Learning Disabilities Research & Practice, *27*(4), 189–202

A Comparison of Responsive Interventions on Kindergarteners' Early Reading Achievement

Mary E. Little
University of Central Florida

D'Ann Rawlinson
UCF Morgridge International Reading Center

Deborah C. Simmons, Minjung Kim, Oi-man Kwok, Shanna Hagan-Burke, Leslie E. Simmons, Melissa Fogarty, and Eric Oslund
Texas A&M University

Michael D. Coyne
University of Connecticut

This study compared the effects of Tier 2 reading interventions that operated in response-to-intervention contexts. Kindergarten children ($N = 90$) who were identified as at risk for reading difficulties were stratified by school and randomly assigned to receive (a) Early Reading Intervention (ERI; Pearson/Scott Foresman, 2004) modified in response to student performance or (b) their schools' typical supplemental reading intervention (regrouping and curriculum pacing adjustments). In both conditions, intervention was provided 30 minutes per day in small groups for approximately 100 sessions. Results indicated no statistically significant group differences on any outcome measures. Between-group effect sizes revealed substantively important differences (Valentine & Cooper, 2003) favoring the ERI responsive condition on multiple measures with effect sizes ranging from .35 to .59. Overall, findings indicated that the majority of students in both Tier 2 intervention conditions performed above the 30th percentile on posttest measures of word reading measures.

Over the past decade, Response to Intervention (RTI) has emerged to promote proactive, data-based instructional decisions to preempt and intercept early learning difficulties. RTI provides a framework for assisting students who require support beyond typical classroom instruction in reading or mathematics (Lembke, McMaster, & Stecker, 2010). Although features vary across RTI models (Fuchs, Mock, Morgan, & Young, 2003; Marston et al., 2007; Tilly, 2006; Vaughn & Chard, 2006), the RTI framework includes several common features. RTI models are organized in multiple tiers of instructional support, beginning with Tier 1 classroom instruction and progressing to increasingly more intensive and/or explicit interventions based on student response to Tier 2 and 3 instruction. Screening and progress monitoring assessments provide data that enable teachers to quantify students' RTI. Beyond these universal features, researchers have identified a range of variables that can be manipulated to intensify intervention efficacy (e.g., Mellard, McKnight & Jordan, 2010).

Despite the growing use of RTI in schools, few experimental studies have compared students' response to Tier 2 interventions that were systematically intensified using student performance data. This article reports findings of an experimental study that compared school-designed intervention to a commercial intervention, both of which were conducted in response-to-intervention contexts.

RTI Research and Practice Context

Even as researchers continue to develop and study the dimensions and implementation of RTI, many schools, districts, and states are moving quickly toward wide-scale implementation (Vellutino, Scanlon, Small, & Fanuele, 2006). According to a recent national study on implementation of the Individuals with Disabilities Education Act, 71 percent of elementary schools in the United States use a version of RTI in reading/language arts (Bradley et al., 2011). In practice, RTI components are largely based on findings from controlled research studies, but additional evidence supporting comprehensive RTI models is needed from applied settings (Lembke et al., 2010). While RTI has become part of educational nomenclature and practice, questions remain regarding how to effectively intensify intervention for children who do not evidence satisfactory response to early reading intervention.

Requests for reprints should be sent to Mary E. Little, University of Central Florida. Electronic inquiries should be sent to Mary.Little@ucf.edu.

The primary grades provide a critical window of opportunity in which early intervention differentially accelerates reading growth compared to later intervention for children with early reading risk (Vellutino et al., 1996; Vellutino, Scanlon, & Tanzman, 1998). In the IES practice guide on RTI in the primary grades, Gersten et al. (2009) concluded there was strong evidence to their recommendation of providing supplemental, intensive, and systematic small group instruction for students who score below benchmarks on universal screening. However, the research base is less clear on how to intensify intervention for students who do not demonstrate adequate response to Tier 2 interventions (Gersten et al., 2009).

One of the challenges of RTI research is isolating the components that are associated with or accountable for student response. By design, RTI approaches, or independent variables, are instructional packages that bundle multiple components. Mellard and colleagues (2010), for example, recommended that educators consider 10 variables when determining how to intensify instruction. "These variables include three dosage-related elements (minutes of instruction, frequency, and duration), as well as instructional group size, immediacy of corrective feedback, the mastery requirements of the content, the number of response opportunities, the number of transitions among contents or classes, the specificity and focus of curricular goals, and instructor specialty and skills" (p. 219). At issue is which combination of alterable variables positively improve student achievement and how to operationalize them in instruction. In addition to the various instructional components that are manipulated in response to student performance, the processes used to modify and intensify intervention vary considerably (Gersten et al., 2005).

The two most prominent RTI models are the problem-solving and the standard protocol. The problem-solving model emphasizes individualized interventions that focus equal attention on the learner, the environment, and the curriculum (Tilly, 2006). Thus, teams within a school collaboratively review student performance data and develop interventions to solve problems (Marston et al., 2007). The standard protocol model involves the delivery of evidence-based, multi-component programs with a strong research base focused on specific skill areas (Fuchs et al., 2003). Standard protocols are designed to be structured and explicit and to be delivered to small groups of students. Groups are identified by matching student needs to the particular protocol. Educators use a specified set of criteria or procedures to adjust intervention. For example, students may receive a specified intervention for a period of 8 weeks, and instructional adjustments (e.g., intensify intervention, dismiss from intervention) are made based on student performance data (Little, 2009).

An emerging model for RTI is the hybrid or blended model, which uses components and processes from both the problem-solving and the standard model (Denton et al., 2010; VanDerHeyden, 2011; VanDerHeyden & Snyder, 2006). Hybrid models maintain technical adequacy with a particular focus on decision-making analyses of contextual and instructional variables. An example hybrid approach to RTI was recently investigated in an efficacy study by Denton and colleagues (2010). In this study, the authors used an evidence-based intervention (Responsive Reading Instruction; Denton, 2001; Denton & Hocker, 2006) that incorporated "program manualization which included instructions for using student assessment data to plan lessons" (Denton et al., 2010, p. 400). The results of this responsive intervention indicated that the treatment group had a significantly higher end-of-year status and rate of growth in oral reading fluency than peers receiving typical practice; however, the majority of students in the comparison condition did not receive supplemental intervention.

Simplifying the Complexity of RTI

Given the complexities of RTI, we were interested in determining how to streamline the various variables in a model in a way that is both effective and feasible. Our approach began with (a) an intervention that has demonstrated positive effects in previous experimental studies and (b) formative assessments that evaluated students' mastery of content taught every 4 weeks. Specifically, we evaluated student performance every 4 weeks, identified skills and content not mastered by individual children, and used a standard protocol or procedure to intensify instruction that responded to student performance. For example, student performance was evaluated, and children with similar performance profiles were grouped and regrouped dynamically at seven points during the intervention. At each measurement point, assessment data were used to modify not only grouping, but curriculum pacing and instructional foci. Students who demonstrated strong and consistent performance were accelerated in the program and placed in a different instructional group. Students who demonstrated weak response received modified curriculum by repeating specific lessons and reinforcing specific content or skills for all students in their small group. Thus, we sought to address the needs of individual students using standardized methods of dynamic regrouping and curricular pacing.

In a previous study (authors), we compared the effects of adjusting ERI instruction based on student performance to ERI instruction without adjustments in a randomized control trial. Participants included 103 kindergarten students identified as at risk for reading difficulties. Results of multilevel hierarchical linear modeling (HLM) analyses conducted on assessments administered at the end of kindergarten indicated that students in the experimental group outperformed those in the comparison group on all posttest measures, with statistically significant differences on measures of letter knowledge, phonemic blending, word reading, spelling, and oral reading fluency. Effect sizes on the full battery of measures ranged from .22 to .69. Findings indicated that instructional adjustments enhance program efficacy beyond what would be gained from program implementation alone. The current study was designed to extend this research to determine whether different types of Tier 2 interventions that were allowed to modify instruction in response to student performance would differentially impact student performance.

