Exceptional Children

Vol. 80, No. 1, pp. 25-44.

Adjusting Beginning Reading Intervention Based on Student Performance: An Experimental Evaluation

MICHAEL D. COYNE
University of Connecticut

DEBORAH C. SIMMONS
SHANNA HAGAN-BURKE
LESLIE E. SIMMONS
OI-MAN KWOK
MINJUNG KIM
MELISSA FOGARTY
ERIC L. OSLUND
AARON B. TAYLOR
Texas A&M University

ASHLEY CAPOZZOLI-OLDHAM
SHARON WARE
University of Connecticut

MARY E. LITTLE
D'ANN M. RAWLINSON
University of Central Florida

ABSTRACT: *This experimental study evaluated a model in which the delivery of a supplemental beginning reading intervention was adjusted based on student performance. Kindergarten students identified as at risk for reading difficulties were assigned to one of two versions of the Early Reading Intervention (ERI; Pearson/Scott Foresman, 2004). Students assigned to the experimental condition received the intervention with systematic adjustments based on student performance. Students in the comparison condition received the same intervention without instructional modifications. The experimental group outperformed the comparison group on all posttest measures at the end of kindergarten. Follow-up analyses at the end of first grade revealed a continued advantage for the experimental group. Findings suggest that systematically adjusting intervention support in response to student performance may be feasible and efficacious.*

Schools are increasingly implementing response to intervention (RTI) or tiered systems of support in an attempt to more effectively and efficiently meet the needs of all students, especially those who are at risk for early reading difficulties (Samuels, 2011). Typical features of RTI include (a) providing comprehensive high-quality classroom instruction to students, (b) administering universal screening measures to identify students who are at risk for future reading difficulties, (c) implementing targeted supplemental intervention for students who need additional support, (d) monitoring student progress, and (e) intensifying the level of support for students who do not respond to supplemental intervention based on progress-monitoring data (Bradley, Danielson, & Doolittle, 2005; National Center on Response to Intervention, 2010).

However, the amount and quality of research evidence supporting the various components of RTI is mixed (Lembke, McMaster, & Stecker, 2010). For example, in an Institute of Education Sciences (IES) practice guide on RTI for reading in the primary grades, Gersten et al. (2009) reported moderate evidence supporting universal screening and strong evidence supporting the provision of supplemental, or Tier 2, intervention for students experiencing reading difficulties. Conversely, Gersten et al. reported that there is little current evidence to guide schools in how to best intensify supports for students who show minimal response to Tier 2 intervention.

The purpose of this study was to experimentally evaluate a model for adjusting the intensity of a Tier 2 kindergarten reading intervention based on student performance. Specifically, the goal was to enhance the efficacy of Tier 2 intervention through timely and feasible instructional adjustments. To that end, we compared the effects of the Early Reading Intervention (ERI; Pearson/Scott Foresman, 2004), a supplemental beginning reading intervention, implemented conventionally (i.e., lessons delivered sequentially without modifications [ERI comparison condition]) with the same intervention systematically adjusted over the year based on student performance (ERI experimental condition). We examined student outcomes at the end of kindergarten and again a year later at the end of first grade.

RTI AND ADJUSTING INTERVENTION

A central feature of RTI is adjusting the level of intervention support based on students' response to instruction. In the standard approach to RTI that has been widely promoted and adopted (Fuchs & Fuchs, 2009; National Center on Response to Intervention, 2010), schools typically make instructional adjustments for students receiving intervention every 8 to 12 weeks, and these adjustments are informed by student achievement assessed through curriculum-based measurement (CBM).

Specifically, in this standard approach to RTI, students who are identified as needing additional support receive a structured intervention implemented for a predetermined amount of time (e.g., 8–12 weeks). The intervention is implemented consistently and with fidelity. During the intervention, student progress is monitored using alternate forms of CBM probes, which assess student growth on a curriculum independent general outcome (e.g., oral reading fluency). At the end of approximately eight to ten weeks, decisions related to increasing or decreasing the level of instructional intensity are made based on well-defined decision rules related to the level and/or slope of students' CBM data. For example, a group of students who have not reached a specified benchmark may be provided with another 8 to 12 weeks of a more intensive intervention.

Strengths of this approach to RTI include efficient use of resources by identifying and implementing a small number of evidence-based interventions that are carefully selected to best meet the collective needs of the largest number of students. This model also emphasizes high levels of implementation fidelity, enabling consistent instructional quality and dosage for all students receiving intervention. A limitation, however, is the lack of opportunities for teachers to customize interventions for students and make ongoing instructional adjustments and refinements based on how students respond to a particular intervention (e.g., students' mastery of content introduced in the intervention).

In this study, we set out to create a model for adjusting intervention support that would maintain the emphasis on standardized implementation found in the typical approach to RTI and also enable interventionists to make timely and individualized instructional decisions. Our goal was to develop a framework that would allow ongoing modifications to instruction based on an analysis of student mastery of intervention-specific content measured with curriculum-embedded assessments. However, we also wanted to provide teachers with a clear structure for how and when to make these adjustments.

A large body of research has evaluated the effects of supplemental beginning interventions (e.g., Wanzek & Vaughn, 2007), including interventions implemented with kindergarten students (e.g., Gunn, Biglan, Smolkowski, & Ary, 2000; Torgesen et al., 1999; Vadasy, Sanders, & Peyton, 2006; Vellutino, Scanlon, Small, & Fanuele, 2006). However, few of these studies examined the benefit of adjusting the level of instructional support based on student response to instruction. Instead, most studies have evaluated interventions implemented with standard and consistent procedures for all students.

Our goal was to develop a framework that would allow ongoing modifications to instruction based on an analysis of student mastery of intervention-specific content measured with curriculum-embedded assessments.

In one large experimental study that evaluated an intervention designed to differentiate instruction based on student performance, Denton et al. (2010) investigated the effects of a small-group supplemental intervention, Responsive Reading Instruction (RRI; Denton & Hocker, 2006). Participants were first-grade students at risk for reading difficulties. Interventionists assessed each student weekly to determine curriculum mastery and identify instructional needs. Based on these performance data, interventionists then selected instructional activities for students from a menu of options. Although interventionists made decisions about which activities to implement, the RRI instructor's manual included specific guidance for selecting activities from the menu and detailed protocols to guide implementation. Compared to students in the control group, 43% of whom received an alternative school-designed comparison intervention, students in the RRI treatment had statistically significant higher outcomes, with moderate effect sizes on a range of reading measures, including word reading, spelling, reading fluency, and comprehension. Denton et al. concluded that a beginning reading intervention designed to be responsive to student needs was associated with positive student outcomes.

A FRAMEWORK FOR ADJUSTING INTERVENTION

In the IES RTI practice guide, Gersten et al. (2009) offered six methods to intensify intervention for students who do not respond to Tier 2 supports: providing concentrated instruction focused on a select number of target skills, teaching to mastery, adjusting the pace of instruction, scheduling multiple and extended daily intervention sessions, providing plentiful opportunities to respond, and delivering intervention one-on-one. Similarly, Mellard, McKnight, and Jordan (2010) outlined a number of dimensions, including curricular, instructional, dosage, grouping, and pacing variables, that can be adjusted to increase or decrease the intensity of supports within an RTI framework.

When developing our framework for adjusting instruction based on student performance, we decided to focus on a small number of variables to simplify the decision-making process for interventionists. We also wanted to identify instructional adjustments that would be feasible for schools to carry out using existing time, resources, and expertise. We chose to experimentally manipulate three dimensions of the ERI intervention in the experimental condition: (a) the frequency with which we collected progress-monitoring data, (b) the opportunities to regroup students based on curriculum mastery, and (c) the progression in which students moved through the intervention program.

