Exceptional Children

Vol. 77, No. 2, pp. 207-228.

Effects of Supplemental Reading Interventions in Authentic Contexts: A Comparison of Kindergarteners' Response

DEBORAH C. SIMMONS
Texas A&M University

MICHAEL D. COYNE
University of Connecticut

SHANNA HAGAN-BURKE

OI-MAN KWOK

LESLIE SIMMONS

CAITLIN JOHNSON

YUANYUAN ZOU

AARON B. TAYLOR
Texas A&M University

ATHENA LENTINI MCALENNEY

MAUREEN RUBY

YVEL C. CREVECOEUR
University of Connecticut

ABSTRACT: *This study compared the effects of 2 supplemental interventions on the beginning reading performance of kindergarteners identified as at risk of reading difficulty. Students (*N = *206) were assigned randomly at the classroom level either to an explicit/systematic commercial program or to a school-designed practice intervention taught 30 min per day in small groups for approximately 100 sessions. Multilevel hierarchical linear analyses revealed statistically significant effects favoring the explicit/systematic intervention on alphabetic, phonemic, and untimed decoding skills with substantive effect sizes on all measures except word identification and passage comprehension. Group performance did not differ statistically on more advanced reading and spelling skills. Findings support the efficacy of both supplemental interventions and suggest the benefit of the more explicit/systematic intervention for children who are most at risk of reading difficulty.*

Converging evidence suggests that individual differences in phonemic awareness and alphabetic knowledge in kindergarten have important implications for learning to read and predicting success in later grades (Elbro & Petersen, 2004; Foorman & Torgesen, 2001; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). Equally important are findings showing that

code-oriented intervention in the primary grades positively mediates early reading outcomes for many students who exhibit early reading risk (Cavanaugh, Kim, Wanzek, & Vaughn, 2004; Foorman, Breier, & Fletcher, 2003). Based on the accumulating empirical knowledge base in early reading intervention, schools have begun to provide more systematic code-based supplemental reading support in the primary grades, often as part of a response to intervention (RTI) process.

RTI is a multitiered system that recognizes the role of preventive supplemental intervention (Tier 2) to extend learning opportunities for students who score below benchmarks on universal screening measures (Gersten et al., 2008). Central to RTI is the assumption that the impact of reading difficulties and disabilities for many students can be moderated through early and progressively more intensive levels of instruction and intervention (Coleman, Buysse, & Neitzel, 2006; Fuchs & Fuchs, 2006). The RTI process in reading is designed to decrease the incidence and prevalence of children who experience reading difficulties and ultimately are identified with reading disabilities—outcomes that have direct implications for special education.

In the multitier RTI process, Tier 2 intervention in combination with effective Tier 1 instruction is designed to preempt the need for more intensive intervention for many children (Mellard & Johnson, 2008). Tier 2 involves supplemental intervention in small groups focused on a few reading components that allow for extended learning opportunities (Al Otaiba & Fuchs, 2006). Through the provision of effective Tier 1 and 2 interventions, the number of students who require more extensive reading instructional support is reduced, thereby enabling schools to use special education resources more strategically and effectively.

Emerging evidence suggests that timing is an important dimension related to the efficacy of supplemental beginning reading interventions (Foorman et al., 2003). Specifically, interventions that begin earlier are generally more effective than interventions that begin later (Foorman et al., 2003). The benefit of providing beginning reading supports in kindergarten has been the focus of several recent investigations, and findings consistently suggest that kindergarten intervention provides an important jump start for developing phonologic and alphabetic proficiency (Scanlon, Vellutino, Small, Fanuele, & Sweeney, 2005; Vadasy, Sanders, & Peyton, 2006; Vellutino et al., 1996).

[F]indings consistently suggest that kindergarten intervention provides an important jump start for developing phonologic and alphabetic proficiency.

The good news is that accumulating evidence documents the potential of supplemental Tier 2 intervention to reduce the prevalence and incidence of reading difficulties and disabilities among primary-grade children (Cavanaugh et al., 2004; Gersten et al., 2008). However, the reality is that many of the supplemental interventions that are widely used to address reading risk in kindergarten have not been rigorously evaluated using research standards recognized by educational and prevention science (Gersten et al., 2005; Society for Prevention Research, 2004; What Works Clearinghouse [WWC], 2008). Even fewer experimental interventions have been implemented in authentic school contexts and compared to typical intervention practices currently implemented in schools (Scammaca, Vaughn, Roberts, Wanzek, & Torgesen, 2007; WWC, 2007).

PRIOR RESEARCH ON SUPPLEMENTAL READING INTERVENTIONS

A growing body of research supports implementation of reading interventions for kindergarten students; however, relatively few studies have been conducted in conditions that parallel real school settings (Cavanaugh et al., 2004). A limitation of the evidence of early reading intervention is that the majority of studies had multiple factors that varied between comparison and experimental conditions, making it difficult to discern those that were responsible for effects. As a result, based on their review of 27 kindergarten reading interventions, Cavanaugh et al. called for "next steps" research that (a) investigates critical variables

through controlled experiments and (b) tests whether schools can achieve similar effects when interventions are implemented with available resources.

In a review of extensive reading interventions that met high-quality research criteria, Scammaca et al. (2007) identified four studies that focused on kindergarten children with identified early reading risk (Gunn, Biglan, Smolkowski, & Ary, 2000; Lennon & Slesinski, 1999; Schneider, Roth, & Ennemoser, 2000; Torgesen, Wagner, Rashotte, Rose, et al., 1999) and shared considerable commonality in the focus of intervention. All included at least one experimental condition that focused on explicit phonemic and code-based instruction. Findings documented that kindergarten interventions significantly enhanced phonemic and decoding skills, particularly for children who are among the 25% most at risk of reading difficulty (Scammaca et al., 2007).

Despite the commonalities, however, there was substantial variability in the delivery contexts and nature of the comparison conditions—variability that may be important to understanding implications for practice. Across the studies, supplemental interventions were delivered within 1-1 or 1-2 tutoring, a combination of 1-1 tutoring and small-group instruction, and small-group instruction only. Three studies used combinations of researcher-hired interventionists and teachers to deliver instruction, and one used general education teachers only. Intervention sessions ranged from 10 weeks of daily 30-min sessions to 88 hr of intervention distributed over 2.5 years. Comparison and control conditions varied from no-treatment comparisons to multiple-intervention comparisons.

RATIONALE FOR STUDY OF TIER 2 SUPPLEMENTAL INTERVENTIONS IN AUTHENTIC AND REPLICABLE CONTEXTS

Often, kindergarten intervention studies are designed to evaluate newly developed or unstudied interventions under highly controlled conditions. Such studies have helped shape our knowledge of the efficacy of specific curricular elements, but they do not allow us to estimate the effectiveness of the interventions in real school settings under typical conditions. For example, 1-1 supplemental tutoring is resource intensive and not feasible in many schools. As the next generation of scientific intervention research moves knowledge to practice, it is important to evaluate the degree to which interventions and their effects can be transferred to school contexts and the feasibility of implementation by school personnel.

Although national initiatives supporting early literacy have attracted critical attention, the supply of evidence-based programs that satisfy current methodological rigor and replicate the realities of implementation in today's schools is meager. Moreover, the majority of kindergarten studies have compared supplemental interventions to no-treatment controls, making it difficult to determine whether effects are due to the nature of the experimental reading intervention being evaluated or increased opportunity to learn provided through supplemental intervention or tutoring. As schools adopt RTI models and implement supplemental interventions, it is important to understand the relative effectiveness of the practices they typically use to address the needs of kindergarteners who are identified as at risk for experiencing reading difficulties.

