemic awareness, phonemic segmentation, sound blending, word attack), while the replication study produced no statistically significant findings and near zero effect sizes. In fact, they stated, “Multilevel hierarchical linear analyses revealed no statistically significant differences between conditions [treatment and control group membership] on any measure.” They concluded that context matters and quality of instruction, instructional conditions, teachers’ experiences, and degree of implementation, but not research group membership, predicted outcomes.

Overall, the results of the IES-funded research (Simmons, et. al, 2013) indicated that in multiple studies with sample sizes of approximately 100-200 students carried out in three states comparing ERI with other school-based Tier 2 interventions, context features predicted end-of-kindergarten outcomes with moderate effect sizes, regardless of research group membership. In second grade with the same students, initial kindergarten letter identification skills still predicted outcomes better than initial group membership or other literacy skills measures. However, second grade outcomes for students who were at greater risk due to initial literacy skills and English learner status were moderated by the type of intervention (i.e., more or less systematic and explicit), though not by ERI or comparison group status, per se.

ERI requires extensive use of manipulatives which must be accurately pulled from kit materials prior to each lesson. Teachers reported that the preparation required was about 20 minutes per lesson (which included placing the previous lesson’s materials back in the kit).

All teachers and paraprofessionals had received extensive training for ERI from the publisher’s regional training representatives prior to the school year preceding the study, and most had been using ERI at the school for at least the previous two school years. Additionally, school staff had



attended local Utah reading conferences that included additional training for implementing ERI with fidelity. The literacy coaches and several teachers had attended this state-provided training and continued to provide in-school professional development and coaching to reading staff.

### C. Setting

The study was conducted in a single elementary school in an urban district in Utah. With a total population in Utah nearing 2 million, the Wasatch Front region between Brigham City and Spanish Fork covers 80% of Utah's population and includes Ogden, Salt Lake City, and Provo/Orem. The participating school is geographically near the center of this long but relatively narrow urban region. The National Center for Education Statistics (NCES) data show that this school had approximately 550 students in kindergarten through 6<sup>th</sup> grade in 2013-2014, with a student-teacher ratio of 23. There are between 80 and 100 students per grade in kindergarten through 6<sup>th</sup> grades. The school receives Title I funds and is considered the most highly impacted school in the district with the greatest diversity in student ethnicity. Over 60% of the students in the school are from homes considered economically disadvantaged, with the NCES listing the percentage of students eligible for free or reduced lunch as 99%. Over 40% of the students are non-White including a large percentage of immigrants and refugees from Mexico and many other countries, including Middle Eastern, African, and Latin American communities. Over 16% are English language learners (ELL) and 14% receive special education services.

### D. Participants

All children enrolled in 1<sup>st</sup> grade who were enrolled in school during beginning of year DIBELS benchmark testing and who scored below grade level (strategic or intensive needs in reading based on the DIBELS beginning of year benchmark assessment) were included in the study, per the school's agreement with families about conducting research to best meet students' needs. Students who enrolled after beginning of year assessments were assigned to appropriate small groups per placement testing results, with new students alternately assigned to treatment or comparison groups to ensure consistent sizing of groups and to minimize disruption from new students. However, students who enrolled after random assignment to groups were not included in the analysis reported here.

Students with disabilities were also included in the study, with these students randomly assigned to treatment or comparison groups separately from students not identified with disabilities. Students with disabilities who were best taught in resource room settings were also randomly assigned to treatment or comparison groups, with those students taught either one-on-one or in small groups of two students.

Students who were assigned to treatment or comparison groups based on DIBELS beginning of year benchmark composite scores included 30% Hispanic and 8% refugee or immigrant students from other countries. Over 25% spoke a home language other than English. According to the NCES, 99% of students qualified for free and reduced lunch in the 2013-2014 school year. Approximately 11% of the students who participated in the study had been identified for special services other than speech language services.

Only children from whom signed parental permission was received participated in the additional Woodcock Johnson testing at the end of the year, though DIBELS end of year benchmark scores were obtained for all students who participated in the study.