The first variable that we manipulated in the experimental condition was the frequency with which intervention-specific progress-monitoring data was collected and used to make instructional decisions. Four unit tests are included with the published ERI intervention materials that assess student mastery of content introduced during each of the four sections of the program. We also developed four additional midunit tests so that student response to the ERI intervention could be assessed approximately every three to four weeks in the ERI experimental condition. Members of the research team met with each interventionist individually to share student mastery data, which were used to make decisions about regrouping and progression adjustments. Students in the comparison condition were assessed with only the four unit tests embedded within the ERI program. These data were shared with interventionists but were not used to make any instructional adjustments.

In the ERI experimental condition, students had the opportunity to be regrouped every 3 to 4 weeks with regrouping decisions informed by student mastery data. The goal of regrouping was to ensure that interventionists worked with homogeneous groups of students performing at similar skill levels. For example, students were generally regrouped based on one of three profiles: (a) students showing strong response to the intervention (i.e., mastery of more than 90% of content), (b) students showing good response (i.e., mastery of between 70%–89% of content), and (c) students showing weaker response (i.e., mastery of less than 70% of content). In contrast, students in the ERI comparison condition stayed with their initial groups throughout the school year; although, in some cases, groups were adjusted slightly at midyear.

The final variable that we manipulated was the progression of students through the ERI program. Curriculum progression decisions were standardized and based on student mastery data. Groups in which students on average demonstrated mastery of at least 90% of the ERI content based on mid- and end-of-unit tests were accelerated through the program, generally skipping every third or fourth lesson. Groups that demonstrated mastery of 70% to 89% of the content progressed through the lessons at a typical pace; however, letter names or letter sounds that were not mastered were incorporated into daily review. Groups that showed weaker response to the intervention, based on less than 70% content mastery, repeated specified lessons in which nonmastered skills were introduced and received targeted letter name and sound review. Decisions related to progression through the program were dynamic in that they were targeted specifically for each group and revisited every 3 to 4 weeks when new mastery data were collected. (Additional details of the regrouping and curriculum progression procedures are described in the methods section.) Students in the ERI comparison condition received consecutive lessons over the course of the intervention (i.e., one lesson per day with no adjustments). Other than the three variables that we manipulated (i.e., progress-monitoring mastery data, regrouping, and program progression), the ERI experimental and comparison conditions were identical.

METHODS

Participants

Participating students (N = 103) were enrolled in 40 classrooms from nine schools across three sites: south-central Texas (TX; n = 2 schools, 16 classrooms, 44 students), eastern Connecticut (CT; n = 5 schools, 13 classrooms, 42 students), and central Florida (FL; n = 2 schools, 11 classrooms, 17 students). Institutional Review Board approval was received at all sites. All participating schools were eligible for Title I funding and had high percentages of children who qualified for free or reduced-price lunch service (range = 50%–81%). School enrollments ranged from 266 to 985 students. The core general education reading curricula used in the kindergarten classrooms included Open Court (SRA/McGraw-Hill), Phonemic Awareness in Young Children (Adams, Foorman, Lundberg, & Beeler, 1998), Phonics Lessons (Pinnell & Fountas, 2006), Treasures (Macmillan/McGraw-Hill), and Trophies (Houghton Mifflin Harcourt).

Students and Assignment to Conditions. Our goal was to identify kindergarten students who were at significant risk for early reading difficulties to participate in the study. First, teachers and principals reviewed existing school-collected data to nominate four to six children per kindergarten classroom for further screening. Next, the research team screened nominated students to identify those who were performing below the 30th percentile on the DIBELS Letter Naming Fluency (LNF; Good & Kaminski, 2002) and the Sound-Matching (SM) subtest from the Comprehensive Test of Phonological Processing (CTOPP; Wag-

TABLE 1

Student Demographics by Condition

Variable	*ERI Experimental* (n = *70)* n	(%)	*ERI Comparison* (n = *33)* n	(%)
Gender				
Male	37	52.9	23	69.7
Female	33	47.1	10	30.3
Ethnicity				
American Indian or Alaska Native	1	1.4	0	0.0
Black or African-American	15	21.4	7	21.2
Hispanic or Latino	37	52.9	19	57.6
White	17	24.3	7	21.2
Receive Special Education Services	8	11.4	4	12.1
Bilingual/English Language Learner	11	15.7	9	27.3
Age				
Mean		5.38		5.40
SD		0.33		0.41

Note. ERI = Early Reading Intervention.

ner, Torgesen, & Rashotte, 1999). Finally, we conducted a retrospective analysis of data from an earlier ERI intervention study (Simmons et al., 2011) to determine pretest cut scores that best identified students who remained at risk (i.e., below 30%) after receiving intervention over the course of kindergarten. Based on these cut scores, students in the current study had to score either ≤ the 9th percentile on the letter identification subtest of the Woodcock Reading Mastery Tests - Revised/Normative Update (WRMT-R/NU; Woodcock, 1987/1998) and/or ≤ the 16th percentile on the CTOPP Rapid Object Naming (RON) subtest. Teachers nominated a total of 212 students and 110 of these students were identified to participate in the study.

Students in the final pool were stratified within schools and randomly assigned to either the ERI experimental (n = 70) or ERI comparison (n = 33) conditions. Because the experimental condition involved strategic regrouping that required more intervention groups, approximately twice the number of children were randomly assigned to the ERI experimental condition than to the ERI comparison condition. Student demographics for each condition are described in Table 1. On average, 11% to 12% of the total participants received special education services. In both conditions, more than 50% of students were Hispanic/Latino. To determine bilingual/English language learner (ELL) status, we used teacher-reported information indicating whether children were learning to speak and read English as a second language. All ELL participants were Spanish speakers, and all were receiving reading instruction in English. Group equivalence was evaluated using independent-sample t-tests for continuous variables (e.g., age) and chi-square tests for categorical variables (e.g., gender, ethnicity). Analyses indicated no statistically significant pretest differences in student demographics across the two conditions.

Attrition Analysis. A total of 110 kindergarten students were selected to participate in the study. Of this group, 103 (88%) participated in both pretest and posttest assessments in kindergarten. Analyses indicated no significant relation between condition and attrition; 8% (n = 3) of the ERI comparison and 5% (n = 4) of the ERI experimental students did not complete the study. Eighty-seven (87%) of the students who completed posttests at the end of kindergarten participated in the follow-up assessments at the end of first grade. Statistical comparisons indicated no statistically significant differences on any demographic variables between groups at the follow-up assessment. Attrition between the posttest and follow-up assessments was similar in each condition with 15.15% in the comparison and 15.71% in the experimental conditions.

Interventionists and Assignment to Condition. Seventeen school-based personnel who typically provided supplemental reading instruction served as interventionists and included certified reading specialists (n = 13) and paraprofessionals (n = 4). Four of the interventionists (23.5%) held a high school diploma, three (17.6%) held a bachelor's degree, and 10 (58.8%) held a master's degree. Interventionists had a mean of 18.03 (SD = 7.45) years' teaching experience. Interventionists were randomly assigned to ERI experimental (20 groups) or comparison (10 groups) conditions. Interventionists in some schools taught multiple groups of the same condition, and five interventionists taught groups in both conditions.

Chi-square analyses were conducted on categorical data and independent-sample t-tests were carried out for continuous variables to assess interventionists' demographic equivalence between conditions. Results of these analyses showed no reliable differences between groups on gender, total years of teaching experience, highest degree earned, ethnicity, or primary language.