PRESENT STUDY

Understanding the factors associated with effects is a critical issue in evaluating supplemental interventions. In the present study, we compared the effects of an experimental commercial kindergarten intervention, Early Reading Intervention (ERI; Pearson/Scott Foresman, 2004), to typical intervention practices currently found in schools. In a previous experimental study of a prototype of the ERI program (Simmons et al., 2007), effect sizes comparing ERI and a commercial supplemental intervention were moderate to large for all measures, with students who had the lowest preintervention scores benefiting the most. Effect sizes favoring the ERI condition compared to the commercial sounds and letters module were .82 on word identification and nonsense word fluency, .86 on spelling, and 1.52 on word attack. This study was conducted approximately six years earlier than the current study, prior to the time

when RTI and early supplemental intervention were practices commonly implemented in schools.

In the current study, our goal was to control for contextual variables to make a more meaningful comparison of a specific intervention program with typical kindergarten reading intervention. Both the ERI and typical practice intervention conditions supplemented general classroom instruction; were implemented for 30 min per day, 5 days per week; and were delivered by school-based personnel in small groups with children identified as at risk for future reading problems. This design allowed us to better isolate the impact of a widely available explicit and systematic reading intervention above and beyond typical intervention practices while controlling for many important contextual and implementation variables (e.g., opportunity to learn, interventionist, group size).

We were interested in whether children with low entry-level literacy skills would (a) benefit differentially from the commercial program compared to children in typical school-designed intervention and (b) whether either supplemental intervention would enable students to move out of reading risk. In addition, we explored teachers' perceptions of the ERI program to better understand its social validity.

This study contributes to the current literature concerning the impact of Tier 2 interventions on kindergarteners' early reading risk in two ways. First, it expands the evidence base of supplemental interventions by comparing a widely available and implemented experimental (commercial) intervention with a more rigorous comparison condition than has been employed in the majority of previous supplemental intervention studies. We sought to increase comparability by standardizing the number of intervention days, time allocated to intervention, instructors, and group size. A second contribution of this study is the potential to better understand the impact of these interventions under authentic and replicable conditions that parallel the resources and personnel typically found in schools. Finally, through multilevel analyses, we are able to more accurately understand the amount of variance explained by intervention.

We identified the benchmark goal for these at-risk students as performance at or above the 30th percentile on measures of phonemic awareness, reading, and spelling and compared the performance of individual students to these benchmark goals at posttest. The standardized criterion for risk, < 30th percentile (Torgesen, 2000) or greater than .75 *SD* units from the national mean, was considered a realistic benchmark against which to judge intervention success. We hypothesized that students who participated in the commercial program condition would (a) show greater improvement on a range of early phonemic and alphabetic measures and (b) evidence lower incidence of reading risk than students in the typical practice condition.

To increase the generalizability of findings, we used personnel who typically provided interventions in the schools (i.e., general kindergarten teachers, reading specialists, and trained paraprofessionals) and provided a modest level of professional development (2 days). We compared the effects of the commercial intervention program to typical practice intervention. To equate for extraneous variables, comparison group students were taught in similar-sized groups of three to five for similar amounts of time, rather than a no-treatment control.

Specific research questions included:

1. Are the effects of the explicit/systematic commercial program (ERI) stronger than typical practice intervention on the reading achievement of kindergarten children identified as at risk of reading disability on measures of phonemic awareness, alphabetic understanding, word attack, word identification, spelling, and passage comprehension?
2. What absolute levels of performance do kindergarten children attain relative to norm-referenced and criterion-referenced reading measures? And, does participating in the explicit/systematic commercial program condition reduce the number of children who are at risk of reading difficulty compared to typical practice intervention?
3. What are teacher perceptions of the ERI program (e.g., satisfaction, impact, feasibility)?

METHODS

Schools, Interventionists, and Students

Schools. Research staff met with district personnel and principals to recruit schools from two general regions, South-Central Texas (TX; $N = 4$) and Eastern Connecticut (CT; $N = 8$). A criterion for involvement was that the school typically provided some type of supplemental reading instruction for children with identified reading risk. All participating schools except one in CT received Title I funding and had high percentages of children who qualified for free or reduced-cost lunch services. The percentage of students qualifying for free and reduced-cost lunch ranged from 50% to 81% in CT schools and from 69% to 81% in TX schools. School enrollments ranged from 266 to 749 students in CT schools to 278 to 985 in TX schools.

Interventionists and Assignment to Condition. Research staff met with the principals of each school to explain the study and underscore the need to randomly assign teachers to treatment conditions. Assignment to condition was randomized at the interventionist level. Participating schools designated an interventionist for each kindergarten classroom (classroom teacher, paraprofessional, specialist). Kindergarten classrooms were then randomly assigned within schools to either the ERI experimental ($n = 31$) or the typical practice intervention (TPI) comparison condition ($n = 26$). Both conditions were present in each school. In response to participating schools' requests, when the number of classrooms was not even in a school, the odd classroom was assigned to the experimental condition.

In CT, types of interventionists varied both within and across schools and included classroom teachers, paraprofessionals, special educators, reading tutors, and so on. In TX, interventionists were participating children's respective kindergarten teachers, with the exception of one reading specialist. All teachers who had students who met eligibility criteria agreed to participate in the study.

Interventionists across conditions and sites received a written explanation that the study would compare the effects of different types of intervention on the beginning reading achievement of kindergarten children. Interventionists assigned to the comparison condition were encouraged to maintain their daily small-group instructional supplements for children identified as at risk and agreed to classroom observations.

Among the 12 participating schools, interventionists consisted of 48 certified kindergarten teachers and nine paraprofessionals. Demographic data for interventionists are found in Table 1. Chi-square analyses were conducted on categorical data and independent-sample *t*-tests for continuous variables to assess equivalence between conditions on interventionist demographics. The results of these analyses showed no reliable relations between groups on gender, total years of teaching experience, number of years of kindergarten experience, highest degree earned, ethnicity, or primary language. Further, analyses indicated no significant difference in the number of paraprofessionals in the experimental or comparison conditions or any demographic or instructional variables. Paraprofessionals differed from certified teachers on only one variable—highest degree earned.

Students. During the third week of kindergarten, researchers consulted with school personnel to identify children who (a) were considered in need of Tier 2 reading instruction, (b) were at least 5 years of age, and (c) received reading instruction in English. A two-phase screening process was used to identify participants. First, teachers and principals reviewed existing school-collected data (e.g., Dynamic Indicators of Basic Early Literacy Skills [DIBELS], Texas Primary Reading Inventory [The University of Texas System, 2006]) to nominate five to eight children per kindergarten classroom for further screening. Next, 310 children with parental consent were administered (a) the DIBELS Letter Naming Fluency ([LNF]. Good & Kaminski, 2002); (b) the Letter Identification Checklist (Woodcock Reading Mastery Test–Revised/Normative Update [WRMT-R/NU]; Woodcock, 1987, 1998); and (c) the Sound-Matching subtest from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999). These screening measures were selected based on their use in recent kindergarten intervention studies and their strong predictive validity for end-of-first-grade reading outcomes (Schatschneider et al., 2004).

TABLE 1

Interventionist Demographics by Condition

Variable	*ERI Experimental (n = 31)*		*TPI Comparison (n = 26)*	
	n	*%*	n	*%*
Female Teachers	29	93.54	25	96.15
Ethnicity				
Black or African American	2	6.45	0	0.00
Hispanic or Latino	4	12.90	3	11.54
White	25	80.65	23	88.46
Primary Language				
English	29	93.55	24	92.31
Spanish	2	6.45	2	7.69
Highest Degree Earned				
High School/ Less than Bachelor's	4	12.90	5	19.23
Bachelor's	17	54.84	14	53.85
Master's	10	32.25	6	23.08
Ed.S.	0	0.00	1	3.85
Total Years Teaching Experience				
Mean		12.16		14.12
Standard Deviation		9.69		11.51
Number of Years Teaching Kindergarten				
Mean		5.93		7.88
Standard Deviation		6.98		8.88

Note. ERI = Early Reading Intervention; TPI = typical practice intervention.