## II. Study Design and Analysis

### A. Sample Formation

After beginning of year DIBELS benchmark assessments were completed, researchers met with literacy specialists at the school to randomly assign students to groups. All students who performed below grade level based on the DIBELS composite score were selected. Those students were grouped according to similar patterns of scores on the DIBELS subtests. For example, first grade students needing strategic support based on the composite score were grouped together based on levels of risk on the individual subtests: Phonemic Segmentation and Nonsense Words Fluency. Those who needed strategic support on both subtests were placed in one group, while those who needed intensive support on Nonsense Words Fluency but strategic support on Phonemic Segmentation Fluency were placed in another group. Those with intensive needs on both subtests were placed in yet another group. Students were ordered within groups according to DIBELS composite scores. Then, the first two students from a group were selected. A coin was flipped to determine if the first student was assigned to the treatment group (heads) or the comparison group (tails). The other student was assigned to the other group. Then the next two students were selected and again, the coin was tossed. If there was an odd number of students in the original group, the coin was tossed to determine the assignment of the last unpaired student. In this way, the odds for each student of being assigned to the treatment group was 50%.

### B. Outcome Measures

The DIBELS Next benchmark assessments were used as both a beginning of year measure to stratify students prior to random assignment to groups, and as a covariate in statistical analyses. Middle and end of year DIBELS benchmark assessments were used as outcome measures. A trained district-level team traveled from school to school in the district to administer DIBELS benchmark assessments. Both composite and subtest scores were considered in the analyses.

The DIBELS Next assessments are standardized measures widely used throughout the U.S. as screening, progress monitoring and outcome measures. Good, et al. (2013) extensively document the technical adequacy of the DIBELS Next subtests and composite scores.

DIBELS Next scores used in the outcome analysis for this report include the following:

- Composite score for 1<sup>st</sup> grade
- Nonsense word fluency (NWF) for 1<sup>st</sup> grades
- Oral reading fluency correct words per minute (DORF) for 1<sup>st</sup> grade
- Oral reading accuracy (DORAcc) for 1<sup>st</sup> grade

### C. Analytical Approach

This study included random assignment to treatment and comparison groups of those students who fell below grade level on the beginning of year DIBELS benchmark assessment in one elementary school. As such, the sample size for the number of students participating in the study is adequate to achieve sufficient statistical power when analyzed using student as the unit of analysis. Using cluster sizes of small group or classroom would not be feasible for power requirements with this sample. To help account for teacher effects, teachers and paraprofessionals taught small groups of students using both the treatment (RFAL) and the comparison (ERI) programs, having been trained to use both.

Several analyses were conducted. Based on the assumption that random assignment to groups created “equal” groups, analysis of variance statistical tests were conducted for each outcome measure. Additionally, to account for baseline group differences in beginning of year assessments despite random assignment to groups, linear regression on the outcome variable with beginning of year composite scores and group (treatment or comparison) as independent variables was conducted to determine the proportion of variability in the outcome that could be accounted for by beginning of year reading skills and group membership.

### D. Statistical Adjustments

Sample sizes, means and standard deviations for study data are included in section III. Because groups were equivalent at baseline (after random assignment to groups), no statistical adjustments were made. Additionally, no adjustments were calculated to account for clustering. A Bonferroni correction could be used to account for multiple comparisons within the same domain. A Bonferroni correction adjusts the p-value (or confidence interval) by dividing by the number of comparisons. However, actual p-values are reported below, and because all are larger than .05 and, hence, not statistically significant when uncorrected, a Bonferroni correction was not necessary.

#### *Sample Attrition (Table 0a)*

Table 0a. Sample Attrition Rates

Time	Measure	Intervention Group		Comparison Group		Difference in Attrition Rates	Overall Attrition Rates
		Sample Size	Attrition Rate	Sample Size	Attrition Rate		
Pre-Intervention (BOY)	DCS (1 <sup>st</sup> grade)	16		17			
Post-Intervention (MOY)	DCS (1 <sup>st</sup> grade)	16	0%	17	0%	0	0%

### E. Missing Data

There were no missing data. No attrition in the samples occurred between beginning and middle of year.