Intervention Procedures

Overview of the Early Reading Intervention. Early Reading Intervention (Pearson/Scott Foresman, 2004) is a supplemental small-group intervention program with evidence of efficacy (e.g., Simmons et al., 2011) that explicitly teaches phonologic, alphabetic, decoding, spelling, and sentence-reading skills for kindergarten students with early reading risk. The 126-lesson program is comprised of four units that systematically progress from early phonemic and alphabetic skills to more complex regular and irregular word reading, spelling, and sentence-reading skills. A typical 30-min lesson consists of seven activities, each designed to last 3 to 5 min. The first 15 min of the lesson focus on phonological awareness and alphabetic understanding; the second 15 min integrate writing and spelling with previously taught phonemic and alphabetic skills. Lessons include detailed scripting to ensure clear and consistent communication of information. Scheduled instruction, review, and feedback are explicitly incorporated into the program. Lessons that introduce new information include a specified number of instructional interactions in which the teacher first models the information to students. Students practice the new skill with the teacher and then apply it to discrimination or generalization tasks. The ERI program was implemented conventionally as written in the comparison condition and modified to be responsive to student progress in the experimental condition.

ERI Experimental Condition. In the ERI experimental condition, we adjusted the intervention based on students' ongoing responses. The instructional adjustments developed for this condition consisted of three features that were systematically manipulated between conditions: (a) use of in-program assessments at the middle and end of each of the four units of the program to measure curriculum mastery and identify specific skills for targeted instruction; (b) student regrouping based on curriculum mastery; and (c) instructional adjustments, which included curriculum progression (i.e., repetition or acceleration in the program) and targeted review of critical skills. ERI experimental students were initially grouped according to pretest assessment data and began at Lesson 1 of the program.

Approximately every four weeks, members of the research team administered ERI in-program assessments to evaluate students' mastery of skills taught within that 4-week period. Research team members then met with interventionists to make regrouping decisions and instructional adjustments based on the student performance data from the in-program assessments. First, intervention groups were reconstituted as needed to ensure that students with similar levels of curriculum mastery were grouped together. For example, if a school had three ERI experimental groups of four students each, research team members and interventionists would meet and regroup students so that each group included students with comparable letter-sound knowledge and phonemic awareness skills. In some cases, student numbers were adjusted to better group students homogeneously (e.g., group sizes of 3, 4, and 5). In general, groups were reconstituted primarily during the beginning of the intervention and remained stable after that, with only a few students changing groups during the last half of the intervention period.

Next, research team members and interventionists used in-program assessment data to make instructional adjustment decisions about whether groups would repeat lessons or accelerate and about what skills groups would review. Groups with students primarily scoring 90% or above on in-program assessments were considered as exhibiting a strong response and continued with the typical lesson progression. Groups of students with scores of 90% or higher on two consecutive in-program assessments were accelerated in the curriculum using guidelines provided in the ERI program manual. Acceleration schedules generally involved students completing two of every three lessons until they reached the final unit in which they were taught each lesson. Groups with students scoring 70% to 89% were considered to demonstrate moderate response, and continued with the typical lesson progression with additional targeted review of letter-sound incorporated into lessons. Finally, groups with students scoring 69% and below were considered as exhibiting a weak response. The instructional adjustments for these groups included reteaching targeted lessons (i.e., based on incorrect letter-sound knowledge) and adding a modified letter-sound review activity to lessons. Groups resumed typical lesson progression after reviewing the targeted lessons and until the next in-program assessment. Working together, research team members and interventionists made all instructional adjustments using a standardized management plan that included systematic decision rules for determining lesson repetition and acceleration and skill review, based on students' performance on curriculum-embedded measures.

In the ERI experimental condition, 56 students (80.0%) were in groups that repeated two or more lessons. For these students, the range of lessons repeated was 2 to 25 (M = 10.84, SD = 6.26). Additionally, 29 students (41.4%) were in groups that were accelerated for two or more lessons. For these students, the range of lessons accelerated was 4 to 27 (M = 17.17, SD = 7.07). Because students were regrouped during the course of the intervention, and because groups' performance differed across in-program assessments, it was possible for students to be in groups that both repeated some lessons and accelerated through other lessons.

ERI Comparison Condition. Students in the ERI comparison condition were initially grouped according to similar entry-level skills based on pretest assessment data and started the program at Lesson 1. Interventionists progressed through the program sequentially. Students generally stayed with their initial groups throughout the school year, although in a very few cases, groups were adjusted slightly at midyear. Approximately every eight weeks, members of the research team administered ERI end-of-unit assessments to evaluate students' mastery of taught skills. End-of-unit assessment results were given to interventionists; however, unlike in the experimental condition, research team members did not make instructional recommendations based on student performance.

Common Intervention Components. To increase comparability between conditions, a number of common instructional components were standardized across the ERI experimental and comparison conditions. Both conditions implemented the ERI program (previously described) as a supplemental reading intervention. Groups in both conditions were comprised of three to five students. Interventionists in both conditions were asked to meet with their groups for 30 min, 5 days per week over the course of the study for an equivalent number of total sessions.

ERI Training. The same professional development was provided to all interventionists in both conditions in two sessions at their respective sites using standardized procedures. The first session, which occurred in September prior to intervention, familiarized interventionists with lesson structure and the scope and sequence of Parts I and II of the four-unit ERI curriculum. Interventionists viewed publisher-developed video clips of lesson elements and engaged in small-group practice of lesson components to gain familiarity with the lesson structure and materials. The second day of professional development training was provided at implementation midpoint and took place in late January. This session focused on Parts III and IV of the program and followed a format similar to that of the first professional development day.

The goal of the professional development was to enable high-quality implementation of the ERI activities and lessons, which was a shared focus of both experimental and comparison conditions. The unique features of the ERI experimental condition were primarily structural (e.g., data sharing, regrouping, reteaching lessons, acceleration) and were guided by research personnel who collected progress-monitoring data; shared student performance with interventionists; identified lesson repetition and acceleration schedules; and, in consultation with the interventionists throughout the year, regrouped students with similar performance scores.

ERI Fidelity. Each interventionist in both conditions was observed three times by research staff over the course of the school year using the same observation procedures. Lessons were also videotaped and coded to evaluate observer reliability. Observers coded fidelity for each lesson activity using a 7-point scale (1 = low, 2–3 = medium-low, 4–5 = medium, 6–7 = nearly perfect). The fidelity measure captured the interventionists' adherence to the ERI program and rated the extent to which they implemented the program according to the teacher guide and managed materials efficiently. Interrater reliability was established using a second observer for 23% of the videotaped observations. Interrater reliability calculated using percent agreement was 94%. Dosage data and student attendance were also collected through online teacher logs.

Measures and Assessment Procedures

Screening and Pretests. All screening and pretest assessments were completed in September and October, approximately six weeks into the school year and prior to the start of intervention. Screening and pretest measures assessed phonological awareness, letter-name knowledge, rapid automatized naming, and receptive vocabulary.

Phonological awareness was assessed using the SM subtest of the CTOPP (Wagner et al., 1999). This untimed subtest requires a child to select a picture from an array of three that contains the same initial or final sound as the target item orally presented by the administrator. The internal consistency coefficient (Cronbach's alpha) for SM is .93 for children ages 5 and 6.

Letter name knowledge was assessed using two measures. The letter ID subtest of the WRMT-R/NU (Woodcock, 1987/1998) assesses the ability to name upper- and lowercase letters. Split-half reliability for the first-grade normative sample is .94. The LNF subtest of the DIBELS (Good & Kaminski, 2002) assesses fluency in naming upper- and lowercase letters in 1 min. Alternate-form reliability for the kindergarten sample is .88.