Overview and Purposes of the Study

Our interest in this study was to extend our understanding of Tier 2 kindergarten interventions, specifically examining the impact of an RTI approach that used a standardized protocol to adjust instruction based on student performance. We were interested in whether a standardized intervention that systematically used data to adjust instruction would be more beneficial than a Tier 2 intervention that varied during implementation by school personnel. Despite the adoption of research-based reading interventions with similar components that met state and district policies and mandates, our research interest focused on fidelity of classroom implementation and impact on student achievement. In addition, despite the emerging evidence base on RTI, there is limited research regarding how to effectively intensify intervention (RTI practice guide, Gersten et al., 2009). Our research was motivated by the limited experimental research that has investigated methods to intensify intervention using progress monitoring data to adjust instruction for students in Tier 2 intervention. To study Tier 2 interventions, the standard response protocol was used in conjunction with a commercial reading program (Early Reading Intervention; Pearson/Scott Foresman, 2004) and compared to school-designed interventions that operated within an RTI framework. We sought to extend prior Tier 2 intervention research (e.g., Denton et al., 2010) by involving all students in the comparison group in interventions that were allowed to vary by school.

The research questions addressed in the study were:

1. What are the statistical and practical effects of Tier 2 interventions (commercial program-ERI) and school-designed (SDI) on measures of kindergarten students' phonemic awareness, alphabetic understanding, word attack, word identification, and spelling?
2. What are the effects of the Tier 2 interventions (ERI and SDI) on the absolute level of student achievement of kindergarten children identified as at risk of reading difficulty?

TABLE 1
Interventionist Demographics by Condition

	SDI		*ERI & SDI*		*ERI*	
	($N = 5$)		($N = 1$)		($N = 15$)	
Variable	*N*	%	*n*	%	*n*	%
Female Teachers	5	100.0	1	100.0	15	100.0
Ethnicity						
Black or African American	0	0.0	0	0.0	1	6.7
White	5	100.0	1	100.0	14	93.3
Latino or Hispanic	0	0.0	0	0.0	0	0.0
Highest Degree Earned						
High School/Less than Bachelor's	1	20.0	0	0.0	0	0.0
Bachelor's	3	60.0	1	100.0	6	40.0
Master's	1	20.0	0	0.0	8	53.3
Ed.S.	0	0.0	0	0.0	1	6.7
Primary Language						
English	5	100.0	1	100.0	15	100.0
Total Years' Teaching Experience						
Mean		16.00		4.00		10.13
SD		9.41		–		5.83

Note: One interventionist who implemented both conditions was not included in the chi-square difference test.There were no significant differences on any variables after adjusting Benjamini-Hochberg correction. ERI = Early Reading Intervention; SDI = school-designed intervention.

METHOD

Schools, Interventionists, and Students

The study was conducted in a large school district in central Florida. Schools represented a mixture of suburban and rural neighborhoods. All schools received Title I services. The percentage of students who were eligible for free or reduced-price lunch ranged from 63 to 92 percent. School enrollments ranged from 401 to 832. Schools in the participating district had established practices for implementing problem solving teams within an RTI model. Bi-weekly, teachers, instructional reading coaches, and curriculum specialists held "data meetings" to review student assessment data from multiple sources. The teams' role was to study individual, class-wide, or school-wide academic problems. These school-based teams were intervention-driven; members designed and continuously monitored the effects of instruction and interventions based upon assessment data of individual targeted students. Schools were at various stages of implementation during the experimental study period.

In total, 21 reading interventionists and 90 students from 8 elementary schools participated. Research staff met with the principals, reading interventionists, and kindergarten teams of each school to explain the study, the research design, and the need for randomization. Participating schools designated two to three interventionists for their kindergarten team. Five out of the eight schools implemented ERI in previous years. Two levels of randomization occurred. First, students were randomly assigned to either ERI or SDI, then reading interventionists were assigned within their school to either the ERI ($n = 15$) or SDI ($n = 5$) based on their previous experience and familiarity with the ERI program. One interventionist participated in both conditions.

Both intervention conditions were present in six of the eight schools. Two of the schools only participated in the ERI condition as the ERI curriculum was already being used as their school-designed intervention. All students participating in the study were then randomly assigned within schools to either treatment (ERI; $n = 16$) or comparison (SDI; $n = 6$) groups. By design, the number of groups was greater for the ERI condition, because a minimum of two groups were needed to implement regrouping to intensify intervention. Reading specialists ($n = 12$) and kindergarten classroom teachers ($n = 9$) served as interventionists. Demographic data on interventionists are presented in Table 1.

A two-phase process was used to identify children who were eligible to participate in the study. In the second month of kindergarten, researchers consulted with school personnel to identify children who (a) were considered in need of supplemental, small-group reading instruction; (b) were at least 5 years of age; and (c) received reading instruction in English. School personnel examined existing school-administered reading measures, the Florida Assessments for Instruction in Reading (FAIR). This state-wide K-12 reading assessment system, developed by the Florida Department of Education (FLDOE), provides teachers with screening, diagnostic, and progress monitoring information that is essential to guiding instruction (FLDOE, 2009).

Kindergarten teachers nominated up to six children most at risk of reading difficulty per kindergarten classroom for further assessment based upon individual student results on the FAIR. Permission forms were distributed to all nominated students. Of 234 nominated students, 227 permission forms were returned. Next, all students with consent were screened using the Letter Naming Fluency (LNF) subtest from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) and the Sound-Matching (SM) subtest from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999). Screening measures were selected based on strong predictive validity for the end-of-first-grade reading outcomes and their use in recent kindergarten intervention studies (Authors; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004).

Students were determined to be at risk and qualified to participate in the study by meeting either of the following criteria: (a) a score at or below the 33rd percentile on the DIBELS LNF measure (i.e., fewer than 6 letters correctly named in 1 minute; Good & Kaminski, 2002), or (b) a score below the 37th percentile on the CTOPP SM subtest (Wagner et al., 1999). These measures were selected as initial screeners based on their use in recent kindergarten intervention studies and their long term predictive validity for end-of-first-grade reading outcomes (Schatschneider et al., 2004). Pretest scores indicated that while some students exceeded the benchmark on phonemic awareness, their letter naming scores were low (i.e., mean pretest score of less than one correct letter). Therefore, we elected to include children who were low letter identification as it is a reliably strong predictor of reading difficulty (Schatschneider et al., 2004. To progress to the next step of the screening process, students had to perform $\leq$ the 36th percentile on DIBELS letter naming fluency and $\leq$ 37th percentile on CTOPP sound matching. All students in the study were then administered the Peabody Picture Vocabulary Test-Third Edition (PPVT-III; Dunn & Dunn, 1997) to determine their baseline level of receptive vocabulary knowledge. Twenty-six percent of students receiving interventions received school designed interventions (SDI; $n = 23$). Within our sample, there were 6 total students with disabilities in ERI condition (4 speech/language, 1 developmental delay, 1 specific learning disability) and 3 total in SDI (all speech/language). Student demographic data are presented in Table 2.

TABLE 2
Student Demographics by Treatment Condition

	ERI		*SDI*	
	($N = 67$)		($N = 23$)	
Variable	*n*	(%)	*n*	(%)
Gender				
Male	30	44.8	9	39.1
Female	37	55.2	14	60.9
Ethnicity				
Black or African-American	25	37.3	5	21.7
Hispanic or Latino	13	19.4	8	34.8
White	24	35.8	10	43.5
Other	5	7.5	0	0.0
Receive Special Education Services	6	9.0	3	13.0
Bilingual/English Language Learner	11	16.4	5	21.7
Age				
Mean	5.51		5.55	
SD	0.29		0.35	

Note: ERI=Early Reading Intervention; SDI=school-designed intervention.