### III. Study Data

The following abbreviations are used for subtest names:

- DIBELS Composite Score (DCS)
- DIBELS Nonsense Words Fluency Correct Letter Sounds (CLS)
- DIBELS Nonsense Words Fluency Whole Words Read (WWR)
- DIBELS Oral Reading Fluency (DORF)
- DIBELS Oral Reading Accuracy (DORAcc)

#### A. Pre-Intervention Data, Baseline Sample (Table 1)

Table 1. Pre-Intervention Sample Sizes and Characteristics\* for the Baseline Sample

Baseline Measures	RFAL Intervention Group				ERI Comparison Group			
	Sample Sizes		Sample Characteristics		Sample Sizes		Sample Characteristics	
	Unit of Assignment	Unit of Analysis	Mean	Standard Deviation	Unit of Assignment	Unit of Analysis	Mean	Standard Deviation
DCS.BOY	16	16	88.56	25.822	17	17	87.29	27.499
CLS.BOY	16	16	20.06	7.188	17	17	20.12	8.659
WWR.BOY	16	16	0.44	0.727	17	17	0.59	1.004
DORF.BOY	Not administered at beginning of year							
DORAcc.BOY	Not administered at beginning of year							

#### B. Pre-Intervention Data, Analysis Sample (Table 2)

For the following tables, random assignment to groups was maintained from beginning of year to middle of year.

Table 2a.  
Pre-Intervention Sample Sizes and Characteristics for the Middle of Year Analytic Sample

Baseline Measures	RFAL Intervention Group				ERI Comparison Group			
	Sample Sizes		Sample Characteristics		Sample Sizes		Sample Characteristics	
	Unit of Assignment	Unit of Analysis	Mean	Standard Deviation	Unit of Assignment	Unit of Analysis	Mean	Standard Deviation
DCS.BOY	16	16	88.56	25.822	17	17	87.29	27.499
CLS.BOY	16	16	20.06	7.188	17	17	20.12	8.659
WWR.BOY	16	16	.44	.727	17	17	.59	1.004
DORF.BOY	Not administered at beginning of year							
DORAcc.BOY	Not administered at beginning of year							

Independent samples t-test results indicate there are not statistically significant differences at beginning of year between treatment and comparison groups on DIBELS subtests.

Independent Samples T-Test											
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI of the Difference		
										Lower	Upper
DCS.BOY	EVA	.035	.853	.136	31	<b>.892</b>	1.268	9.300	-17.700	20.237	
CLS.BOY	EVA	.271	.606	-.020	31	<b>.984</b>	-.055	2.780	-5.725	5.615	
WWR.BOY	EVA	2.046	.163	-.491	31	<b>.627</b>	-.151	.307	-.777	.475	

EVA: Equal variances assumed    CI: Confidence Interval

### C. Post-Intervention Data and Findings (Table 3)

For the following tables, random assignment to groups was maintained from beginning of year to middle of year.

Table 3a. Post-Intervention Outcomes for Analytic Sample and Estimated Effects

Outcome Measures	RFAL Treatment Group		ERI Comparison Group		Estimated Effects	
	Mean	Standard Deviation	Mean	Standard Deviation	SMDES	p-value
DCS.MOY	123.56	78.490	96.47	55.948	.40	.260
CLS.MOY	52.50	20.106	51.12	21.459	.07	.850
WWR.MOY	11.56	9.352	8.29	9.866	.34	.337
DORF.MOY	24.63	22.894	12.71	7.556	.71	.063
DORAcc.MOY	65.06	19.726	55.47	21.881	.46	.197

SMDES: Standardized Mean Difference Effect Size Using Pooled Standard Deviation

Independent samples t-test results indicate there are not statistically significant differences between treatment and comparison groups on DIBELS subtests, though differences approached statistical significance (without a Bonferroni correction) for DIBELS oral reading fluency ( $p = .063$  given not equal variances), in favor of the RFAL treatment group.

		Independent Samples T-Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI of the Difference	
								Lower	Upper	
DCS.BOY	EVA	1.672	.206	1.147	31	<b>.260</b>	27.092	23.615	-21.071	75.255
CLS.BOY	EVA	.007	.936	.191	31	<b>.850</b>	1.382	7.250	-13.405	16.170
WWR.BOY	EVA	.028	.868	.975	31	<b>.337</b>	3.268	3.351	-3.566	10.103
DORF.MOY	EVA	7.589	<b>.010</b>	2.034	31	.051	11.919	5.860	-.033	23.871
	EVNA			1.983	18.056	<b>.063</b>	11.919	6.010	-.704	24.542
DORAcc.MOY	EVA	.036	.850	1.320	31	<b>.197</b>	9.592	7.268	-5.231	24.415

EVA: Equal variances assumed    EVNA: Equal variances not assumed    CI: Confidence Interval

Figure 1 on the next page shows treatment and comparison group means for beginning and end of year DIBELS scores, as well as on grade level cutoffs at each time point.