Rapid automatized naming ability was measured using the RON subtest of the CTOPP battery (Wagner et al., 1999). On this timed measure, children name five familiar objects presented multiple times in random order on a display page. The procedure is repeated with a second stimulus page; the total time required to name all objects on both pages represents the child's test score. Alternate-form reliability coefficients are between .79 and .82 for ages 5 to 7.

Finally, *receptive vocabulary* was assessed with the Peabody Picture Vocabulary Test-III (PPVT-III; Dunn & Dunn, 1997). This measure of receptive vocabulary asks the child to identify one of four pictures that represents the term presented orally by the examiner. The internal consistency reliability (Cronbach's alpha) for Form A is .95 for ages 5 to 6; the test-retest reliability is .92 to .93.

Curriculum-Embedded Progress Measures. ERI progress assessments are criterion-referenced measures that evaluate students' mastery of skills taught in each unit of the program (i.e., alphabetic, phonemic, word reading, spelling). Students in the ERI experimental condition were assessed at the midpoint and the end of each of the four units of the program. Students in the ERI comparison condition were assessed only at the end of each unit. Mid- and end-of-unit measures sampled taught skills. For example, the assessment at the end of Unit 1 measured letter names, letter sounds, phonemic awareness of first and last sounds in words presented orally, and ability to select a letter that corresponded to the first and last sound in a word.

Kindergarten Posttest Only. Posttest assessments were administered in April and May of kindergarten within 2 weeks of intervention completion as follows:

Letter knowledge was assessed with the Supplementary Letter Checklist of the WRMT-R/NU (Woodcock, 1987/1998). This checklist is a complete inventory of all lowercase letters. It was administered twice, once to assess children's ability to produce letter names and a second time to produce letter sounds. Scores reflect the total number of correct names or sounds.

Phonological awareness was assessed with the SM (previously described) and Blending Words (BW) subtests of the CTOPP (Wagner et al., 1999). The BW is an untimed measure that asks a child to blend audio-recorded isolated sounds into words. The internal consistency coefficient (Cronbach's alpha) is .88 for age 5 and .89 for age 6. In addition, the Phoneme Segmentation Fluency (PSF) DIBELS subtest (Good & Kaminski, 2002) was used to measure a child's ability to segment three- and four-phoneme words. Scores indicate the number of sound segments correctly produced in 1 min. Alternate-form reliability for the kindergarten is .88.

Kindergarten Posttest and First-Grade Follow-up

The following measures were administered at kindergarten posttest as well as April and May of first grade to examine the durability of effects:

Decoding was assessed through the Word Attack subtest of the WRMT-R/NU (Woodcock, 1987/1998), an untimed measure of pseudoword reading. The split-half reliability coefficient is .94 for first grade. The Nonsense Word Fluency (NWF) subtest of the DIBELS (Good & Kaminski, 2002) was also administered to measure letter-sound correspondence and blending fluency. The score for this measure is the number of letter sounds produced correct, at either the sound or word level, in 1 min. The alternate-form reliability is .88 for kindergarten.

Word reading was assessed with the Word Identification subtest of the WRMT-R/NU (Woodcock, 1987/1998), an untimed measure of words of increasing difficulty. The split-half reliability coefficient is .98 for the first-grade sample.

Oral reading fluency was assessed with a decodable passage leveled for use in early first grade. The passage entitled "Mac Gets Well" (Makar, 1995) was presented to children, and the number of words a child read correctly in 1 min was used as a raw score. A prior study (Vadasy et al., 2006) used the same passage to measure kindergarten oral reading fluency and estimated internal-consistency reliability with a Cronbach's alpha of .93.

Spelling was measured using the Test of Written Spelling-4 (TWS-4; Larsen, Hammill, & Moats, 2005). This untimed measure assesses student ability to spell a list of dictated words. Total raw score is the number of correctly spelled words. The TWS-4 was administered to establish baseline information for kindergarten students; however, norm-referenced scores for the measure begin in first grade. The alternate-form reliability is .86 for the age 6. The internal-consistency reliability (Cronbach's alpha) is .87 for age 6 on test Form A.

Reading comprehension was assessed at the first-grade follow-up only using the Passage Comprehension subtest of the WRMT-R/NU (Woodcock, 1987/1998), which assesses students' reading comprehension of sentences and short passages. In this cloze task, students provide a key word that is missing in the text. The split-half reliability coefficient is .94 for first grade.

Procedures

All assessments were administered by trained data collectors who were blind to study condition. Data collectors included graduate students trained by members of the research team in two 4-hr sessions consisting of a review of general assessment procedures, modeling of the specific test protocols, paired practice, and supervised independent practice. Each data collector met the criterion of 90% accuracy in assessment administration and recording. After data collection, two trained individuals scored each testing protocol independently.

DATA ANALYSES

Due to the multilevel structure in our data (i.e., 103 students nested within 30 different intervention groups further nested within nine schools), we used a multilevel modeling approach (Hox, 2002). Specifically, we used a cross-classified random effects model (CCREM; Beretvas, 2008; Luo & Kwok, 2009; Meyers & Beretvas, 2006) to analyze the data given that students in the ERI-Experimental condition were regrouped based on their mastery of the curriculum during the intervention. All models were analyzed using HLM (V6.08; Raudenbush, Bryk, Cheong, & Congdon, 2004). Full-information maximum likelihood (FIML) was used to estimate all the models, and two-tailed tests were used to evaluate intervention effects.

Intraclass correlations (ICCs) based on the unconditional model were calculated for the posttest measures (Raudenbush & Bryk, 2002). Intraclass correlations, which measure the magnitude of dependency between observations, may also be viewed as the average correlation between any pair of students within a group. In addition to the tests of significance (i.e., *t*-test) of the difference between the ERI experimental and comparison conditions on the posttest and follow-up measures, we calculated effect sizes based on an approach developed by Hedges for cluster-randomized designs (δ_T; Hedges, 2007), which can be used to interpret practical differences between the two conditions. We used the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995), as recommended by the What Works Clearinghouse (WWC; 2008), to control for the potential inflated type I error rate due to the multiple comparisons. In accordance with the approach recommended by the WWC, we report comparisons that are statistically significant and also interpret treatment effects that are greater than .25 but not statistically significant as "substantively important" (p. 22).

RESULTS

Pretests

Descriptive statistics for pretest, posttest, and first-grade follow-up measures for the ERI experimental and comparison conditions are presented in Table 2. Using multilevel analyses to test group comparability at pretest, we found no statistically significant differences between conditions on any measure. The average performance of students in both groups was below the 25th percentile on the CTOPP sound-matching measure (M = 22nd percentile; 9.94 *SD*) and DIBELS letter-naming fluency (M = 13th percentile; 10.23 *SD*) as well as the 23rd percentile on the PPVT-III receptive vocabulary. On average, students named only 1.3 letters in 1 min.

Posttests

Treatment Effects. We examined posttest and follow-up mean differences between conditions (i.e., intervention effect) using SPSS Mixed with the following set of models:

Level 1 (Student-Level) Model

$$\begin{aligned}Posttest_{ijk} = \beta_{0jk} + \beta_{1jk}Pretest_{ijk} + \beta_{2jk}PPVT_{ijk} + \\ \beta_{3jk}Age_{ijk} + \beta_{4jk}Gender_{ijk} + \beta_{5jk}Hispanic_{ijk} + \\ \beta_{6jk}African\_American_{ijk} + \beta_{7jk}Special\_ed_{ijk} + \\ \beta_{8jk}Bilingual_{ijk} + \beta_{9jk}Last\ Lesson\ Number_{ijk} + \\ \beta_{10jk}Number\ of\ Lessons_{ijk} + \\ \beta_{11jk}Adherence\ to\ Program_{ijk} + r_{ijk}\end{aligned} \qquad (1)$$

where i represents the i-th student and j represents the j-th group at the beginning of the study and k represents the k-th group at the end of the study. Posttest_{ijk} is the score of one of the posttest measures.