Procedures

Core Reading Instruction

All elementary schools in this Florida school district implemented a consistent core reading program, Macmillan/McGraw-Hill *Treasures* (Bear et al., 2009). Macmillan/McGraw-Hill *Treasures* materials included with the program are the textbook anthology, leveled texts for differentiated instruction, and *Triumphs* intervention kits for additional reading instruction and support. By state and district policies, all kindergarten students received district-prescribed core reading instruction for 90 minutes each day.

Common Intervention Components

To increase comparability between conditions, a number of common instructional components were standardized across the ERI and SDI conditions. That is, groups in both conditions were comprised of five students. Interventionists in both conditions were asked to meet with their groups for 30 minutes, 5 days per week over the course of the intervention period for an equivalent number of total sessions. Content in both conditions focused on early literacy skills, with ERI interventionists implementing the ERI curriculum and comparison interventionists implementing school-designed intervention focusing on early literacy skills. The interventions were implemented outside of the 90-minute required reading block and did not interfere with the core reading instruction.

ERI Condition

The ERI condition is based on the ERI curriculum and complemented by standardized RTI modifications. The ERI curriculum is comprised of 126 daily lessons. Each lesson is designed to be delivered in 30 minutes, and consists of seven instructional activities that focus on phonological awareness, alphabetic understanding, writing, spelling, and reading. The ERI curriculum is organized into four major components and follows a systematic instructional sequence. Thus, the lesson design includes detailed scripting that guides the interventionist in incorporating modeling, practice, review, and feedback into the daily activities. Consistent language is embedded throughout the curriculum to ensure clear communication of information and reduce variability in implementation.

All ERI students entered the study beginning at Lesson 1 of the curriculum and were grouped according to assessment data from screening and pretest measures. An ERI instructional management plan that included standardized decision rules was used to make instructional adjustments based on students' performance. Instructional adjustments in ERI consisted of four components: (a) use of in-program assessments at the middle and end of each part of the program to evaluate content mastery and identify specific skills for targeted instruction, (b) student regrouping based on curriculum mastery, (c) curriculum progression (i.e., repetition or acceleration in the program), and (d) targeted review of critical skills.

Approximately every 4 weeks, members of the research team administered ERI in-program assessments to determine if students were making adequate progress. Instructional adjustment decisions were based on students' assessment percentages and subtest performances.

Students scoring 90 percent or above on in-program assessments were considered as demonstrating a strong response, and therefore continued with the normal lesson progression. Students with scores of 90 percent or higher on two consecutive in-program assessments were accelerated in the curriculum using guidelines provided in the program. Acceleration schedules generally involved students completing two of every three lessons until they reached the final program component, in which they were taught each lesson. Students scoring 70 percent to 89 percent were considered to demonstrate a moderate response, and continued with the standard lesson progression. If students in this group did not know letter sounds taught in previous lessons, targeted letter sound instruction and review was incorporated in subsequent lessons. Students scoring 69 percent and below were considered as demonstrating a weak response. Instructional adjustments for this group included reteaching targeted lessons and adding a modified letter sound review activity. Students returned to standard lesson progression after reviewing the targeted lessons until the next in-program assessment.

Once the assessment data and additional factors had been reviewed, the interventionists and members of the research team met to identify whether and how to adjust instruction. Adjustments included student regrouping, lesson reteaching, lesson acceleration, and targeted skill reviews.

ERI Professional Development

Professional development was designed to approximate the type of training that is typical when schools adopt published programs, and was limited to 2 days. The first professional development occurred before the start of the intervention in late September. The goals of this session were to (a) develop an awareness of the purpose of the research project, (b) orient the interventionists to the design of the ERI program, (c) introduce the ERI curriculum and materials, (d) provide guidance for implementation of the intervention, and (e) introduce the assessment and instructional management plan that was unique to the ERI condition. The professional development familiarized interventionists with lesson structure and the scope and sequence of Parts 1 and 2 of the four-part ERI curriculum. Interventionists viewed publisher-developed video clips of lesson elements and participated in hands-on sessions with the curriculum materials. They were also shown how to set up and manage materials and student groups. Finally, time was devoted to addressing critical instructional techniques such as giving immediate corrective feedback and providing both group and individual turns to students.

The second professional development session was held after the interventionists had completed half of the ERI curriculum. This took place in late January. The professional development focused on the lessons and materials for Parts 3 and 4 of the program and followed a format similar to that of the first professional development day. Project ERI research team members provided both professional development sessions.

Fidelity of ERI Implementation

The fidelity-of-implementation measure used for the study focused on procedural fidelity, or adherence to the ERI program (Dane & Schneider, 1998; Gersten et al., 2005). Adherence to the ERI intervention was assessed by direct observations using a 4-point Likert scale (1 = poor, 2 = fair, 3 = good, 4 = excellent). For each of the seven activities within an ERI lesson, observers evaluated whether interventionists completed each component of the activity and whether they used specified wording prescribed in the activity. Each interventionist was observed three times by research staff over the course of the school year. Dosage of the ERI program was documented through interventionist logs that were collected and double checked weekly by research staff. Interrater reliability was calculated using percent agreement and averaged 86.9 percent ($SD = 10.7$).

The ERI interventionists were rated based on whether or not they completed all components of each activity with the scripted wording used. The average across all lessons was used to create an adherence score. The mean adherence score was 3.18 ($SD = 0.51$). Dosage was calculated as the total number of lessons received, which averaged 107 days ($SD = 9.02$).

SDI Intervention

Reading interventionists in the comparison condition (i.e., SDI) were asked to provide typical school-designed beginning reading intervention to identified students for 30 minutes daily, in groups of 5 students. Systematic RTI methods were in place in all participating schools at varying levels of implementation. These procedures focused on formative assessment, grade-level data meetings, and Tier 2 intervention support. Two of the six SDI interventionists implemented commercial intervention programs; *Reading Mastery Plus* (Engelmann & Hanner, 2001) and *Road to the Code* (Blachman, Ball, Black, & Tangel, 2000). The other four SDI interventionists reported that the implemented interventions were teacher constructed and integrated the district's core curriculum materials. The types and frequency of adjustments in the SDI condition were allowed to vary and were determined by each school. That is, interventionists adapted their instruction based on (a) results from informal reading assessments and school-administered progress monitoring through the use of FAIR (FLDOE, 2009) and DIBELS (Good & Kaminski, 2002); and (b) instructional feedback from classroom teachers during grade-level data meetings. Due to insufficient availability of professional personnel, no systematic efforts were involved in regrouping. Interventionists in the SDI condition did not receive any additional professional development from research staff; however, all interventionists in the district were experienced in providing supplemental beginning reading intervention and had previously received extensive professional development and resources in evidence-based reading instruction methods, both at the district and school site, by literacy coaches within the school district and state of Florida.

The SDI interventionists were formally observed and videotaped on three separate occasions to determine the instructional content of each activity within the lessons. Observations occurred at the beginning, middle, and end of the intervention period (Observations 1–3). Two trained observers independently coded the instructional content of each activity within the SDI lessons from the videotapes, focusing on early reading skills that fell under the following main categories: phonological awareness, alphabetics/phonics, writing, vocabulary, and comprehension. Observational data were coded at 1-minute intervals for each of the three observations for each interventionist. Interrater reliability between the observers was 84.7 percent. These procedures allowed us to determine the percentage of SDI lessons that included specific content in the areas of early reading intervention. Table 3 provides summary data of the percentage of SDI lessons that included specific content from each observation period. The mean number of days children participated in SDI intervention was 106.29 ($SD = 15.35$).