To estimate risk, we identified scores on each measure that were as close to the 30th percentile as possible. In total, 232 children qualified to participate by being among the five lowest performers on screening measures within their class and meeting one or both of the following criteria: (a) a score at or below the 33rd percentile on the DIBELS LNF measure (i.e., fewer than six letters correctly named in 1 min; Good & Kaminski, 2002); or (b) a score below the 37th percentile on the CTOPP Sound-Matching subtest (Wagner, Torgesen, & Rashotte, 1999). Of the 206 students who completed the study, 136 met both criteria, 64 met the phonemic awareness criterion only, and four qualified only on the letter naming fluency measure.

Student demographics for each condition are described in Table 2. As illustrated, 15 children (13.39%) in the experimental condition received special education services. Fourteen were served for speech/language impairments and one under the category of developmental delay. Nine children (9.57%) in the TPI comparison group received special education services. Eight received speech and language services (one of the eight also had a hearing impairment, and another had a visual impairment). The remaining child in the TPI group received special education services for developmental delay.

For the bilingual/English language learner classification, we used teacher-reported information indicating that children were learning to speak and read English. All children with this designation 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 significant relation between student characteristics and condition.

Attrition Analysis. A total of 232 kindergarten students were selected to participate in the study. Of this group, 206 (89%) participated in both pretest and posttest assessments; 26 (11%) did not complete the study and were not available at

TABLE 2

Student Demographics by Condition

Variable	*ERI Experimental (n = 112)*		*TPI Comparison (n = 94)*	
	n	%	n	%
Gender				
Male	59	52.68	47	50.00
Female	53	47.32	47	50.00
Ethnicity				
Asian	1	0.89	0	0.00
American Indian or Alaska Native	1	0.89	0	0.00
Black or African American	24	21.43	15	15.96
Hispanic or Latino	44	39.29	42	44.68
White	42	37.50	37	39.36
Receives Special Education Services	15	13.39	9	9.57
Bilingual/English Language Learner	24	21.43	29	30.85
Age				
Mean		5.49		5.39
Standard Deviation		0.32		0.27

Note. ERI = Early Reading Intervention; TPI = typical practice intervention.

posttest. The primary reason for attrition was students moving to another school.

To determine any differential attrition between students who exited the study and those who remained, we compared the two groups (i.e., the 26 attritors vs. the 206 nonattritors) on all demographic variables, including gender, age, bilingual, special education status, group (ERI and TPI conditions), and all screening and pretest measures. No significant differences were found between the two groups on any variable except gender, with the attritor group containing a higher proportion of male than female students. Analyses indicated no significant relation between condition and attrition; 7% of TPI and 15% of experimental intervention students did not complete the study.

PROCEDURE

Common Intervention Components. To increase comparability between conditions, a number of common instructional components were standardized across both conditions. Groups in both conditions consisted 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 intervention period, for an equivalent number of total sessions. Schools coordinated intervention delivery times according to their respective schedules; all intervention supplements occurred during the regular school day. Content in both conditions focused on early literacy skills, with experimental interventionists implementing the ERI curriculum and comparison interventionists providing typical practice reading intervention focusing on early literacy skills.

Experimental Condition. Students in the experimental condition received the ERI program, which includes 126 daily lessons. A typical 30-min lesson consists of seven activities, each designed to last 3 to 5 min. The first 15 min of the lesson focuses on phonological awareness and alphabetic understanding; the second 15 min integrates writing and spelling with previously taught phonemic and alphabetic skills.

The program is organized into four parts. Part I: Learning Letters and Sounds consists of 42 lessons and introduces 11 letter names and sounds and the phonemic skills of first- and last-sound isolation. Part II: Segmenting, Blending, and Integrating includes 30 lessons and continues with the introduction of five new letter names and sounds while introducing phonemic blending and segmenting using letter tiles. Part III: Reading Words introduces six new letter names and sounds with primary instruction focusing on

word decoding in vowel–consonant and consonant–vowel–consonant words. This component consists of 24 lessons in which instruction integrates oral segmenting and blending with real-word decoding and the introduction of irregular word reading. Part IV: Reading Sentences and Storybooks consists of 30 lessons. Instruction in this final section focuses on combining alphabetic skills and strategies with irregular word reading to read sentences and short storybooks.

The 126 lessons are highly specified and include detailed scripting to ensure clear and consistent communication of information. When introducing a new skill, the teacher models the information several times using consistent wording. In addition, skills are carefully integrated to enhance learning. Scheduled instruction, review, and feedback are explicitly incorporated into the program. Each lesson that introduces new information includes a specified number of instructional interactions in which the teacher first models the information. Students practice the new skill with the teacher and then apply it to discrimination or generalization tasks.

ERI Training. Professional development was designed to approximate training typically provided when schools adopt published programs and was limited to 2 days. The first session familiarized interventionists with lesson structure and the scope and sequence of Parts I and II 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 day of professional development was provided at implementation midpoint and took place in late January in CT and early February in TX. This session focused on Parts III and IV of the program and followed a format similar to that of the first professional development day.

ERI Fidelity. Each ERI interventionist was observed three times by research staff over the course of the school year. Observers coded procedural fidelity for each lesson activity on two items using a 1–4 scale (1 = poor, 2 = fair, 3 = good, 4 = excellent). The first item rated the extent to which a teacher completed all components (e.g., modeled examples, provided practice opportunities). The second item rated the extent to which a teacher was fluent with lesson wording and activities.

Typical Practice Interventions. Interventionists in the comparison condition were asked to provide typical school-designed beginning reading instruction for 30 min daily, in groups of three to five students. The comparison interventionists received no professional development; however, schools were not prohibited from providing additional professional development. The content and instructional approaches were allowed to vary naturally. Some schools had kindergarten intervention options already in place, others used participation in this research study as an opportunity to develop and implement Tier 2 interventions. Although some schools coordinated the intervention at a school level, most allowed interventionists to design and deliver the intervention individually. Teachers used a range of materials with 43% relying on commercial materials exclusively, 54% using self-developed lessons and materials, and 3% using a combination of commercial and teacher-developed materials. Of the commercial materials teachers used, 13% were Saxon Phonics, 28% SRA/Open Court, 10% Harcourt, 21% Lindamood Bell, 20% Lindamood Bell and Fountas and Pinnell combinations, and 8% Fundations. Typical practice interventionists kept implementation logs documenting when they met with their groups, session length, and student attendance. Typical practice intervention groups were observed informally throughout the study and were evaluated formally twice (once in fall and once in spring).

During formal observations, observers recorded the instructional content of each activity within TPI lessons using the following codes: (a) concepts of print, (b) phonological awareness (syllable/word level), (c) phonemic awareness (i.e., phoneme level), (d) sound blending and segmenting, (e) reading/decoding letter names, (f) reading/decoding letter sounds, (g) reading decodable words, (h) reading sight words, (i) reading connected text, (j) writing letters, (k) writing sounds, (l) spelling/encoding, (m) vocabulary, and (n) listening comprehension. During 29% of the TPI observations, two observers independently coded

the instructional content of all lesson activities using the categories listed previously. Using Kappa to calculate percentage of agreement, the reliability of codes for focus of instruction was $K = .73$.

Across the two observations, TPI lessons focused on a range of early reading skills. The majority of lessons (90.2%) included alphabetic and decoding activities, with 67% of lessons including instruction on letter names and 63% instruction on letter sounds. The second most prominent component was phonological awareness, with 49% of the lessons (53% in fall and 43% in spring) dedicating instructional time to tasks related to phonological processing. On average, 41% of the instruction focused on phoneme-level activities. Writing and spelling were observed in 39% of the lessons, whereas vocabulary and comprehension activities were observed in 14%. A substantial shift in emphasis in vocabulary and comprehension activities took place over time. That is, in the fall, fewer than 10% of lessons included instruction in vocabulary and comprehension, whereas 29% of the lessons observed during the spring incorporated vocabulary instruction and 24% included listening comprehension.

In addition to content, observers coded the quality of instruction on two items using a 1–4 scale (1 = poor, 2 = fair, 3 = good, 4 = excellent). The first item rated the extent to which the teacher's instructional delivery of content maximized the efficacy of curricular activities and was responsive to learner performance. The second item evaluated the quality of instructional time and the extent to which instructional time was maximized and free from interruption.