In this student-level model, covariates included the corresponding pretest score (Pretest_{ijk}), PPVT score (PPVT_{ijk}) as an indicator of the student's receptive language ability, plus demographic variables consisting of student age (Age_{ijk}), gender (Gender_{ijk}), ethnicity (represented by two dummy-coded variables: Hispanic_{ijk} and $\text{African-American}_{ijk}$), special education services (Special_ed_{ijk}), bilingual status (Bilingual_{ijk}), two dosage variables, the number of lessons students received ($\text{Number of Lessons}_{ijk}$) the last lesson number ($\text{Last Lesson Number}_{ijk}$), and a fidelity of implementation measure ($\text{Adherence to Program}_{ijk}$).

The within-group random error is e_{ijk}, and the corresponding variance, $V(e_{ijk}) = \sigma^2$, captures the within-group variation. For posttests for which there were no corresponding pretests, we examined correlations to identify the pretest most highly associated with posttest performance. The untimed letter ID pretest score was highly correlated with WRMT-R/NU Letter Name and Letter

TABLE 2

Pretest, Posttest, and Follow-Up Means and Standard Deviations by Condition

	ERI Experimental (Posttest n = 70) (Follow-up n = 59)						ERI Comparison (Posttest n = 33) (Follow-up n = 28)					
	Pretest		Posttest		Follow-up		Pretest		Posttest		Follow-up	
Measure	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
PPVT-III	82.37	19.46	–	–			78.35	18.38	–	–		
CTOPP Rapid Object Naming	6.88	2.02	–	–			6.88	1.98	–	–		
WRMT-R/NU Letter ID	79.97	8.57	–	–			77.03	7.27	–	–		
WRMT-R/NU Letter Name Checklist	–	–	24.90	4.35	–	–	–	–	22.03	6.12	–	–
WRMT-R/NU Letter Sound Checklist	–	–	24.81	6.05	–	–	–	–	21.36	7.69	–	–
DIBELS Letter Naming Fluency	12.66	9.94	–	–			13.84	10.96	–	–		
CTOPP Sound Matching	7.58	1.09	9.60	2.55	–	–	7.42	1.28	9.03	1.92	–	–
CTOPP Blending Words	7.71	1.65	10.76	2.84	–	–	7.70	1.67	9.19	3.08	–	–
DIBELS Phonemic Segmentation Fluency	–	–	48.43	32.31	–	–	–	--	34.00	31.95	–	–
DIBELS Nonsense Word Fluency	–	–	46.66	26.25	–	–	–	--	43.19	26.48	–	–
WRMT-R/NU Word Attack	94.00	0.00	106.83	11.74	107.69	14.95	93.61	2.26	102.67	11.57	101.93	14.77
WRMT-R/NU Word ID	83.35	11.87	105.06	14.11	107.81	16.12	82.03	4.77	95.73	13.48	99.50	15.29
Test of Written Spelling	–	–	86.14	8.20	90.78	12.78	–	–	84.00	5.58	84.86	9.17
Oral Reading Fluency	–	–	10.40	10.01	42.83	32.78	–	–	7.47	6.72	26.68	21.22
WRMT-R/NU Passage Comprehension	–	–	–	–	99.63	12.37	–	–	–	–	92.14	13.48

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

Sound Checklists, TWS-4, and oral reading fluency (ORF) and therefore was used as the pretest covariate for these measures at posttest and follow-up. Letter identification pretest scores were also used as a covariate for WRMT-R/NU passage comprehension at follow-up.

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

Level 2 (Group-Level) Models

$$\beta_{0jk} = \gamma_{00} + \gamma_{01}\, \textit{ERI-Experimental}_{jk} + \gamma_{02}\, \textit{Interventionist's Experience}_{jk} + \gamma_{03}\, \textit{School}_{1jk} + \gamma_{04}\, \textit{School}_{2jk} + \ldots + \gamma_{010}\, \textit{School}_{8jk} + U_{0jk} + W_{0jk}$$

and

$$\beta_{1jk} = \gamma_{10};\ \beta_{2jk} = \gamma_{20};\ \beta_{3jk} = \gamma_{30};\ \beta_{4jk} = \gamma_{40};\ \beta_{5jk} = \gamma_{50};\ \beta_{6jk} = \gamma_{60};\ \beta_{7jk} = \gamma_{70};\ \beta_{8jk} = \gamma_{80};\ \beta_{9jk} = \gamma_{90};\ \beta_{10jk} = \gamma_{100} \qquad (2)$$

In the first equation, ERI-Experimental$_{jk}$ is a dummy-coded variable with 1 as the experimental condition and 0 as the comparison condition. The interventionists' years of teaching experience was also controlled in the group level. Additionally, because of the small number of schools participating (a total of nine schools), we treated school as a fixed factor and created eight dummy variables to represent the school effect, rather than adding another level to the model with an additional random effect. Treating schools as fixed covariates (i.e., the eight dummy variables: School$_{1j}$, School$_{2j}$ . . . School$_{8j}$) not only takes the school effect into account, but also increases the statistical power (Bloom, Richburg-Hayes, & Black, 2007; Cook, 2005). There are two between-group random effects, namely U_{0jk} and W_{0kj}, and the corresponding variances are $V(U_{0jk}) = \tau_{00(\text{initial})}$ and $V(W_{0jk}) = \tau_{00(\text{last})}$, respectively. These two random effect variances capture the between-group variation at the beginning and the end of the study.

The target effect, γ_{01}, represents the magnitude of the intervention effect between ERI experimental and regular ERI intervention groups on the posttest measure after controlling for all other variables, including the corresponding pretest covariate, demographic variables, and school effect. Given that the intervention groups, particularly for those in the ERI-Experimental condition, were changed for regrouping throughout the study, we calculated two different ICCs; namely, initial-group ICC and last-group ICC to capture the level of dependency in the data at these time points.

Results are presented in Table 3. The second and third columns of Table 3 present the corresponding ICCs for each of the outcome measures based on the unconditional model (or random intercept model in which no predictors are included in the model). The ICCs by the initial groups ranged from .00 to .22 and the ICCs by the last groups ranged from .18 to .60. The nonzero ICCs provided support for the nonindependent/correlated observations and the need to use multilevel models to adequately handle the nonindependency in our data. Moreover, the change in the magnitude of the ICCs did reflect the change in the homogeneity of the groups, specifically for the groups in the ERI-Experimental condition in which students were regrouped based on their curriculum mastery, and we expected that students with very similar progress would be grouped together at the end of the study which could result in higher ICCs.

After adjusting for the pretest covariate, demographic variables, and the teacher and school effects, and controlling for the comparison-wise type I error rate using the Benjamini-Hochberg correction, it was found that the experimental group scored statistically significantly higher than the comparison group on the following posttest measures: WRMT-R/NU Letter Name and Letter Sound Checklists, Word ID, and the ORF test. We found substantively important effects (i.e., nonstatistically significant effects with effect sizes greater than .25) on all other posttest measures including DIBELS PSF and NWF, CTOPP Sound Matching and Blending Words, WRMT-R/NU Word Attack, and TWS-4. Secondary analyses indicated that there were no statistically significant differences on any student outcomes for groups taught by paraprofessionals compared to groups taught by certified teachers regardless of the intervention conditions. At end-of-first-grade follow-up, the experimental group scored statistically significantly higher than the comparison group on all measures including WRMT-R/NU Passage Comprehension, WRMT-R/NU Word Attack and Word ID, TWS-4, and ORF.