TABLE 3
Percentage of Lessons Observed That Included Specific Content

	SDI Year 04		
	y*Observation 1*	*Observation 2*	*Observation 3*
Phonological Awareness			
PA at Word/Syllable Level	16.6	0.0	0.0
Blending/ Segmenting	33.3	33.3	16.6
ID # of Sounds	16.6	16.6	0.0
Sound Isolation	33.3	33.3	16.6
Alphabetics/Phonics			
Letter Names	50.0	83.3	83.3
Letter Sounds	50.0	83.3	83.3
Reading Words	50.0	66.6	50.0
Sight Word Work	16.6	66.6	66.6
Reading Connected Text	50.0	33.3	66.6
Writing			
Writing Letter Names	33.3	20.0	33.3
Writing Letter Sounds	33.3	20.0	0.0
Writing Words	0.0	33.3	500
Vocabulary	33.3	33.3	33.3
Comprehension			
Listening	50.0	50.0	16.6
Reading	16.6	16.6	16.6
Read Aloud	16.6	33.3	16.6

Measures

Six assessment measures were administered throughout the study, and assessed a range of early literacy and beginning reading skills. Screening and pretest assessments were administered in September-October prior to the start of the intervention. Post-testing occurred within 2 weeks after the completion of the intervention in May-June. Pretest and posttest measures were administered in one day, distributed in one to two sessions. All assessments were administered by trained graduate assistants and research staff.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002)

DIBELS served as one of the screening, pretest, and posttest measures for the study. Three of the DIBELS subtests were administered: The Letter Naming Fluency (LNF) subtest, the Phoneme Segmentation Fluency (PSF) subtest, and the Nonsense Word Fluency (NWF) subtest. The LNF was one of the initial screening measures. This subtest measures a student's ability to rapidly name upper- and lowercase letters presented in 1 minute. Alternate-form reliability for the kindergarten sample is .89. The split-half reliability coefficient is .94 for the first-grade sample (coefficient for kindergarten sample not reported). The PSF subtest measures a student's ability to fluently segment three- and four-phoneme words. Scores indicate the number of sound segments correctly identified in 1 minute. Alternate-form reliability for the kindergarten sample is .88. The PSF measure was administered at both pre- and posttest. Finally, the NWF subtest measures a student's knowledge of letter-sound correspondence and the ability to fluently blend letters into words in which letters represent their most common sounds. Scores indicate the number of letter sounds produced correctly in 1 minute. Because the measure is fluency-based, students receive a higher score if they read the nonsense word as a whole word and a lower score if they give letter sounds in isolation. The alternate-form reliability for this subtest for the kindergarten sample is .88. The NWF measure was administered at both pre- and posttest.

Woodcock Reading Mastery Test-Revised/NU (WRMT-R; Woodcock, 1987, 1988)

The WRMT-R was used for pre- and post-testing. Four of the subtests were administered in the study: the Letter

Identification (LI) subtest, the Supplementary Letter Checklist, the Word Identification (WI) subtest, and the Word Attack (WA) subtest. The LI subtest was administered at pretest. It is a standardized, untimed test battery that assesses a student's ability to name upper- and lowercase letters presented in various fonts. The Supplementary Letter Checklist was administered at posttest to assess children's ability to correctly identify the letter name and produce the appropriate sound for each presented lowercase letter. The WI subtest was administered at both pre- and posttest. It measures a student's skill in reading a list of real words presented in isolation; it is untimed. The split-half reliability coefficient is .98 for the first-grade sample (coefficient for kindergarten sample not reported). Finally, the WA subtest is an untimed measure of a student's skill in reading a list of non-words (e.g., "tet") presented in isolation. The raw score is the number of non-words read correctly, which is converted into a standard score. The split-half reliability coefficient is .94 for the first-grade sample (coefficient for kindergarten sample is not reported).

Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999)

The CTOPP served as one of the screening, pre-, and posttest measures for the study. Three of the CTOPP subtests were administered: the Sound-Matching (SM) subtest, the Rapid Object Naming (RON) subtest, and the Blending Words (BW) subtest. The SM measures a child's ability to select one of three pictures that has the same initial or final sound as the first word. This is an untimed measure. The internal consistency coefficient (Cronbach's alpha) is .93 for both the age 5 and age 6 samples. The SM subtest was used as a screening measure and was also administered at posttest. The RON subtest measures the speed with which a student can name common objects. The student is presented 72 pictures, and the score is the total number of seconds it takes to complete the task. Alternate-form reliability is .82 for 5-year-olds and .81 for 6-year-olds. The RON measure was used as a screening measure. Finally, the BW subtest is an untimed measure that requires the student to combine sounds into real words. Sounds are presented on a compact disc. The internal consistency coefficient (Cronbach's alpha) is .88 for the age 5 sample and .89 for the age 6 sample. The BW measure was administered at both pre- and posttest.

Peabody Picture Vocabulary Test-III (PPVT-III; Dunn & Dunn, 1997)

Vocabulary was assessed at pretest only, using the PPVT-III, an individually administered oral test of receptive vocabulary. For each test item, the student is presented with four black-and-white illustrations and asked to select the picture considered to best illustrate the meaning of a word presented orally by the examiner. The internal-consistency reliability (Cronbach's alpha) for form IIIA is .95 for the 5- to 6-year-old group, and the test-retest reliability ranges from .92 to .93.

Test of Written Spelling-4 (TWS-4; Larsen, Hammill, & Moats, 2005)

Spelling was assessed at posttest only using the TWS-4. This measure uses a dictated word format and is untimed. The words to be spelled are drawn from 10 basal spelling programs and popular graded word lists. Scores indicate the number of words spelled correctly. Although norm-referenced scores begin with first grade, the TWS-4 was used to gather baseline information for kindergarten students. The alternate-form reliability for the age 6 sample is .86, and the internal-consistency reliability (Cronbach's alpha) for the age 6 sample for test form A is .87.

"Mac Gets Well" (Makar, 1995; Vadasy, Sanders, & Peyton, 2006)

Oral reading fluency was measured at posttest using a decodable passage at a beginning of first-grade reading level entitled "Mac Gets Well" (Makar, 1995; Vadasy et al., 2006). Scores on this measure indicate the number of words read correctly in 1 minute. Words were not counted if they were read incorrectly or skipped, or if the student took longer than 3 seconds to begin reading the word (the word was then supplied). Estimates of internal-consistency reliability for this passage in a previous study (Vadasy et al.) indicated a Cronbach's alpha of .93 for the kindergarten sample.

Data Analysis

Because of the non-independent observations due to the nesting data structure in our study (i.e., 90 students nested within 23 different intervention groups), multilevel modeling (Hox, 2010) was applied to analyze the data. All multilevel models were analyzed using the Mixed routine in SPSS (version 18; SPSS, Inc., 2009). Restricted maximum likelihood (REML) was used to estimate all the models. Intra-class correlations (ICCs) based on the unconditional model were calculated for the posttest measures (Raudenbush & Bryk, 2002). ICC, which measures the magnitude of dependency between observations, may also be viewed as the average correlation between any pair of students within an intervention group. Due to the disproportional sample size between the two intervention conditions which may violate the homogeneity assumption, we tested the heterogeneity of variances across all outcome measures before conducting the group comparison using multilevel modeling. Results showed none of the variances were heterogeneous, which allowed us to compare the two groups using MLM regardless of the different sample size. In addition to the statistical tests of significance (i.e., *t*-test) of the difference between different intervention conditions on the posttest measures, we calculated the effect size specifically for cluster-randomized designs (δ_T; Hedges, 2007). This type of effect size may be used to evaluate the practical significance of differences between the two intervention conditions. Moreover, we used the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995), as recommended by Valentine and Cooper (2003), to control for the potential inflated type I error rate due to the multiple comparisons. In accordance with this approach, we report as "substantively important" (Valentine & Cooper, p. 22) (a) all treatment effects that are statistically significant and (b) any treatment effects that are greater than .25 but not statistically significant.