Measures and Assessment Procedures

The assessment battery included instruments designed to (a) assess normative and criterion-referenced growth on a range of early literacy constructs (e.g., phonemic, alphabetic), (b) evaluate growth on fluency and nonfluency-based measures, and (c) represent the progression of skills from the beginning to the end of kindergarten. Pretests were completed prior to the start of the intervention (approximately six weeks into the school year; September–October). Posttesting occurred within 2 weeks after the completion of the intervention (April–May).

Letter Knowledge. Two letter identification measures were used for screening only. The Letter Identification subtest of the WRMT-R/NU (Woodcock, 1987, 1998) is a standardized, untimed test battery that assesses a student's ability to name upper- and lowercase letters presented in various fonts. The Letter Naming Fluency subtest of the DIBELS (Good & Kaminski, 2002) measures a student's ability to rapidly name upper- and lowercase letters presented in 1 min. Alternate-form reliability for the kindergarten sample is .89. The split-half reliability coefficient for this subtest is .94 for the first-grade sample (coefficient for kindergarten sample not reported). The Supplementary Letter Checklist of the WRMT-R/NU (Woodcock, 1987, 1998) was administered at pretest and posttest to assess children's ability to correctly identify the letter name and produce the appropriate sound for each presented lowercase letter.

Phonological Awareness. The Blending Words (BW) subtest of the CTOPP (Wagner et al., 1999) is an untimed measure that requires the student to combine sounds into real words. Sounds are presented on an audiocassette tape. The internal consistency coefficient (Cronbach's alpha) is .88 for the age 5 sample and .89 for the age 6 sample. The Sound-Matching (SM) subtest of the CTOPP (Wagner et al., 1999) 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 Phoneme Segmentation Fluency (PSF) DIBELS subtest (Good & Kaminski, 2002) measures a student's ability to fluently segment three- and four-phoneme words. Scores indicate the number of sound segments correctly identified in 1 min. Alternate-form reliability for the kindergarten sample is .88.

Decoding. The Word Attack subtest of the WRMT-R/NU (Woodcock, 1987, 1998) is an untimed measure of a student's skill in reading a list of nonwords (e.g., "tet") presented in isolation. The raw score is the number of nonwords read correctly, which is converted into a standard score. The split-half reliability coefficient for this subtest is .94 for the first-grade sample (coefficient for kindergarten sample not reported). The

Nonsense Word Fluency (NWF) subtest of the DIBELS (Good & Kaminski, 2002) 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 min. 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 Phonemic Decoding Efficiency subtest of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) measures the number of nonwords students can pronounce in 45 s from a list that gradually increases in difficulty. The alternate-form reliability coefficient is .97 for the age 6 sample, and the test– retest reliability for ages 6 through 9 on form A is .90.

Word Identification. The Word Identification subtest of the WRMT-R/NU (Woodcock, 1987, 1998) measures a student's skill in reading a list of real words presented in isolation and is untimed. The split-half reliability coefficient is .98 for the first-grade sample (coefficient for kindergarten sample not reported). The Sight Word Efficiency (SWE) subtest of the TOWRE (Torgesen, Wagner, & Rashotte, 1999) measures the number of high-frequency words a student can read in 45 s. Presented in a list, words gradually increase in difficulty. The alternate-form reliability coefficient is .97 for the age 6 sample and the test–retest reliability for ages 6 through 9 on form A is .97.

Spelling. Spelling was measured using the Test of Written Spelling-4 (TWS-4; Larsen, Hammill, & Moats, 2005). 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.

Reading Comprehension. Reading comprehension was assessed using the Passage Comprehension subtest of the WRMT-R/NU (Woodcock, 1987, 1998). This measure utilizes a modified cloze procedure to assess a student's ability to read a short passage and supply a missing key word. The split-half reliability coefficient is .94 for the first-grade sample (coefficient for kindergarten sample not reported).

Vocabulary. Vocabulary was assessed at pretest only, using the Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1997), 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.

Procedures. Pre- and posttest assessments were conducted with individual students, who were seated next to an examiner at a desk or table in quiet environments outside of classrooms, including the offices of support personnel, unused classrooms, and cafeterias. All assessments were administered by trained data collectors. 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 of each test. Each data collector met the criteria of 90% accuracy in recording scores for a modeled administration of each measure. After data collection, each testing protocol was independently scored by two trained individuals.

DATA ANALYSIS

Because nonindependent observations were used due to the nesting structure in our data (i.e., 206 students nested within 57 different intervention groups), multilevel modeling (Hox, 2002) was applied to analyze the data, as it takes nonindependence into account. All multilevel models were analyzed using hierarchical linear models (HLM; V6.04; Raudenbush, Bryk, Cheong, & Congdon, 2004). Restricted maximum likelihood (REML) was used to estimate all the models, and two-tailed tests were used to evaluate intervention effects.

Intra-class correlations (ICCs) based on the unconditional model were calculated for the posttest measures and are included (Raudenbush & Bryk, 2002). ICC, which measures the magnitude of dependency between observations, can 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 curriculum and the TPI on the posttest measures, we used the corresponding effect size (δ_T; Hedges, 2007) to evaluate the practical significance of differences in cluster-randomized designs. We used the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) as recommended by the WWC (2008) to control for the potential inflated type I error rate that is due to the multiple comparisons. In accordance with the approach recommended by the WWC (2008), 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 of pretest and posttest measures for the experimental condition and the school-designed comparison group are presented in Table 3. Measures without corresponding pretest or posttest data indicate those considered too difficult for beginning of kindergarten or measures used for screening but not administered at posttest. Using HLM to test 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 and below the 36th percentile on DIBELS letter naming fluency. On average, students could name 10 to 11 letter names on an untimed measure and 5.7 letters in 1 min, and performed at the 22nd percentile on the PPVT-III measure of receptive vocabulary.

Posttests

Treatment Effects. At posttest, the experimental group obtained higher mean scores than the school-designed comparison group on all measures except WRMT-R/NU passage comprehension. We further examined posttest mean differences between conditions (i.e., intervention effect) using HLM with the following set of steps:

Level 1 (student-level) model:

$$\begin{aligned}\text{Posttest}_{ij} = {} & \beta_{0j} + \beta_{1j}\,\text{Pretest}_{ij} + \beta_{2j}\,\text{PPVT}_{ij} + \\ & \beta_{3j}\,\text{Age}_{ij} + \beta_{4j}\,\text{Gender}_{ij} + \beta_{5j}\,\text{Hispanic}_{ij} + \\ & \beta_{6j}\,\text{African\_American}_{ij} + \beta_{7j}\text{Special\_ed}_{ij} + \\ & \beta_{87j}\text{Bilingual}_{ij} + e_{ij} \qquad (1)\end{aligned}$$

where i represents the i-th student (i = 1…206), and j represents the j-th group (i.e., j = 1…57). Posttest_{ij} is the score of one of the posttest measures for the i-th student in the j-th group.