TABLE 3

Tests of Statistical Significance and Effect Sizes Between ERI Experimental and Comparison Groups

Measure	$ICC_{initial}$[a]	ICC_{last}[b]	γ_{01}[c]	t-*Value*	††*Two-tailed* p	*Effect Size* (δ_T)
Alphabet Knowledge						
WRMT-R/NU Letter Name Checklist	0.00	0.44	4.07	2.15	0.039††	0.57
Letter Sound Knowledge						
WRMT-R/NU Letter Sound Checklist	0.01	0.60	5.30	2.62	0.013††	0.54
Phonemic Awareness						
DIBELS Phonemic Segmentation Fluency	0.07	0.54	10.27	1.85	0.084	0.29
CTOPP Sound Matching	0.15	0.52	1.54	1.99	0.058	0.38
CTOPP Blending Words	0.15	0.44	1.40	2.33	0.028	0.28
Word Attack						
DIBELS Nonsense Word Fluency	0.07	0.27	8.68	2.02	0.052	0.37
WRMT-R/NU Word Attack	0.03	0.34	6.15	1.68	0.109	0.34
WRMT-R/NU Word Attack – Follow-up	0.20	0.45	9.95	2.57	0.015††	0.39
Word Identification						
WRMT-R/NU Word Identification (ID)	0.00	0.45	15.69	4.22	0.000††	0.76
WRMT-R/NU Word Identification (ID) –Follow-up	0.03	0.52	14.36	3.08	0.006††	0.58
Spelling						
TWS-4	0.22	0.27	3.73	1.93	0.065	0.29
TWS-4 – Follow-up	0.01	0.24	11.88	3.12	0.004††	0.69
Oral Reading Fluency	0.00	0.32	5.83	2.52	0.014††	0.46
Oral Reading Fluency – Follow-up	0.00	0.18	25.25	2.81	0.007††	0.61
Passage Comprehension						
WRMT-R/NU Passage Comprehension – Follow-up	0.10	0.31	13.50	4.54	0.000††	0.64

Note. ERI = Early Reading Intervention; CTOPP = Comprehensive Test of Phonological Processing; DIBELS = Dynamic Indicators of Early Literacy Skills; ICC = Intra-class correlation based on the unconditional model (i.e., without any predictors in the model); TWS-4 = Test of Written Spelling-4; WRMT-R/NU = Woodcock Reading Mastery Test-Revised/Normative Update.

[a]$ICC_{initial} = \tau_{00(initial)}/(\tau_{00(initial)}+\sigma^2)$ (i.e., $\tau_{00(initial)}$ = between-group variance at the initial time point).

[b]$ICC_{last} = \tau_{00(last)}/(\tau_{00(last)}+\sigma^2)$ (i.e., $\tau_{00(last)}$ = between-group variance at the last time point).

[c]γ_{01}, as shown in equation (2), is the difference between experimental and comparison groups on the posttest measure when holding the effects of other variables as constant. Positive value indicated that experimental group, on average, scored higher on the posttest measure than the comparison group.

††Significant effect after Benjamini-Hochberg correction.

Given that there is no standardized effect size measure available under the CCREM framework, we created an effect size measure based on Hedges' (2007) effect sizes (i.e., δ_T) specifically designed for cluster-randomized studies to evaluate the practical significance of the intervention effect between conditions. It is computed using the following equation:

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

where γ_{01} is the treatment effect (i.e., the mean difference between the experimental and comparison conditions) after taking into account teacher and school effects and adjusting for the demographic and pretest covariates. Accordingly, σ^2 is the within-group variance and $\tau_{00(\text{initial})}$ and $\tau_{00(\text{last})}$ are the between-group variances from the unconditional model. As shown in Table 3, standardized effect sizes favored the ERI experimental condition on all measures and ranged from .29 to .76 at posttest and from .39 to .69 at first-grade follow-up.

Treatment Response. All children who participated in the study were considered to be at risk at pretest (i.e., below the 30th percentile) on either or both the CTOPP SM and DIBELS LNF measures. To examine absolute response, we used posttest percentile rankings on normative and criterion-referenced measures, including DIBELS phonemic segmentation fluency (PSF) and nonsense word fluency (NWF), CTOPP sound-matching and blending words, and WRMT-R/NU Word Attack and Word ID. We established the "out-of-risk" criterion as at or above the 30th or closest corresponding percentile by these measures (Torgesen et al., 1999; Vadasy et al., 2006).

Table 4 presents the percentages of students at each of the percentile intervals (e.g., 15–29th) by condition for each measure. Findings indicated not only different levels of response between groups but also varied levels of response by measure used. Normative-referenced measures of untimed performance generally yielded a smaller percentage of nonresponders continuing to score below the 30th percentile (e.g., 5.7% of the ERI experimental group students and 9.1% of the ERI comparison group students on WRMT-R/NU Word Attack) than timed measures (35.8% of the ERI experimental and 36.3% of the ERI comparison group on DIBELS NWF). Patterns of response indicated strong response (i.e., performance at or above the 50th percentile) by many students in both conditions (e.g., 52.8% and 30.3%, respectively, for the ERI experimental and comparison groups on the PSF measure). However, substantial percentages of children failed to move out of significant risk (i.e., < 15th percentile).

To illustrate the distribution of responses in each condition, we constructed box plots for all measures with corresponding percentile rankings for kindergarten students (see Figure 1). These plots allow us to examine the range of performance within and between conditions and to better interpret level of response. The box plots revealed that a majority of students in both intervention conditions were performing above the 30% at posttest. Results of multilevel analyses indicated that, for most measures, the ERI experimental condition produced a stronger response than the ERI comparison condition, and the box plots revealed that the experimental condition was particularly beneficial for students at the lower end of the distribution. For example, the plot for word identification indicated a stronger effect of the experimental intervention overall, with students at the lower quartile of the distribution performing at approximately the 40th percentile compared to those at the lower quartile of the comparison condition, who performed at the 10th percentile.

Fidelity of Implementation. Across observations of interventionists implementing the ERI experimental condition, adherence to the program averaged 5.63 (*SD* = .26) on a 7-point scale (1 = low; 7 = nearly perfect). The average number of lessons received, or dosage, for the students in the experimental condition was 99 (*SD* = 13.68). For the comparison condition, adherence to program fidelity averaged 5.31 (*SD* = .65) on a 7-point scale (1 = low implementation; 7 = nearly perfect). The average number of lessons was 105 (*SD* = 11.55).