#### IV. Conclusions

This small sample study with random assignment to groups included a large enough sample for sufficient statistical power with student as the unit of analysis. Teacher effects were controlled for by having teachers teach both treatment and comparison intervention groups: RFAL and ERI. Random assignment to groups was maintained for half the school year, at which point the school decided to change most students to RFAL groups. This decision was based on (a) greater gains in the RFAL treatment group and (b) teachers' preferences for RFAL, despite their extensive training and use of ERI prior to this study. They preferred RFAL because preparation time was substantially less each day, they felt lessons were easier to deliver with better results, and they liked the sentence variation in RFAL that required students to read without being able to infer or guess based on sentence patterns.

Because groups were equivalent at baseline based on both composite and subtest scores, findings reported here included only t-tests of group differences (which were not statistically significantly different) and calculation of standardized mean difference effect sizes (SMDDES). Though the group differences were not statistically significantly different, all differences favored the RFAL group. Additionally, SMDDES were large for DIBELS Oral Reading Fluency, and moderate for the DIBELS Composite Score and Nonsense Word Fluency. Finally, both groups exceeded the benchmark cutoffs for the Nonsense Words Fluency subtests. Though neither group, on average, reached grade level cutoffs at middle of year, the RFAL group approached the benchmark cutoff for reading on grade level based on the DIBELS Composite Score and exceeded the Oral Reading Fluency benchmark cutoff, while the ERI group was, on average, much further below these cutoffs.

Study findings suggest that RFAL helped to change struggling students' trajectories quickly with potentially better outcomes based on the short duration of the study. Additionally, given the wide spread use of ERI and the reputation of the curriculum's authors (Kame'enui and Simmons), initial hopes were that RFAL might do as well as ERI, though study teachers initially thought otherwise. Findings suggest that participation in a larger scale study could be warranted.