In this 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 age (Age_{ij}), gender (Gender_{ij}), ethnicity (represented by two dummy-coded variables: Hispanic_{ij} and $\text{African\_American}_{ij}$), 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 for which there were no corresponding pretests, we examined correlations to identify the pretest most highly associated with posttest performance. The untimed letter identification pretest score was highly correlated with word attack, word identification, and passage comprehension posttest measures and was used as a covariate. We also specified the classroom-level or group-level models as shown here:

Level 2 (classroom-level/group-level) models:

$$\begin{aligned}&\beta_{0j} = \gamma_{00} + \gamma_{01}\,\text{ERI}_j + \gamma_{02}\,\text{School}_{1j} + \\ &\gamma_{03}\,\text{School}_{2j} + \ldots. + \gamma_{012}\,\text{School}_{11j} + U_{0j} \text{ and} \\ &\beta_{1j} = \gamma_{10};\ \beta_{2j} = \gamma_{20};\ \beta_{3j} = \gamma_{30};\ \beta_{4j} = \gamma_{40}; \\ &\beta_{5j} = \gamma_{50};\ \beta_{6j} = \gamma_{60;}\ \beta_{7j} = \gamma_{70;}\ \beta_{8j} = \gamma_{80}. \quad (2)\end{aligned}$$

In Equation 2, ERI_j is a dummy-coded variable with 1 as the ERI experimental group and 0 as the TPI comparison intervention. In addition, because of the small number of schools participating (a total of 12 schools), we treated school as a fixed factor and created 11 dummy variables to represent the school effect, rather than adding

TABLE 3

Pretest and Posttest Means and Standard Deviations by Condition

	ERI Experimental (n = 112)				*TPI Comparison (n = 94)*			
	Pretest		*Posttest*		*Pretest*		*Posttest*	
Measure	M	SD	M	SD	M	SD	M	SD
PPVT-III	87.55	15.03	—	—	88.74	15.33	—	—
CTOPP Rapid Object Naming	7.65	3.16	—	—	8.26	2.99	—	—
WRMT-R/NU Letter ID	87.04	10.21	—	—	88.09	11.52	—	—
DIBELS Letter Naming Fluency	5.76	8.00	—	—	5.83	8.91	—	—
CTOPP Sound Matching	7.68	1.10	9.64	1.97	7.69	1.15	9.05	2.06
CTOPP Blending Words	8.52	1.66	10.49	1.96	8.46	1.80	10.00	2.33
WRMT-R/NU letter name checklist	10.46	8.05	25.61	4.28	11.04	8.06	25.00	4.80
WRMT-R/NU Word Attack	94.73	2.66	108.09	9.20	94.78	3.12	105.16	10.99
WRMT-R/NU Word ID	83.56	6.71	104.10	11.15	84.29	8.17	103.26	14.75
WRMT-R/NU letter sound checklist	—	—	24.93	6.00	—	—	22.86	7.50
WRMT-R/NU Passage Comprehension	—	—	97.83	9.29	—	—	99.09	11.34
DIBELS Nonsense Word Fluency	—	—	25.74	11.67	—	—	22.40	14.40
DIBELS Phonemic Segmentation Fluency	—	—	32.483	15.07	—	—	26.20	18.93
Test of Written Spelling-4	—	—	2.37	2.02	—	—	1.76	1.99
TOWRE Phonemic Decoding Efficiency	—	—	4.14	3.91	—	—	3.31	3.81
TOWRE Sight Word Efficiency	—	—	8.22	5.99	—	—	8.10	7.35

Note. CTOPP = Comprehensive Test of Phonological Processing; WRMT-R/NU = Woodcock Reading Mastery Test–Revised/Normative Update; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; TOWRE = Test of Word Reading Efficiency; PPVT-III = Peabody Picture Vocabulary Test; ERI = Early Reading Intervention; TPI = typical practice intervention.

another level to the model with an additional random effect. Treating schools as fixed covariates (i.e., the 11 dummy variables: $School_{1j}$, $School_{2j}$ …$School_{11j}$) not only takes the school effect into account, but also increases the statistical power (Bloom, Richburg-Hayes, & Black, 2007; Cook, 2005). The target effect, γ_{01}, represents the magnitude of different intervention effects between experimental and school-designed intervention groups on the posttest measure after controlling for all other variables, including the corresponding pretest covariate, demographic variables, and school effect. Results are presented in Table 4. The second column of Table 4 presents the corresponding ICC for each of the outcome measures based on the unconditional model (or random intercept model in which no predictors .11 to .41, which is very common in educational research [Hox, 2002]). The non-zero ICCs provided support for the nonindependent/correlated observations and the need to use multilevel models to adequately handle the nonindependency in our data.

After adjusting for the pretest covariate, demographic variables, and school effect, and controlling for the comparison-wise type I error rate using the Benjamini-Hochberg correction, the experimental group was found to score statistically significantly higher than the school-designed group on the following posttest measures: WRMT-R/NU letter name and letter sound checklists, DIBELS PSF, CTOPP sound matching and blending words, and WRMT-R/NU word attack. No statistically significant between-group differences were found on DIBELS NWF, TOWRE phonemic decoding and sight word effi-

TABLE 4

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

Measure	*ICC*	γ_{01}	t-*value*[a]	††*Two-tailed* p	*Effect Size* (δ_T)
Alphabet Knowledge					
WRMT-R/NU letter *name* checklist	.23	1.23	2.15	0.032††	0.36
Letter Sound Knowledge					
WRMT-R/NU letter *sound* checklist	.41	2.32	2.08	0.038††	0.41
Phonemic Awareness					
DIBELS Phonemic Segmentation Fluency	.27	7.07	2.60	0.010††	0.47
CTOPP Sound Matching	.21	0.74	2.62	0.008††	0.42
CTOPP Blending Words	.24	0.73	2.33	0.020††	0.40
Word Attack					
DIBELS Nonsense Word Fluency	.34	3.77	1.71	0.088	0.33
WRMT-R/NU Word Attack	.26	4.54	2.82	0.004††	0.51
TOWRE Phonemic Decoding Efficiency	.21	0.92	1.49	0.136	0.26
Word Identification					
WRMT-R/NU Word ID	.42	2.62	1.32	0.186	0.25
TOWRE Sight Word Efficiency	.41	0.24	0.25	0.802	0.05
Spelling					
Test of Written Spelling-4	.33	0.57	1.73	0.084	0.34
Passage Comprehension	.11	0.53	0.43	0.668	0.07

Note. ICC = Intra-class correlation based on the unconditional model (i.e., without any predictors in the model); CTOPP = Comprehensive Test of Phonological Processing; WRMT-R/NU = Woodcock Reading Mastery Test–Revised/Normative Update; DIBELS = Dynamic Indicators of Early Literacy Skills; TOWRE = Test of Word Reading Efficiency; ERI = Early Reading Intervention; TPI = typical practice intervention.

γ_{01}, as shown in Equation 3, is the difference between ERI and TPI groups on the posttest measure while holding the effects of other variables as constant. Positive value indicated that ERI group on average scored higher on the posttest measure than the TPI group.

[a]The approximate degrees of freedom (from HLM) for all the *t*-tests are equal to 44.

††Significant effect after Benjamini-Hochberg correction.

ciency, the Test of Written Spelling, WRMT-R/NU word identification, or passage comprehension. Further, 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. This finding should be interpreted with caution, however, given the small number of paraprofessionals serving as interventionists. Finally, we also examined whether the effect of the intervention varied by site (TX vs. CT) by adding a treatment condition by site interaction term to each of the models. None of these interaction terms were statistically significant (all $p > .20$), supporting the decision to analyze data jointly.

To evaluate the practical significance of the intervention effect between experimental and comparison groups, we used one of the effect sizes (i.e., δ_T) specifically designed for cluster-randomized studies (Hedges, 2007). It can be computed using the following equation:

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

where γ_{01} is the treatment effect (i.e., the mean difference between experimental and school-designed groups) after adjusting for the school effect and the demographical and pretest covariates. Similarly, σ^2 is the within-group variance, and τ_{00} is the between-group variance after adjusting for the school effect and the same set of

TABLE 5

Posttest Performance Percentile Distribution by Condition

	Percentile				
Measure	*<15th (%)*	*15–29th (%)*	*30–49th (%)*	*50–74th (%)*	*≥75th (%)*
Phonemic Awareness					
DIBELS Phonemic Segmentation Fluency					
ERI (*n* = 112)	16.07	26.79	33.93	18.75	4.46
TPI (*n* = 93)	36.56	22.58	19.35	13.98	7.53
CTOPP Sound Matching					
ERI (*n* = 110)	3.64	26.36	21.82	28.18	20.00
TPI (*n* = 93)	11.83	29.03	12.90	35.48	10.75
CTOPP Blending Words					
ERI (*n* = 109)	6.42	6.42	10.09	44.95	32.11
TPI (*n* = 93)	12.90	7.53	10.75	44.09	24.73
Word Attack					
DIBELS Nonsense Word Fluency					
ERI (*n* = 112)	8.04	28.57	26.79	33.04	3.57
TPI (*n* = 94)	24.47	20.21	21.28	25.53	8.51
WRMT-R/NU Word Attack					
ERI (*n* = 110)	1.82	0.01	20.00	16.36	61.82
TPI (*n* = 93)	2.15	1.08	36.56	18.28	41.94
Word Identification					
WRMT-R/NU Word ID					
ERI (*n* = 112)	8.93	6.25	15.18	36.61	33.04
TPI (*n* = 94)	21.28	5.32	4.26	28.72	40.43
Passage Comprehension					
ERI (*n* = 112)	8.04	17.86	25.89	38.39	9.82
TPI (*n* = 94)	14.89	10.64	18.09	38.30	18.09

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

covariates. As shown in Table 4, the standardized effect of the explicit/systematic condition ranged from .25 through .51 on all measures except measures of sight word reading and passage comprehension.