RESULTS

Ninety kindergarten students (67 ERI-R and 23 SDI) were nested within 21 intervention groups. A comparison of the two groups of students on their demographic variables (i.e., gender, ethnicity, English language learner status, and bilingual status) showed no significant differences on any variable. Further, an examination of between-condition differences on interventionists' demographic variables (i.e., ethnicity, highest degree earned, and total years of teaching experience) found none that were statistically significant.

Pretests

Descriptive statistics of both pretest and posttest measures for the ERI-R condition and the SDI comparison group are presented in Table 4. Measures without corresponding pretest or posttest data were those considered too difficult for beginning of kindergarten or measures used for screening but not administered at posttest. Using *t*-tests to analyze group comparability at pretest, we found no statistically significant differences on any measure. The average performance of students in both groups was below the 25th percentile on the CTOPP sound-matching measure. On average, students could name 1.16 letters in 1 minute, and performed below the 25th percentile on the PPVT-III measure of receptive vocabulary.

Posttests

We examined the intervention effect by comparing the posttest mean differences on the outcome measures between conditions using the SPSS Mixed routine with the following set of models:

Level 1 (student-level) model.

$$\begin{aligned}\text{Posttest}_{ij} = {} & ß_{0j} + ß_{1j}\text{Pretest}_{ij} + ß_{2j}\text{PPVT}_{ij} + ß_{3j}\text{Age}_{ij} \\ & + ß_{4j}\text{Gender}_{ij} + ß_{5j}\text{Hispanic}_{ij} \\ & + ß_{6j}\text{African\_American}_{ij} + ß_{7j}\text{Other\_ethnicity}_{ij} \\ & + ß_{8j}\text{Special\_ed}_{ij} + ß_{9j}\text{Bilingual}_{ij} + e_{ij} \qquad (1)\end{aligned}$$

where i represents the i-th student ($i = 1 \ldots 90$) and j represents the j-th intervention group (i.e., $j = 1 \ldots 23$). Posttest_{ij} is the score of one of the posttest measures for the i-th student from the j-th group.

In the student-level model, covariates included the corresponding pretest score (Pretest_{ij}), PPVT score (PPVT_{ij}) as an indicator of the student's receptive language ability, plus demographic variables consisting of student's age (Age_{ij}), gender (Gender_{ij}), ethnicity (represented by three dummy-coded variables with the Caucasian students as the reference group: Hispanic_{ij}, $\text{African-American}_{ij}$, and $\text{Other\_ethnicity}_{ij}$), special education services (Special_ed_{ij}), and bilingual status (Bilingual_{ij}). The within-group random error is e_{ij}, and the corresponding variance, $V(e_{ij}) = \sigma^2$, captures the within-group variation. For posttests with no corresponding pretests, we used the untimed Letter Identification pretest score as the pretest covariate given that it has been used in previous studies (e.g., Coyne et al., in press) and was generally positively correlated with WRMT-R/NU Letter Name and Letter Sound Checklists, TWS-4, and "Mac Gets Well" oral reading fluency test.

The intervention group-level models were specified as shown below:

Level 2 (group-level) models.

$$\begin{aligned}ß_{0j} = {} & \gamma_{00} + \gamma_{01}\text{ERI\_}_j + \gamma_{02}\text{Years\_of\_experience}_j \\ & + \gamma_{03}\text{Ethnicity}_j + \gamma_{04}\text{Higher\_degree1}_j \\ & + \gamma_{05}\text{Higher\_degree2}_j + \gamma_{06}\text{Higher\_degree3}_j \\ & + \gamma_{07}\text{School}_{1j} + \gamma_{08}\text{School}_{2j} + \cdots \\ & + \gamma_{13}\text{School}_{7j} + U_{0j}, \qquad (2)\end{aligned}$$

and

$$\begin{aligned}ß_{1j} & = \gamma_{10}; ß_{2j} = \gamma_{20}; ß_{3j} = \gamma_{30}; ß_{4j} = \gamma_{40}; ß_{5j} \\ & = \gamma_{50}; ß_{6j} = \gamma_{60}; ß_{7j} = \gamma_{70}; ß_{8j} = \gamma_{80}; ß_{9j} \\ & = \gamma_{90}; ß_{10j} = \gamma_{100}; ß_{11j} = \gamma_{110}; ß_{12j} \\ & = \gamma_{120}; ß_{13j} = \gamma_{130}.\end{aligned}$$

In equation (2), ERI_{jk} is a dummy-coded variable with 1 as the ERI group and 0 as the SDI condition. The interventionist's years of teaching experience (i.e., $\text{Years\_of\_experience}_j$), ethnicity (i.e., Ethnicity_j which is a dummy-coded variable with Caucasian teachers as reference group), and highest degree earned (i.e., Higher_degree1_j, Higher_degree2_j, and Higher_degree3_j, which are dummy-coded variables with the Bachelor degree category as the reference group), were also controlled in the group level. Given the small number of schools (a total of eight schools), we treated school as a fixed factor and created seven dummy variables to control for the school effect. As pointed out in previous studies (e.g., Bloom, Richburg-Hayes, & Black, 2007; Cook, 2005), treating schools as a fixed factor by creating corresponding dummy variables (i.e., School_{1j}, $\text{School}_{2j} \ldots \text{School}_{7j}$) not only takes the school effect into account but also increases the statistical power.

The target effect, γ_{01}, represents the magnitude of intervention effect between the ERI groups and the SDI groups on the posttest measure after controlling for all the other variables, including the corresponding pretest covariate, demographic variables, and teacher and school effects. The between-group random effect is U_{0j}, and the corresponding variance, $V(U_{0j}) = \tau_{00}$, capture the between-group variation. Results are presented in Table 5.

The second column of Table 5 presents the corresponding ICC for each of the outcome measures based on the unconditional model (or random intercept model in which

TABLE 4
Kindergarten Pretest and Posttest Means and Standard Deviations

	ERI (N = 67)				SDI (N = 23)			
	Pretest		*Posttest*		*Pretest*		*Posttest*	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
PPVT-III	85.27	16.13	–	–	86.26	17.44	–	–
CTOPP Rapid Object Naming	5.29	2.08	–	–	5.09	1.97	–	–
WRMT-R/NU Letter ID	81.34	7.77	–	–	79.13	7.50	–	–
DIBELS Letter Naming Fluency	1.22	1.80	–	–	0.96	1.80	–	–
CTOPP Sound Matching	6.87	1.32	8.93	2.13	6.57	1.41	8.35	2.21
CTOPP Blending Words	7.72	1.77	11.33	2.18	7.52	1.78	10.74	2.78
WRMT-R/NU Word Attack	94.33	1.89	108.31	9.99	94.00	0.00	106.39	9.29
WRMT-R/NU Word ID	82.75	6.18	107.78	10.41	81.70	3.34	105.57	12.32
WRMT-R/NU Letter <u>Name</u> Checklist	–	–	25.88	3.99	–	–	25.87	1.77
WRMT-R/NU Letter <u>Sound</u> Checklist	–	–	26.54	5.22	–	–	25.17	6.31
DIBELS Nonsense Word Fluency	0.00	0.00	30.13	15.87	0.17	0.83	25.52	10.88
DIBELS Phonemic Segmentation Fluency	1.58	5.80	43.69	18.54	1.78	4.42	32.43	21.39
Test of Written Spelling-4	–	–	88.49	6.83	–	–	87.18	7.77
"Mac Gets Well"	–	–	12.28	8.81	–	–	9.55	6.39

Note: PPVT = Peabody Picture Vocabulary Test; CTOPP = Comprehensive Test of Phonological Processing; WRMT-R/NU = Woodcock Reading Mastery Test-Revised/Normative Update; DIBELS = Dynamic Indicators of Early Literacy Skills.

no predictors are included in the model). The ICCs ranged from .00 to .22, which are common ranges in educational research (Hox, 2010). Some of the measures showed zero ICC, meaning that there is no significant variation between mean performance of the instructional groups for many of the measures. The non-zero ICCs in several of the measures provided support for the non-independent/correlated observations and the need to use multilevel models to adequately handle the non-independency in our data.