Social Validity. We assessed teacher perceptions of the ERI program via a postintervention survey completed by all 17 interventionists across both conditions. A 6-point Likert scale was used, with a score of 1 representing "strongly disagree" and a score of 6 representing "strongly agree." Overall, teachers were positive about the ERI program and reported that it improved their teaching abilities and the students' word reading. The average rating of responses to the statement "I have a positive reaction to the ERI program" was 5.41 (*SD* = .795). Teachers scored an average of 5.29 (*SD* = .772) in response to the statement "My ability to teach children with early reading difficulties improved through using the ERI program." Regarding student outcomes, teachers rated the statement "The ERI program con-

TABLE 4

Posttest Performance Percentile Distribution by Condition

	Percentile				
Measure	*<15th*	*15–29th*	*30–49th*	*50–74th*	*≥75th*
Phonemic Awareness					
DIBELS Phonemic Segmentation Fluency					
ERI Experimental (n = 70)	25.7%	11.4%	10.0%	21.4%	31.4%
ERI Comparison (n = 33)	42.4%	18.2%	9.1%	6.1%	24.2%
CTOPP Sound Matching					
ERI Experimental (n = 70)	11.4%	24.3%	14.3%	21.4%	28.6%
ERI Comparison (n = 30)	6.1%	30.3%	21.2%	24.2%	9.1%
CTOPP Blending Words					
ERI Experimental (n = 70)	11.6%	4.3%	12.9%	25.7%	45.7%
ERI Comparison (n = 31)	30.3%	9.1%	0.0%	27.3%	27.3%
Word Attack					
DIBELS Nonsense Word Fluency					
ERI Experimental (n = 70)	12.9%	22.9%	12.9%	32.9%	18.6%
ERI Comparison (n = 32)	24.2%	12.1%	15.2%	30.3%	15.2%
WRMT-R/NU Word Attack					
ERI Experimental (n = 70)	5.7%	0.0%	25.7%	15.7%	52.9%
ERI Comparison (n = 32)	9.1%	0.0%	36.4%	15.2%	39.4%
Word Identification					
WRMT-R/NU Word ID					
ERI Experimental (n = 69)	11.4%	10.0%	10.0%	30.0%	38.6%
ERI Comparison (n = 33)	36.4%	3.0%	12.1%	30.3%	18.2%

Note. DIBELS = Dynamic Indicators of Early Literacy Skills; ERI = Early Reading Intervention; CTOPP = Comprehensive Test of Phonological Processing; WRMT-R/NU = Woodcock Reading Mastery Test-Revised/Normative Update.

tributed to improving my students' word reading skills" with an average score of 5.46 (SD = .624).

DISCUSSION

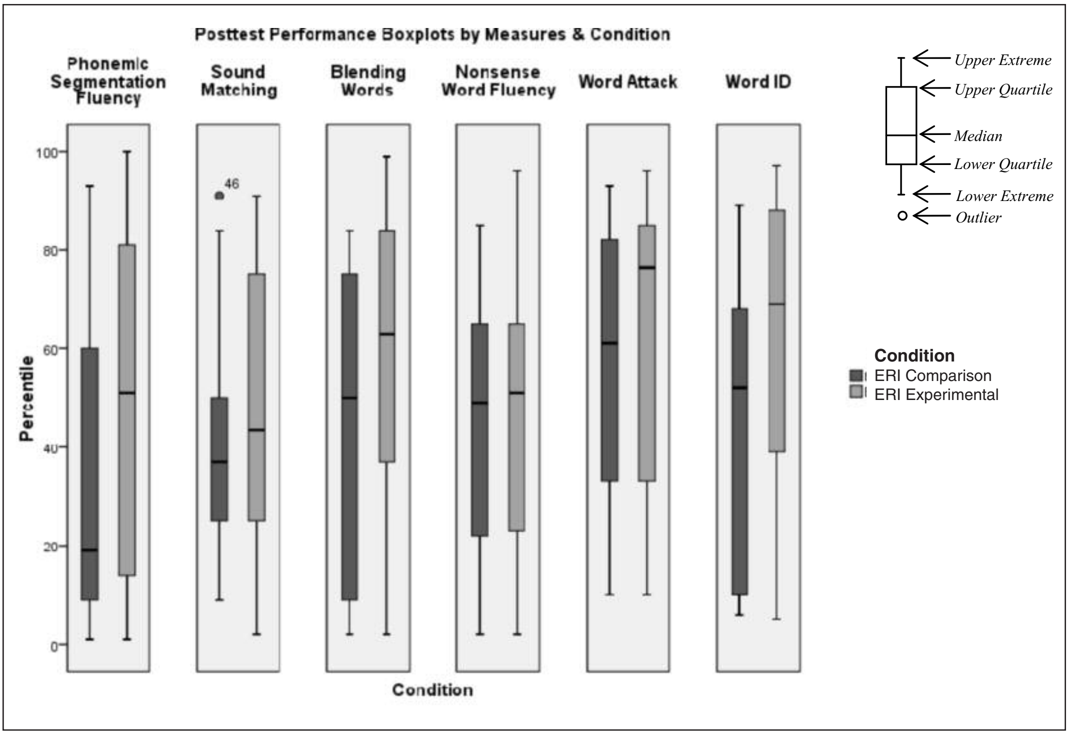
RTI or tiered systems of support have gained considerable traction in schools in recent years. However, many questions remain regarding the effectiveness and feasibility of adjusting intervention based on student performance. Although a number of instructional variables have been identified that can be adjusted to modify the intensity of supports within an RTI approach, few studies have systematically manipulated and examined effects of specific modifications.

In this study, we were interested in experimentally evaluating a framework for adjusting intervention support that focused on instructional variables that schools could reasonably manipulate using existing time, resources, and expertise. Our framework included systematic and standardized procedures for monitoring progress, regrouping students based on curriculum mastery, and adjusting the progression through the curriculum to ensure sufficient content exposure, review, and practice. Specifically, we compared approximately 100 lessons of a supplemental kindergarten beginning reading intervention without modification (ERI comparison condition) with the same intervention modified to enable ongoing instructional adjustments based on student performance and content mastery (ERI experimental condition).

Results of multilevel models indicated that kindergarten students identified as being at risk

FIGURE 1

Posttest Performance Boxplots by Measures and Condition



for reading difficulties who received the ERI intervention that was adjusted based on their performance outperformed those who received the unmodified ERI intervention on all posttest measures, with statistically significant differences on measures of letter names and sounds, word identification, and oral reading fluency. Substantively important effects favored the ERI experimental conditions on all other posttest outcomes including measures of phonemic awareness, word decoding, and spelling. Effect sizes across all posttest comparisons ranged from .29 to .76. Additionally, follow-up analyses revealed that students who received the ERI experimental condition in kindergarten continued to outperform comparison students at the end of first grade, with statistically significant findings on all measures of word reading, spelling, reading fluency, and reading comprehension and with effect sizes comparable to those at posttest.

These results are generally consistent with those of Denton et al. (2010), who observed favorable outcomes among children receiving their Responsive Reading Intervention compared to controls, many of whom were receiving an alternative intervention. However, we believe our findings provide an even stronger test of the effects of adjusting intervention because comparison students received the same base intervention program (i.e., same content, instructional design, delivery features, dosage), the only difference across conditions being the ongoing adjustments.

Students in both conditions made considerable growth from pretest to posttest. Whereas all students were performing below the 30th percentile on measures administered at pretest, a majority of students scored above the 30th percentile on each posttest measure, most of them performing above the 50th percentile. Although both intervention conditions accelerated students' learning, the ERI experimental condition seemed to particularly benefit students at the lower end of the distribution. For example, substantially fewer students continued to perform below the 15th percentile in the ERI experimental condition compared to the ERI comparison condition.

In summary, findings provide support for the efficacy of an essential component of RTI models—adjusting intervention support based on student response to instruction. In addition, findings reveal a clear and lasting advantage for students who received beginning reading intervention that was adjusted based on their ongoing response and curriculum mastery.

Findings provide support for the efficacy of an essential component of RTI models—adjusting intervention support based on student response to instruction.