Treatment Response. All children who participated in the study fell at or below risk benchmarks on either the CTOPP sound-matching and DIBELS letter-naming fluency measures, or both, at pretest. There were no statistically significant between-group pretest differences by measure, and comparable percentages of children performed below the 30th percentile prior to intervention. To examine absolute response, we used posttest percentile rankings on normative and criterion-referenced measures, establishing the "out of risk" criterion as at or above the 30th or closest corresponding percentile by measure (Torgesen, Wagner, Rashotte, Rose, et al., 1999; Vadasy et al., 2006). We were interested in determining not merely whether students cleared the 30th percentile bar but to what degree.

Table 5 presents the percentages of students at each of the percentile intervals (e.g., 15th–29th) by condition by measure. Findings indicated not only different levels of response between groups but also varied levels of response by measure. Normative-referenced measures of untimed performance generally yielded a smaller percentage of nonresponders (e.g., 1.8% of experimental students and 3.3% of comparison students on WRMT-R/NU word attack) than timed

measures (36.6% of experimental and 44.7% of school-designed on DIBELS NWF). Patterns of response further indicated strong response (i.e., performance at or above the 50th percentile) by many students in both conditions (e.g., 24% and 22%, respectively, for experimental and typical practice intervention groups on the PSF measure). Conversely, interval-percentile analyses indicated that a substantial percentage of children failed to move out of significant risk (i.e., < 15th percentile). Chi-square analyses indicated that after adjusting the *p* value for multiple comparisons, differences between conditions were statistically significant at the .001 level on DIBELS PSF and NWF but not the WRMT-R/NU word identification measure. The only measure in which higher percentages of students in the ERI condition remained below the 15th percentile was passage comprehension (15% compared to 8%). Comparisons of the percentage of students who exceeded the 75th percentile indicated no statistically significant differences between conditions. To illustrate the distribution of response 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.

Results of between-group analyses indicated that for the majority of measures, the experimental condition produced stronger response than the typical practice intervention and in particular for students at the lower end of the distribution.

Regarding absolute performance, using the 30th percentile to indicate risk, findings reveal that both intervention conditions were effective for the majority of students. Results of between-group analyses indicated that for the majority of measures, the experimental condition produced stronger response than the typical practice intervention and in particular for students at the lower end of the distribution. For example, the plot for word attack indicated a stronger effect of the commercial intervention overall, with students at the lower quartile of the distribution still performing at approximately the 60th percentile compared to the lower quartile of the typical practice intervention condition who performed at the 30th percentile.

Fidelity of ERI Implementation. Across observations, procedural fidelity averaged 3.07 (*SD* = .64) on a 4-point scale (1 = low implementation; 4 = high implementation). Observers also coded the overall quality of the entire lesson on a 1–4 scale (1 = not at all adequate, 2 = less than adequate, 3 = adequate, 4 = above expectations). The average overall quality of implementation rating was 3.01 (*SD* = .63). Interrater reliability was established using a second independent observer for 25% of the direct observations. Kappa (*K*) was used to calculate percentage of agreement. Interrater reliability for each of the two procedural fidelity items was *K* = .63. Kappa for the overall quality of implementation item was .87.

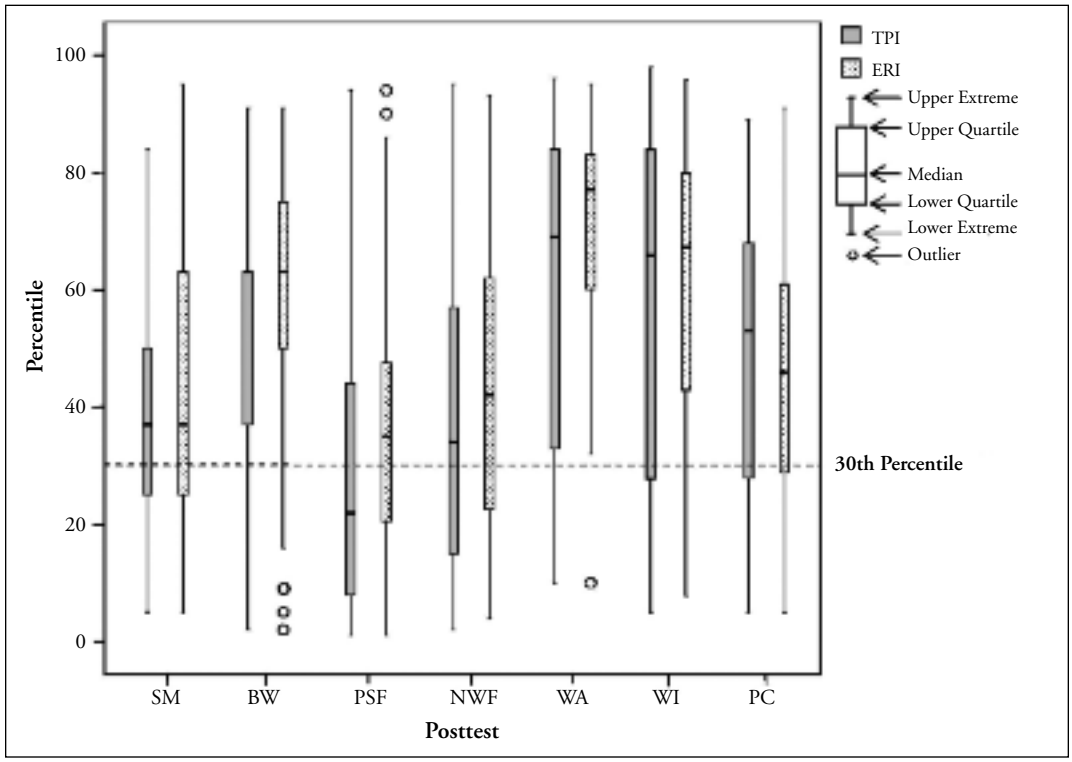
Characteristics of Typical Practice Instruction. Two items were used to describe the characteristics of typical practice instruction. On a 4-point scale, the quality of instructional delivery ratings averaged 2.65 (*SD* = .68) and quality of instructional time ratings averaged 3.41 (*SD* = .49). Interrater reliability was established using a second independent observer for 25% of the direct observations. Interrater reliability for instructional delivery ratings was *K* = 1.00. Kappa for the overall quality of implementation item was .81.

To evaluate whether ERI and TPI conditions differed on contextual variables that could affect outcomes, we evaluated group comparability on the following factors that were common to both conditions: (a) group size, (b) length of lessons, and (c) teachers' use of instructional time. No statistically significant between-group differences were detected on any of the measures. Specifically, group sizes for both conditions averaged 4 students, average lesson length ranged from 27 to 32 min, and teachers in both conditions maximized their use of instructional time with means ranging from 3.40 (TPI; *SD* = .74) to 3.51 (ERI; *SD* = .66) on a 4-point scale.