After adjusting for the pretest covariate, demographic variables and both teacher and school effects, and controlling for the comparison-wise type I error rate using the Benjamini-Hochberg (1995) correction, we found no statistically significant difference between the ERI group and the SDI group on any posttest measures. To further evaluate the practical significance of the intervention effect between the ERI and the SDI conditions, we used one of the effect sizes (i.e., δ_T) specifically designed for cluster-randomized

TABLE 5
Tests of Statistical Significance and Effect Sizes Between the ERI and SDI Groups

Measure	*ICC*	γ_{01}	*p-value*	*Effect Size* (δ_T)	*Power*
Alphabet Knowledge					
WRMT-R/NU Letter <u>Name</u> Checklist	0.13	0.36	0.38	0.08	
Letter Sound Knowledge					
WRMT-R/NU Letter <u>Sound</u> Checklist	0.22	0.55	0.41	0.07	
Phonemic Awareness					
DIBELS Phonemic Segmentation	0.05	5.31	0.22	0.22	
Fluency					
CTOPP Sound Matching	0.00	1.28	0.06	**0.59**	0.481
CTOPP Blending Words	0.00	0.01	0.49	0.00	
Word Attack					
DIBELS Nonsense Word Fluency	0.00	5.22	0.17	**0.35**	0.307
WRMT-R/NU Word Attack	0.00	1.92	0.29	0.20	
Word Identification					
WRMT-R/NU Word ID	0.06	1.69	0.31	0.13	
Spelling					
TWS-4	0.00	3.54	0.08	**0.50**	0.427
Oral Reading Fluency	0.08	3.49	0.12	**0.34**	0.371

Note: ICC = intra-class correlation based on the unconditional model (i.e., without any predictors in the model). WRMT-R/NU = Woodcock Reading Mastery Test-Revised/Normative Update; DIBELS = Dynamic Indicators of Early Literacy Skills; CTOPP = Comprehensive Test of Phonological Processing.
γ_{01}, as shown in equation (3), is the difference between the ERI and SDI groups on the posttest measure while holding the effects of other variables as constant. Positive value indicated that the ERI group, on average, scored higher on the posttest measure than the SDI group.

studies (Hedges, 2007). It can be computed using the following equation:

$$\delta_T = \frac{\mu_\bullet^T - \mu_\bullet^C}{\sigma_T} = \frac{\gamma_{01}}{\sqrt{\sigma^2} + \sqrt{\tau_{00}},} \quad (3)$$

where γ_{01}, as presented in equation (2), is the target intervention effect (i.e., the mean difference between the ERI and the SDI conditions) after taking into account both teacher and school effects and adjusting for the demographical and pretest covariates. Accordingly, σ^2 is the within-group variance and τ_{00} is the between-group variance before adjusting for both teacher and school effects and the same set of covariates. As shown in Table 5, the standardized effect sizes comparing the performances of the two different conditions ranged from 0.00 (standard deviation) to 0.59 (standard deviation). Measures with differences considered substantively important (What Works Clearinghouse, 2008) included CTOPP Sound-Matching subtest ($\delta_T = .59$), TWS-4 ($\delta_T = .50$), DIBELS NWF ($\delta_T = .35$), and "Mac Gets Well" oral reading fluency test ($\delta_T = .34$), with findings favoring the ERI group.

To evaluate absolute performance, we used percentile rankings from the normative samples for the WI, WA, BW, PSF, and NWF posttest performance of students in the ERI and SDI conditions. We established "out of risk" as performing at or above the 30th percentile for each of the outcomes.

Figure 1 displays the boxplots for the two conditions by outcome measure and includes the reference line for the 30th percentile. Findings revealed that both intervention conditions were effective for the majority of students. Trends for all measures indicated that the ERI had a higher median percentile score on each of the outcome variables. Boxplots further revealed that the distribution of performance differed between groups. The impact of the ERI condition on the lowest performing students is demonstrated by examining children in the bottom quartile of each condition. The students in the bottom quartile of the ERI condition are above the normative samples' 50th percentile on the BW, WI, and WA measures, whereas the students in the lower quartile of the SDI condition do not score above the 50th percentile on any measure. Chi-square results indicated that the percentage of children above the 30th percentile on NWF and PSF was significantly greater for the ERI condition than the school-designed. On NWF, 76 percent of the students where above the 30th percentile in the ERI condition compared to 61 percent in the school-designed group, which was statistically significant, $\chi^2(1) = 5.21$, $p = .02$. The difference was greater for PSF, with 76 percent of children above the 30th percentile in the ERI condition and 57 percent in the school-designed group, $\chi^2(1) = 8.10$, $p = .004$.

DISCUSSION

The purpose of this study was to examine whether a Tier 2 intervention, ERI, that standardized instructional modifications based on student performance would produce statistically and practically significant effects on the reading achievement of kindergarten students identified with early reading risk over schools' typical Tier 2 interventions. In the ERI condition, we assessed student performance on curriculum-embedded assessments approximately every 4 weeks and used a standardized response protocol to systematically adjust groups and program progression based on performance on curriculum-embedded measures. To increase the internal validity of the study, students in both experimental and comparison conditions participated in supplemental, small-group intervention, controlled for dosage and size of groups. To address external validity, interventionists were school-based personnel who received modest amounts of professional development.

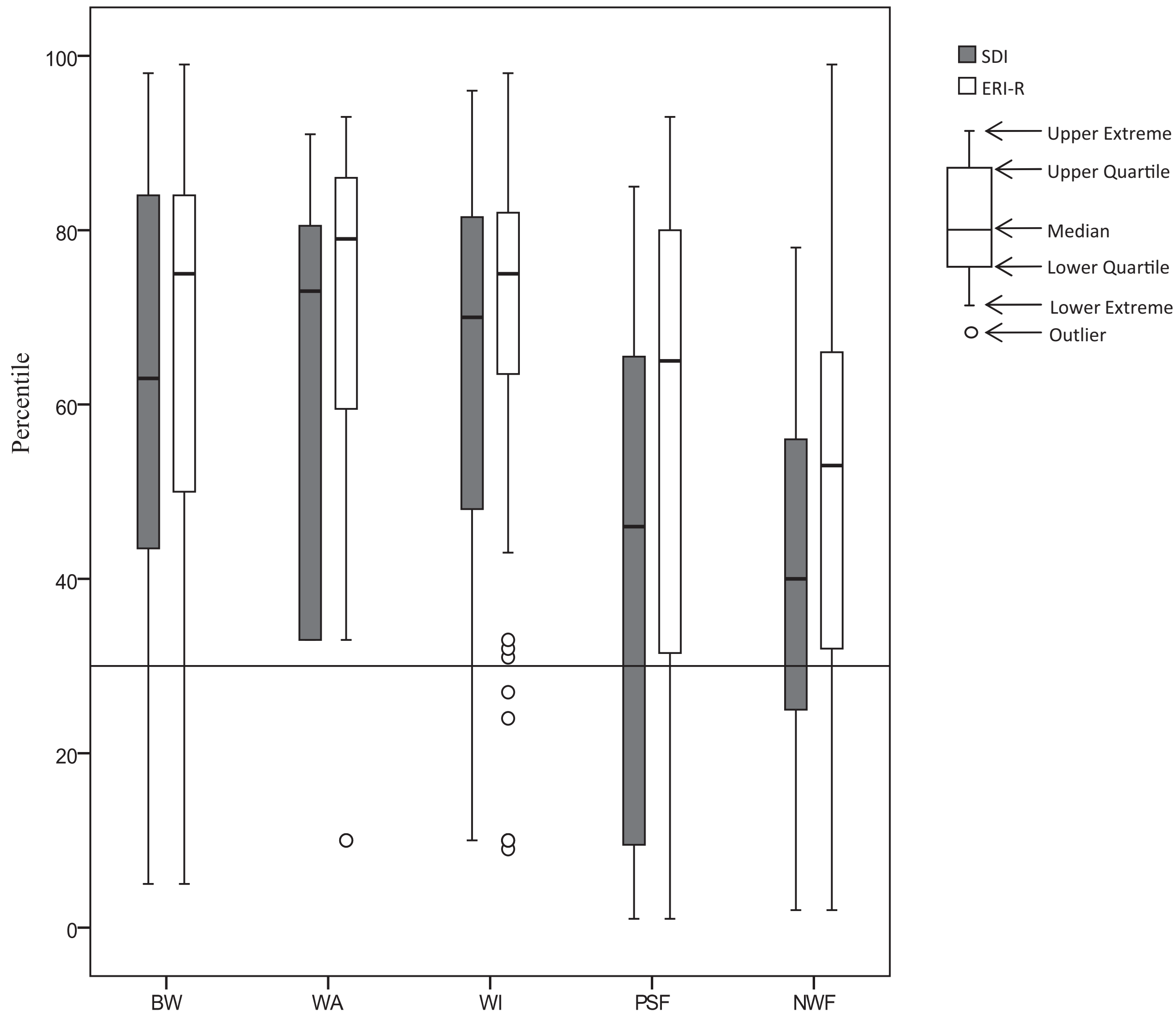
The first question addressed was whether the ERI condition would produce differential achievement benefits over school-designed Tier 2 interventions. Findings indicated no statistically significant differences between the two conditions. These findings contrast with those from a previous study in which intensifying ERI with standardized instructional modification produced statistically significant effects over a standard ERI implementation on measures including word identification, word attack, oral reading fluency, and spelling.