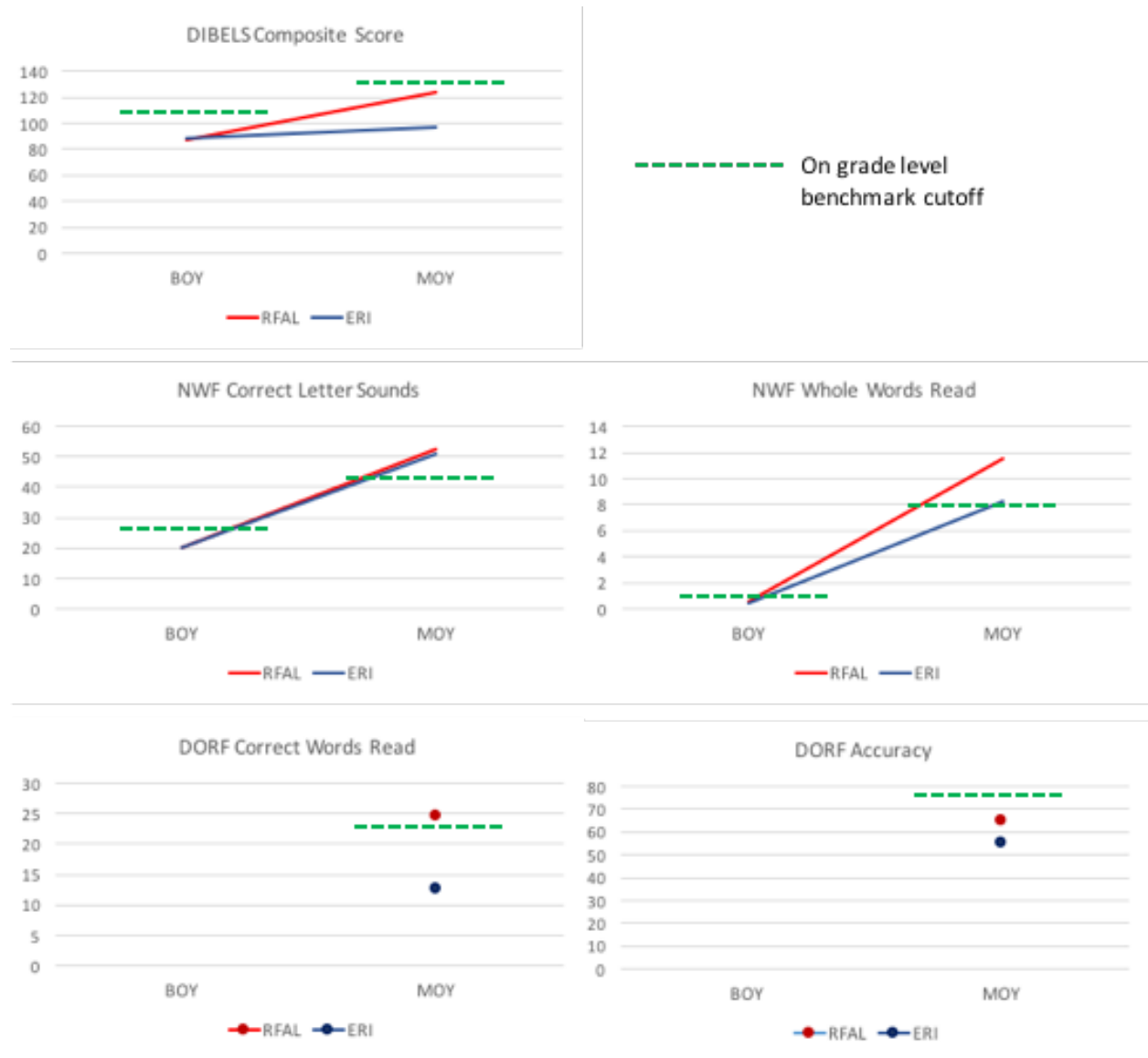


Figure 1. Treatment and comparison group means at beginning and middle of year

## VI. References

- Coyne, M. D., Little, M., Rawlinson, C., Simmons, D., Kwok, O., Kim, M., Simmons, L., Hagan-Burke, S., & Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, 6(1), 1-23. DOI: 10.1080/19345747.2012.706694
- Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Simmons, L. E., Kwok, O., Kim, M., Fogarty, M., Oslund, E., Taylor, A., Capozzoli-Oldham, A., Ware, S., Little, M. E., & Rawlinson, D. M. (2013). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children*, 80, 25–44.
- Hagan-Burke, S., Coyne, M. D., Kwok, O., Simmons, D. C., Kim, M., Simmons, L. E., Skidmore, S. T., Hernandez, C. L., & Ruby, M. M. (2011). The effects and interactions of student, teacher, and setting variables on reading outcomes for kindergarteners receiving supplemental reading intervention. *Journal of Learning Disabilities*, 46(3), 260–277. DOI: 10.1177/0022219411420571
- Hanson, R. A., & Farrell, D. (1995). The long-term effects on high school seniors of learning to read in kindergarten. *Reading Research Quarterly*, 30(4), 908-933.
- Lignugaris/Kraft, B., Findlay, P., Major, J., Gilberts, G. & Hofmeister, A. (2001). The association between a home reading program and young children's early reading skill. *Journal of Direct Instruction*, 1(2), 117-136.
- National Reading Panel (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*. Washington D.C.: National Institute of Child Health and Human Development, National Institutes of Health.
- Simmons, D. C., Taylor, A. B., Oslund, E. L., Simmons, L. E., Coyne, M. D., Little, M. E., Rawlinson, D. M., Hagan-Burke, S., Kwok, O., & Kim, M. (2014). Predictors of at-risk kindergarteners' later reading difficulty: Examining learner-by intervention interactions. *Reading and Writing*, 27(3), 451-479.
- Skenandore, K., Hofmeister, A., & Willis, P. (2003, November). *The Cibecue Community School project: A report on the design, implementation and results*. Paper presented at the Bureau of Indian Affairs Office of Indian Education Programs National Meeting, Denver, CO.