Social Validity. Table 6 depicts teachers' responses to the Teacher Satisfaction Survey administered only to teachers who taught the commercial program. As illustrated, teachers were generally positive about (a) their level of fluency

FIGURE 1

Posttest Performance Box Plots by Measure and Condition



Note. SM = Comprehensive Test of Phonological Processing Sound Matching; BW = Comprehensive Test of Phonological Processing Blending Words; PSF = DIBELS Phonemic Segmentation Fluency; NWF = DIBELS Nonsense Word Fluency; WA = WRMT-R/NU Word Attack; WI = WRMT/NU Word Identification; PC = WRMT-R/NU Passage Comprehension; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; WRMT-R/NU = Woodcock Reading Mastery Test–Revised/Normative Update; ERI = Early Reading Intervention; TPI = typical practice intervention.

with implementing the program, (b) their overall reaction to the ERI program, (c) the influence of the ERI program on improving their ability to teach children with early reading difficulties, and (d) their ability to transfer ERI strategies to teach other children.

DISCUSSION

This study assessed the relative and absolute effects of two types of supplemental interventions on the reading performance of children identified as being at risk of reading difficulty in their kindergarten classrooms. In the context of RTI, this study first sought to extend existing research on supplemental interventions by examining the effects of an explicit/systematic experimental (commercial) intervention with a treated comparison condition (schools' typical Tier 2 intervention) under comparable instructional conditions (e.g., interventionists, group size, duration). Second, it investigated the impact of both interventions under authentic and replicable conditions that parallel the resources and personnel typically found in schools. In addition, we examined teachers' perceptions of the experimental intervention and its impact on their ability to teach children with early reading difficulties.

Relative Effects of ERI Commercial Program Compared to Typical Practice Intervention

With respect to relative effects, performance trends favored students who participated in the

TABLE 6

Teacher ERI Satisfaction Survey

Survey Statement	Mean	SD
I used the ERI program with my small-group almost daily.	1.24	0.51
I became more fluent with the ERI program over time.	1.27	0.64
My ability to teach children with early reading difficulties improved through using the ERI program.	1.57	0.90
I have more confidence in my ability to teach kindergarten children with early reading difficulties.	1.72	0.96
I use strategies learned from the Early Reading Intervention when I teach other children.	1.47	0.97
I have changed the way I teach children with early reading difficulties.	1.93	0.96
Overall, I have a positive reaction to the ERI program.	1.53	0.86
I would recommend ERI to other teachers who teach struggling early readers.	1.53	0.90

Note. Numbers based on a scale ranging from 1 (Completely Agree) to 6 (Strongly Disagree). ERI = Early Reading Intervention.

ERI experimental condition compared to students in the TPI comparison condition when considering effects that were both statistically significant (i.e., 6 of 12 comparisons) as well as those that were not statistically significant but substantively important (i.e., 3 out of 12 comparisons). Effect sizes on measures of alphabet knowledge, letter sound knowledge, phonemic awareness, word attack, and spelling ranged from .26 (decoding efficiency) to .51 (word attack). Effect sizes for word identification and passage comprehension did not exceed the WWC (2008) .25 ES benchmark. Present findings corroborate and extend the efficacy of preventive supplemental intervention that teaches children identified with early reading risk to isolate and segment sounds in words and to systematically decode using letter–sound correspondences. A more detailed examination of findings suggests that the order and magnitude of effects corresponded with the scope and sequence of the ERI curriculum. That is, skills introduced earlier in the ERI curriculum and on which students had greater opportunity for instruction and practice (e.g., phonemic awareness, letter–sound knowledge, untimed decoding) evidenced more reliable and greater differential growth than those emphasized later in the program (e.g., decoding efficiency, whole-word spelling).

On advanced measures of word identification, spelling and comprehension, effects failed to reach statistical significance. Regarding word identification, students in both groups averaged eight words correct on the TOWRE measure of sight word efficiency. Four of the first 10 words on the TOWRE are decodable, whereas six require sight word knowledge, indicating that both groups of students acquired a combination of decoding skills and irregular word knowledge. The average ERI student received 103 lessons, the point in the curriculum when five irregular words had been introduced. Toward that end, findings indicate that compared to typical practice conditions, 103 or 126 lessons of the ERI program was not sufficient to yield statistically significant effects on word identification. With regard to the lack of findings on passage comprehension, we observed a significant increase in attention to reading comprehension in the typical practice interventions in the second half of the year in contrast to the ERI intervention, which included very little emphasis on comprehension. Although the typical practice intervention did not outperform ERI on passage comprehension, this finding suggests that future research should explore the benefits of explicitly targeting comprehension in kindergarten interventions.

Similar to previous kindergarten intervention studies, we found that implementation of an experimental early reading intervention was efficacious for children at early reading risk (Cavanaugh et al., 2004) on measures aligned with the content of instruction. However, the effects of our treatment were, on average, half the magnitude of

previously documented effects (e.g., Vadasy et al., 2006; Vellutino, Scanlon, Small, & Fanuele, 2006). Moreover, effects were lower than in a previous experimental study of the ERI program (Simmons et al., 2007). One possible explanation for the smaller effect sizes in the present study is the strength of our comparison condition and comparability of delivery conditions. That is, instead of a no-treatment control or the general classroom reading curriculum, we compared ERI to school-designed typical practice intervention that was implemented for similar amounts of time, with similar group sizes, and by similar interventionists. We believe that our comparison condition more accurately mirrors the typical intervention practices that are now implemented in many schools that are moving toward a multi-tier or RTI approach to instruction and intervention. Toward that end, effect sizes of .40 to.47 on phonemic awareness and .26 to .51 on word attack and decoding may be interpreted more precisely as program or curriculum impact separated from the impact of other instructional factors such as group size, interventionist, or number of instructional sessions (Scammaca et al., 2007).

Another explanation for the reduction in effect sizes from previous studies may in part be attributable to the statistical methods. That is, in the present study, multilevel models were used to more accurately apportion variance between student and intervention levels resulting in a more accurate but likewise conservative estimate of effects. Therefore, as we compare the effects of supplemental interventions, it is imperative that we consider the nested nature of the data and the statistical methods used to evaluate the effects of interventions.

Absolute Growth: Students' Response to Intervention

Regarding students' absolute growth, we were interested in whether interventions would change children's reading status from at-risk to not-at-risk. We compared students' end-of-year achievement levels in the ERI and TPI conditions to one another and also to national normative samples. All participating students were performing below the 30th percentile on measures of phonological awareness and/or letter knowledge in the fall of kindergarten. After approximately 100 sessions, significant changes in risk status were apparent for both groups on multiple dimensions of reading-related (e.g., letter knowledge, phonemic awareness) and word attack/decoding skills. As indicated in Table 5, students in both the ERI experimental and TPI comparison conditions who were at risk on measures highly predictive of later reading risk (Schatschneider et al., 2004) were achieving above the 30th percentile on the majority of measures at posttest. Thus, findings suggest that both conditions were effective in accelerating most students' learning compared to normative samples of similarly aged children and at levels that parallel those of prior kindergarten research (e.g., Gunn, Smolkowski, Biglan, & Black, 2002; Torgesen, Wagner, Rashotte, Rose, et al., 1999; Vadasy et al., 2006; Vellutino et al., 2006).

Results indicated that across a range of phonemic and word reading measures, two to four times more students in the typical practice intervention remained at reading risk (< 15th percentile) than in the explicit/systematic intervention. On measures of phonemic and nonsense word reading fluency, a significantly larger percentage of children in the TPI condition remained below the 15th percentile than in the ERI condition. Although differential trends are evident by condition, students' performance on fluency-based measures is consistent with previous early intervention research documenting their difficulty in developing automaticity and fluency with the component skills of reading (Torgesen, 2000). Prior research examining learner characteristics has indicated that kindergarten children who do not respond to instruction perform as much as 1.5 standard deviations below responders on measures of naming speed (Al Otaiba & Fuchs, 2006). Even though tasks that require rapid retrieval of phonemic and alphabetic information may differentiate responders from nonresponders, trends from this study suggest that systematic, code-based interventions may moderate this effect.