Regarding practical significance, findings from this study indicated the performance of students in the ERI condition differed substantively from students in the school-designed condition on measures of sound matching, nonsense word fluency, oral reading fluency, and written spelling with effects sizes ranging between .34 and .59. Compared to the previous ERI study, effect sizes were consistently lower in the present ERI to SDI comparison (e.g., word attack: .50 to .20; word identification: .69 to .13; oral reading fluency .66 to .34). While effect sizes suggest the potential of the ERI condition, further research is needed to corroborate this potential.

Our second question focused on the absolute effects of the interventions; that is, how did students perform compared to recognized benchmarks or achievement levels? Findings showed that the majority of students in both conditions performed above the 30th percentile on all performance measures. Trends for all measures indicated that students in the ERI condition had a higher median percentile score on each of the outcome variables, with statistically significant differences in the number of children above the 30th percentile on NWF and PSF.

Interpreting Statistical and Practical Effects

While the study was designed to address issues remaining to be answered regarding RTI, our findings raised more questions. How does one translate findings when between-group differences are non-significant but there are several substantively important effect sizes? Was the absence of statistical difference due to the usual suspect of lack of statistical power due to small sample size? Perhaps. Issues of statistical power often accompany studies that involve small samples of children who are at risk of reading difficulty. Because our sample was restricted to children who performed below the 15th percentile on screening measures, the sample size was modest and may have suppressed effects. Another common concern is implementation. ERI implementation fidelity was within the average range (3.1 on a 4-point scale), and while there was room for improvement, implementation was not demonstrably low.



BW = Comprehensive Test of Phonological Processing Blending Words; WA = WRMT-R/NU Word Attack; WI = WRMT/NU Word Identification; PSF = DIBELS Phonemic Segmentation Fluency; NSF = DIBELS Nonsense Word Fluency. Reference line reflects the 30th percentile for test-based norms.

FIGURE 1 Posttest performance boxplots by measure and condition.BW = Comprehensive Test of Phonological Processing Blending Words; WA = WRMT-R/NU Word Attack; WI = WRMT/NU Word Identification; PSF = DIBELS Phonemic Segmentation Fluency; NSF = DIBELS Nonsense Word Fluency. Reference line reflects the 30th percentile for test-based norms.

A factor that we consider important in interpreting effects is one we refer to as “intervention context.” In this study, the school-designed intervention was an established set of procedures and practices that, in large part, effectively addressed the needs of many children. Analysis of the content of instruction (see Table 3) indicated that school interventionists were targeting appropriate skills over time. Over a period of years, the district had established and implemented a comprehensive and coordinated early reading initiative. State, district, and school procedures included bi-weekly intervention meetings among professionals to review and adjust interventions based upon each student’s individual progress monitoring data within the school-designed intervention. While the new intervention yielded substantively important differences on some measures, the district’s existing RTI approach likewise yielded a significant return on its systematic investment in early intervention.

A final explanation for the absence of statistically significant differences may reside in the 2 × 4’s of instruction. Many of the instructional variables that may be adjusted to increase instructional intensity (Mellard et al., 2010) were not only similar, but were controlled in this research. That is, both conditions received explicit supplemental, small-group instruction for comparable amounts of time. Findings suggest that the nuances that distinguish interventions may not be as important as the instructional 2×4's that are common to effective early reading intervention, consisting of systematic, small-group instruction focused on targeted phonemic, alphabetic, and text reading skills for 30 minutes a day.

Limitations and Implications for Research and Practice

Theoretically, we believed that a standardized Tier 2 intervention that systematically used data to adjust instruction for students would enhance intervention effects of a less systematic approach. What we found in the school-designed intervention was a carefully planned, implemented, and supported system of reading intervention in response to both state and district policies and practices. However, due to school practices and personnel, actual implementation of interventions within the SDI condition was less systematic. We were interested in whether a standardized intervention that systematically used data to adjust instruction would be more beneficial than a Tier 2 intervention that varied during implementation by school personnel.

The study was conducted in a district with a history of commitment to responding to early reading achievement. Therefore, the findings must be interpreted in context, and effects sizes interpreted accordingly. As indicated previously, the sample size of the study may have impacted the findings, and further research is needed to clarify the statistical versus practical effect inconsistencies. Finally, and fundamental to RTI research, we had no measure of the Tier 1 or general kindergarten reading instruction. Understanding the alignment of Tier 1 and Tier 2 intervention may enable researchers and practitioners to develop more compatible and effective instruction.

The study was designed to contribute to the emerging body of literature on how to effectively use assessment and fidelity data to intensify intervention for students in early reading intervention (Denton et al., 2010; Gersten et al., 2009). As noted by Mellard et al. (2010), multiple methods are available to intensify intervention, and this study merely examined the tip of the RTI iceberg. At issue is whether certain modifications are more efficacious than others. Our approach was to manipulate a modest number of structural variables (i.e., regrouping and instructional progress based on learner performance), recognizing that each change exacts significant change in schools' resources. Because of resource limitations, we did not seek to add instructional time, individualize instruction, or schedule multiple intervention sessions. On an intensity continuum, our level of response was moderate and may not have been sufficient to address the needs of students who were least responsive. Studies examining methods to intensify intervention for students who do not respond to Tier 2 intervention are limited (Gersten et al., 2009) and the current findings reinforce the need for, and complexity of, research that unpacks the active ingredients in RTI.

Acknowledgments

This research was supported in part by the Grant Number R324E06067, National Center for Special Education Research, from the Institute of Education Sciences (IES), U.S. Department of Education. The content and positions in this article do not necessarily represent those of the funding agency. We thank our testers and data-entry staff for their contributions. Last, we thank the school districts, administrators, teachers, interventionists, and children for their support and participation in this research.

The intervention used in this study was *The Early Reading Intervention* (Pearson/Scott Foresman, 2004). Because two authors of this manuscript are also coauthors of the intervention program (i.e., Simmons and Coyne), the following steps were implemented to ensure objectivity of findings.