A central feature of RTI, or tiered systems of support, is adjusting the intensity of intervention support based on student performance (National Center on Response to Intervention, 2010). Although teachers and schools are increasingly being advised to make these types of ongoing instructional decisions, there is little direct experimental evidence of the effects of adjusting instruction based on student response. In this study, we found that when a Tier 2 intervention was adjusted in ways often recommended by RTI guidance documents (e.g., teaching to mastery, adjusting the pace of instruction; Gersten et al., 2009), student learning was accelerated, particularly the learning of students most at risk. Therefore, we believe that our findings provide support for the efficacy of an essential component of RTI models—adjusting intervention support based on student response to instruction.

In this study, our goal was to develop a framework that would allow ongoing modifications to instruction based on an analysis of student performance but would also provide a clear structure for how and when to make these adjustments. First, we enabled interventionists to make timely adjustments to instruction based on individual student data of content mastery. We also, however, provided interventionists with a structured framework for making these adjustments based on clearly defined and standardized decision rules. We believe that this approach to adjusting intervention support presents a promising framework that can allow schools to both individualize instructional adjustments and also ensure consistency schoolwide by relying on carefully specified decision rules that can be anchored to a particular intervention program.

Finally, we developed our responsive package so that it was fairly modest in scope and relatively feasible for school personnel to carry out. First, we based instructional decisions on data collected from in-program unit assessments of curriculum mastery—a type of assessment that is often included with commercially available intervention materials. Next, regrouping took place within the existing intervention framework. For example, some schools had resources to run only two or three intervention groups, so students were regrouped as homogeneously as possible among those existing groups. Finally, we developed instructional adjustments to make use of common intervention materials. For example, although curriculum progression was differentiated for different intervention groups, all groups used the same instructional materials from the same intervention program (i.e., ERI). Even with these fairly modest and feasible adjustments, students demonstrated increased learning gains that were both statistically significant and educationally meaningful.

LIMITATIONS

We evaluated a framework for adjusting intervention that included using progress-monitoring data, regrouping students to increase instructional homogeneity, and differentiating students' progression through the curriculum. Although findings revealed that this framework produced favorable results on students' beginning reading outcomes, we were not able to isolate the active ingredient in this framework. Future research should evaluate the effects of other alterable variables that can be adjusted to increase or decrease the intensity of supports within an RTI framework, both in isolation and in various combinations. Additionally, we developed our procedures for adjusting intervention to be used with a specific intervention program (i.e., ERI). Although the types of adjustments we targeted could easily be modified for use with other supplemental beginning reading interventions, our findings cannot be generalized to other interventions. In addition, interventionists were supported by research staff in making instructional decisions. Thus, it is important to investigate whether similar types of adjustments are effective when used with other evidence-based supplemental intervention programs and whether interventionists can engage in this type of instructional decision making without external support. We also made adjustments based on student performance on in-program assessments included in the ERI intervention program. These assessments were developed to provide information about student mastery of content introduced in the ERI program and there is no information about the technical adequacy of these measures. Future research should continue to investigate the benefits and limitations of using curriculum-embedded assessments, which often lack published reliability and validity data, to inform instructional decision making, particularly compared to curriculum-independent assessments (e.g., CBM).

REFERENCES

Adams, M. J., Foorman, B. R., Lundberg, I., & Beeler, T. (1998). *Phonemic awareness in young children.* Baltimore, MD: Brookes.

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 (Methodological)*, *57*, 289–300.

Beretvas, S. N. (2008). Cross-classified random effects models. In A. A. O'Connell & D. Betsy McCoach (Eds.), *Multilevel modeling of educational data* (pp. 161–197). Charlotte, NC: Information Age 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. http://dx.doi.org/10.3102/0162373707299550

Bradley, R., Danielson, L., & Doolittle, J. (2005). Response to intervention. *Journal of Learning Disabilities, 38*, 485–486. http://dx.doi.org/10.1177/00222194050380060201

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

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.

Fuchs, L. S., & Fuchs, D. (2009). On the importance of a unified model of responsiveness to intervention. *Child Development Perspectives, 3*, 41–43. http://dx.doi.org/10.1111/j.1750-8606.2008.00074.x

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 for reading in the primary grades.* (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/

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.

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and Non-Hispanic students in early elementary school. *The Journal of Special Education, 34*, 90–103. http://dx.doi.org/10.1177/002246690003400204

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

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

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 Schools, 47*, 22–35. http://dx.doi.org/10.1002/pits.20449

Luo, W., & Kwok, O. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research, 44,* 182–212. http://dx.doi.org/10.1080/00273170902794214

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

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

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41,* 473–497. http://dx.doi.org/10.1207/s15327906mbr4104_3

National Center on Response to Intervention. (2010, March). *Essential components of RTI – A closer look at response to intervention.* Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention. Retrieved from http://www.rti4success.org/resourcetype/essential-components-rti-closer-look-response-intervention

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

Pinnell & Fountas. (2006). *Phonics lessons*. Portsmouth, NH: Heinemann.

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

Raudenbush, S. W., Bryk, A. S., Cheong, Y., & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Chicago, IL: Scientific Software.

Samuels, C. A. (2011). An instructional approach expands its reach. *Education Week, 30*(22), 2–5. Retrieved from http://www.edweek.org/ew/articles/2011/03/02/22rti-overview.h30.html?qs=an+instructional+approach+expands+its+reach

Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L., Johnson, C., . . . Crevecoeur, Y. C. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children, 77*, 207–228.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579–593. http://dx.doi.org/10.1037//0022-0663.91.4.579

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. http://dx.doi.org/10.1037/0022-0663.98.3.508

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*, 157–169. http://dx.doi.org/10.1177/00222194060390020401

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*, 541–561.

What Works Clearinghouse. (2008). *Procedures and standards handbook, version 2.0.* [Electronic version]. Retrieved 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

MICHAEL D. COYNE (Connecticut CEC), Associate Professor, Department of Educational Psychology, University of Connecticut, Storrs. **DEBORAH C. SIMMONS** (Texas CEC), Professor; **SHANNA HAGAN-BURKE** (Texas CEC), Associate Professor; **LESLIE E. SIMMONS** (Texas CEC), Program Manager; **OI-MAN KWOK** Associate Professor; **MINJUNG KIM,** Postdoctoral Research Assistant; **MELISSA FOGARTY** (Texas CEC), Postdoctoral Research Associate; **ERIC L. OSLUND** (Texas CEC), Postdoctoral Fellow; and **AARON B. TAYLOR**, Assistant Professor, Department of Psychology, Texas A&M University, College Station. **ASHLEY CAPOZZOLI-OLDHAM** (Connecticut CEC), Graduate Assistant; and **SHARON WARE** (Connecticut CEC), Research Associate, Department of Educational Psychology, University of Connecticut, Storrs. **MARY E. LITTLE** (Florida CEC), Professor, Department of Child, Family and Community Sciences; and **D'ANN M. RAWLINSON**

(Florida CEC), Director, Communications, University of Central Florida, Orlando.

Deborah Simmons is a coauthor and Dr. Michael Coyne is consulting author of the Early Reading Intervention, the experimental program investigated in this article. To address perceived conflicts of interest, independent researchers conducted data analyses and reviewed methods, analyses, results, and discussion sections of this article for objectivity and accuracy.

Address correspondence concerning this article to Michael D. Coyne, University of Connecticut Department of Educational Psychology, Neag School of Education, 249 Glenbrook Rd., Unit 2064, Storrs, CT 06269-2064 (e-mail: mike.coyne@uconn.edu).

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324E060067 to Texas A&M University. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

We would like to acknowledge and sincerely thank the school districts, interventionists, data collectors, and editors who assisted in this research and article preparation.

Manuscript submitted October, 2011; accepted July, 2012.