Teacher Perceptions of Early Reading Intervention

It is noteworthy that teachers in the experimental condition not only perceived that their ability to teach children with early reading difficulties improved through using the ERI program but also that they transferred strategies to teach other stu-

dents. These findings are important as we extend research into the contexts of use where teachers make decisions about methods and materials.

Limitations

Several factors must be considered when interpreting the findings. First, although we sought to standardize the number of sessions and amount of instruction students received, our data are based on self-reports by teachers, for which we have no systematic index of reliability. A second factor related to the intervention is the variability among typical practice interventions. Observations documented that the majority of teachers focused on highly important content and used time to maximize the effectiveness of instruction; nonetheless, inconsistency in emphasis and variability in instructional approaches and materials was apparent, which bears further research. Third, the variability in percentages of children identified as at risk varies by measure, and it is important to understand whether certain measures inflate the level of response and underestimate the risk. Thus, schools should carefully consider how they define "risk" and "response," both in terms of the measures they use and the criteria they set. Future research should provide further guidance on the reliability and usefulness of measures used to determine whether students are out of risk. Fourth, whereas random assignment of conditions within the same school strengthens methodological rigor, it also presents the possibility of intervention contamination or transfer. Though there was potential for teachers in the comparison condition to adopt practices from the commercial program, observers did not note this phenomenon at all.

Summary and Implications

As more and more schools provide early reading intervention in kindergarten, educators need evidence about the impact of supplemental interventions delivered in ecologically valid contexts. Prior research on the efficacy of early reading interventions has often compared experimental programs to no-treatment control conditions in which students received no supplemental intervention. Today, business-as-usual in kindergarten often includes some type of school-designed intervention for students who are at risk for reading difficulties. Therefore, it is increasingly important for kindergarten intervention studies to evaluate the efficacy of experimental treatments compared to typical practice interventions, while also controlling for important contextual variables such as group size, instructional time, and type of interventionist (Scammaca et al., 2007).

Our findings have important implications for both researchers and practitioners. First, the results support the efficacy of early preventive intervention and ERI in particular when used as an instructional supplement to improve foundational phonemic and alphabetic skills. The potential effects of ERI on more advanced reading skills were not documented, perhaps because of a truncated implementation. The amount of implementation, nonetheless, is a critical variable. It is possible that the standard implementation of ERI, although necessary for some, was not necessary for all. Whereas typical practice instruction was allowed to progress at a rate determined by the teacher, teachers in the ERI condition progressed at a standard pace through the program. Future research should consider a more responsive schedule of instruction based on students' individual progress on formative assessments. Although the experimental design of the present study did not manipulate or entertain differential progress in the curriculum, findings suggest that some students may have been able to progress through the curriculum at a faster pace. Conversely, some students may have benefited from more responsive and data-based instruction that used individual progress monitoring information to adjust instruction. Finally, a promising implication from this study is that both the ERI and TPI interventions accelerated the learning of many students identified as being at risk of reading difficulties at the beginning of kindergarten. This suggests that the intensity of implementation may be a key feature of effective interventions, beyond the impact of the specific program or materials. It also suggests that we must be cautious when interpreting effect sizes from studies that compare interventions to no-treatment controls. Because schools today often offer a fairly intensive school-designed intervention, the small to medium effects found for ERI may be a more realistic estimate of the value added by implementing a carefully designed, systematic intervention program. Moreover, carefully

designed programs that offer systematically designed instruction may be most important for children with the greatest reading risk.

REFERENCES

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39,* 414–431.

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.

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.

Cavanaugh, C., Kim, A., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal, 2*(1), 9–21.

Coleman, M., Buysse, V., & Neitzel, J. (2006). *Recognition and response: An early intervening system for young children at risk for learning disabilities.* Chapel Hill, NC: FPG Child Development Institute.

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.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test—Third Edition* (PPVT-III). Bloomington, MN: Pearson Assessments.

Elbro, C., & Petersen, D. K. (2004). Long-term effects of phoneme awareness and letter sound training: An intervention study with children at risk for dyslexia. *Journal of Educational Psychology, 96*, 660–670.

Foorman, B. R., Breier, J. I., & Fletcher, J. M. (2003). Interventions aimed at improving reading success: An evidence-based approach. *Developmental Neuropsychology, 24*, 613–639.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research and Practice, 16*, 203–212.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93–99.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading 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.

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.

Gunn, B., Smolkowski, K., Biglan, A., & Black, C. (2002). Supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school: A follow-up. *The Journal of Special Education, 36*, 69–79.

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

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

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

Lennon, J. E., & Slesinski, C. (1999). Early intervention in reading: Results of a screening and intervention program for kindergarten students. *School Psychology Review, 28,* 353–364.

Mellard, D. F., & Johnson, E. S. (2008). *RTI: A practitioner's guide to implementing Response to Intervention.* Thousand Oaks, CA: Corwin Press.

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.

Raudenbush, S. W., Bryk, A. S., Cheong, Y., & Congdon, R. (2004). *HLM 6: Hierarchical linear and non-*

linear modeling. Chicago, IL: Scientific Software International.

Scammaca, N., Vaughn, S., Roberts, G., Wanzek, J., & Torgesen, J. K. (2007). *Extensive reading interventions in grades K–3: From research to practice.* Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Scanlon, D. M., Vellutino, F. R., Small, S. G., Fanuele, D. P., & Sweeney, J. M. (2005). Severe reading difficulties—Can they be prevented? A comparison of prevention and intervention approaches *Exceptionality, 13*, 209–227.

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

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology, 92*, 284–295.

Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Santoro, L. E., . . . Kaufman, N. K. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design specificity. *Journal of Learning Disabilities, 40,* 331–347. doi: 10.1177/00222194070400040401

Society for Prevention Research. (2004*). Standards of evidence: Criteria for efficacy, effectiveness and dissemination* [Electronic Version]. Retrieved from http://www.preventionresearch.org

The University of Texas System, Texas Education Agency (2006). *TPRI Early Reading Assessment.* Austin, TX: Texas Education Agency

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15*, 55–64.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

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.

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.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R. S., & 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.

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.

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

What Works Clearinghouse. (2007). *Beginning reading* [Electronic Version]. Retrieved from http://ies.ed.gov/ncee/wwc/reports/topic.aspx?tid=01

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

DEBORAH C. SIMMONS (Texas CEC), Professor, Department of Educational Psychology, Texas A&M University, College Station. **MICHAEL D. COYNE** (Connecticut CEC), Associate Professor, Department of Educational Psychology, University of Connecticut, Storrs. **SHANNA HAGAN-BURKE** (Texas CEC), Associate Professor, Department of Educational Psychology; **OI-MAN KWOK**, Assistant Professor, Department of Educational Psychology; **LESLIE SIMMONS** (Texas CEC), Project Coordinator, Department of Educational Psychology; **CAITLIN JOHNSON** (Texas CEC), Graduate Research Assistant, Department of Educational Psychology; **YUANYUAN ZOU**, Graduate Research Assistant, Department of Educational Psychology; and **AARON B. TAYLOR**, Assistant Professor, Department of Psychology, Texas A&M University, College Station. **ATHENA LENTINI MCALENNEY** (Connecticut CEC), Graduate Research Assistant; **MAUREEN RUBY** (Connecticut CEC), Graduate Research Assistant;

and **YVEL C. CREVECOEUR** (Connecticut CEC), Graduate Research Assistant, Department of Educational Psychology, University of Connecticut, Storrs.

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.

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 and Kimberly Williams for her support in manuscript preparation. Last, we thank the school districts, administrators, teachers, interventionists, and children for their support and participation in this research.

Correspondence concerning this article should be addressed to Deborah Simmons, Texas A&M University, Department of Educational Psychology, 4225 TAMU, 704 Harrington Tower, College Station, TX 77843-4225 (e-mail: dsimmons@tamu.edu).

Manuscript received April 2009; accepted November 2009.