1. All data analyses were conducted by statisticians who had no financial interest with the Early Reading Intervention.
2. An external consultant with no financial affiliation with the Early Reading Intervention program independently reviewed the manuscript to ensure that (a) data analyses were appropriate, accurate, and objective, (b) reported findings and discussion were accurate, and (c) interpretations were consistent with data analysis.

REFERENCES

Bear, D., Dole, J., Fisher, D., Gibson, V., Echevarria, J., Hasbrouck, *et al*. (2009). *Macmillan treasures kindergarten*. New York, NY: Mc-Graw Hill.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*, 289–300.

Blachman, B. A., Ball, E. W., Black, R., & Tangel, D. (2000). *Road to the code: A phonological awareness program for young children*. Baltimore, MD: Brookes Publishing.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, *29*, 30–59. doi:10.3102/0162373707299550

Bradley, M. C., Daley, T., Levin, M., O'Reilly, F., Parsad, A., Robertson, A., & Werner, A. (2011). *IDEA national assessment implementation study: A final report* (NCEE 2011–4027). Washington, DC: Institute of Educational Sciences. Retrieved from http://ies.ed.gov/ncee/pubs/20114026/index.asp

Cook, T. D. (2005). Emergent principles for the design, implementation, and analysis of cluster-based experiments in social science. *The Annals of the American Academy of Political and Social Science*, *599*, 176–198. doi:10.1177/0002716205275738

Coyne, M. D., Little, M. E., Rawlinson, D. M., Simmons, D. C., Kwok, O., Kim, M., Simmons, L. E., Hagan-Burke, S., & Civetelli, C. (in press). Replicating the impact of a supplemental beginning reading intervention: The role of contextual differences. *Journal of Research on Educational Effectiveness*.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. doi:10.1016/S0272–7358(97)00043–3

Denton, C. A. (2001). *Responsive reading*. Unpublished curricular materials.

Denton, C. A., & Hocker, J. L. (2006). *Responsive reading instruction: Flexible intervention for struggling readers in the early grades*. Longmont, CO: Sopris West.

Denton, C. A., Kethley, C., Nimon, K., Kurz, T., Mathes, P. G., Shin, T. B., & Shih, M. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children*, *76*, 394–416.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test – Third Edition (PPVT-III)*. Bloomington, MN: Pearson Assessments.

Engelmann, S., & Hanner, S. (2001). *Reading mastery plus*. Columbus, OH: SRA/McGraw-Hill.

Florida Department of Education. (2009). *Florida assessments for instruction in reading (FAIR)*. Retrieved from http://www.fldoe.org/schools/foil/FAIR.pdf.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice*, *18*, 157–171. doi:10.1111/1540–5826.00072

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention in the primary grades. A practice guide* (NCEE 2009–4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*, 149–164. Retrieved from http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=15457832&site=ehost-live

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*, 341–370. doi:10.3102/1076998606298043

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Mahwah, NJ: LEA.

Larsen, S. C., Hammill, D. D., & Moats, L. C. (2005). *Test of written spelling* (4th ed.). Austin, TX: Pro-Ed.

Lembke, E. S., McMaster, K. L., & Stecker, P. M. (2010). The prevention science of reading research within a response-to-intervention model. *Psychology in the Schools*, *47*, 22–35. doi:10.1002/pits.20449

Little, M. E. (2009). *Response to intervention for teachers: Classroom instructional problem solving*. Denver, CO: Love Publishing Company.

Makar, B. W. (1995). *Primary phonics*. Cambridge, MA: Educators Publishing Service.

Marston, D., Pickart, M., Reschly, A., Heistad, D., Muyskens, P., & Tindal, G. (2007). Early literacy measures for improving student reading achievement: Translating research into practice. *Exceptionality*, *15*, 97–117. doi:10.1080/09362830701294177

Mellard, D., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research and Practice*, *25*, 217–225. doi:10.1111/j.1540–5826.2010.00319.x

Pearson/Scott Foresman. (2004). *Scott Foresman sidewalks: Early reading intervention*. Glenview, IL: Author.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265–282. doi:10.1037/0022–0663.96.2.265

SPSS Inc. (1999). *SPSS base 10.0 for Windows—User's guide*. Chicago, IL: Author.

Tilly, W. D. (2006). Diagnosing the learning enabled: The promise of response to intervention. *Perspectives*, *32*(1), 20–24.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, *98*, 508–528. doi:10.1037/0022–0663.98.3.508

Valentine, J. C., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse. Retrieved from http://ies.ed.gov/ncee/wwc/references/iDocViewer/Doc.aspx?docId=

VanDerHeyden, A. M. (2011). Technical adequacy of response to intervention decisions. *Exceptional Children*, *77*, 335–350.

VanDerHeyden, A. M., & Snyder, P. (2006). Integrating frameworks from early childhood intervention and school psychology to accelerate growth for all young children. *School Psychology Review*, *35*, 519–534.

Vaughn, S., & Chard, D. (2006). Three-tier intervention research studies: Description of two related projects. *Perspectives*, *32*(1), 30–34.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, *88*, 601–638. doi:10.1037/0022–0663.88.4.601

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities*, *39*(2), 157–169. doi:10.1177/00222194060390020401

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology*, *36*, 367–397. doi:10.1016/S0022–4405(98)00019–3

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed.

What Works Clearinghouse. (2008). *Procedures and standards handbook, version 2.0*. [Electronic Version]. Retreived February 15, 2009, from http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf.

Woodcock, R. W. (1987, 1998). *Woodcock reading mastery tests – Revised/Normative Update*. Bloomington, MN: Pearson Assessments.

About the Authors

Mary E. Little, Ph.D. is a professor and Graduate Program Coordinator in Special Education, Department of Child, Family, and Community Sciences at the University of Central Florida. Her research and scholarship focus on teacher efficacy, research-based instruction and interventions, and systems change that span preschool through high school.

D'Ann Rawlinson, M.Ed. is the Associate Director of the UCF Morgridge International Reading Center. She has over 17 years of experience in the field of literacy education. She directed federal and state research and development projects focusing on literacy intervention and low-progress readers.

Deborah C. Simmons, Ph.D., is a professor of special education in the Department of Educational Psychology at Texas A&M University. Her research and scholarship focus on reading interventions that span preschool through high school.

Minjung Kim is a doctoral candidate of the Educational Psychology department (Research, Measurement, and Statistics program) at Texas A&M University. Her research interests are longitudinal data analyses incorporating multilevel models and structural equation models. Applications of these statistical models to educational and psychological research are also her interest.

Oi-man Kwok, Ph.D., is an associate professor of Educational Psychology at Texas A&M University. His current interests include longitudinal data analysis using both Structural Equation Modeling (SEM) and Multi-Level Modeling (MLM) techniques.

Shanna Hagan-Burke, Ph.D., is an associate professor of special education in the Department of Educational Psychology at Texas A&M University. Her research interests include early literacy and factors that influence the efficacy of academic and behavioral interventions for at-risk learners.

Leslie E. Simmons is a research assistant serving as the coordinator for Project ERI in the Department of Educational Psychology at Texas A& M University. Her research interests focus on literacy assessment and reading intervention development.

Melissa Fogarty is a Ph.D. candidate in Educational Psychology at Texas A&M University. She is interested in issues concerning fidelity such as measuring fidelity, analyzing the relationship between fidelity and reading outcomes, and determining factors that influence program implementation.

Eric Oslund is a doctoral candidate in the Special Education program in Educational Psychology at Texas A&M University. His current interests include statistical analyses and methodologies associated with intervention and assessment of reading growth from kindergarten through high school.

Michael D. Coyne, Ph.D., is an associate professor in the Neag School of Education at the University of Connecticut. His research focuses on early reading and vocabulary instruction and intervention for students at risk of experiencing learning difficulties.