Article

Hammill Institute on Disabilities

Journal of Learning Disabilities
2015, Vol. 48(6) 602–621

DOI: 10.1177/0022219414521664
journaloflearningdisabilities.sagepub.com
SAGE

An Examination of the Efficacy of a Multitiered Intervention on Early Reading Outcomes for First Grade Students at Risk for Reading Difficulties

Hank Fien, PhD[1], Jean Louise M. Smith, PhD[1], Keith Smolkowski, PhD[2], Scott K. Baker, PhD[1], Nancy J. Nelson, PhD[1], and Erin Chaparro, PhD[1]

Abstract

This article presents findings of an efficacy trial examining the effect of a multitiered instruction and intervention model on first grade at-risk students' reading outcomes. Schools (N = 16) were randomly assigned to the treatment or control condition. In the fall of Grade 1, students were assigned to an instructional tier on the basis of *Stanford Achievement Test–10th Edition* scores (31st percentile and above = Tier 1; from the 10th to the 30th percentile = Tier 2). In both conditions, students identified as at risk (i.e., Tier 2; n = 267) received 90 min of whole group instruction (Tier 1) and an additional 30 min of daily small group intervention (Tier 2). In the treatment condition, teachers were trained to enhance core reading instruction by making instruction more explicit and increasing practice opportunities for students in Tier 1. In addition, at-risk readers were provided an additional 30-min daily small group intervention with content that was highly aligned with the Tier 1 core reading program. Results indicate significant, positive effects of the intervention on students' decoding and first semester fluent reading and potentially positive effects on reading comprehension and total reading achievement.

Keywords

efficacy, treatment, intervention, reading

Considerable research on reading has centered on the prevention of reading difficulties and early intervention (Snow, Burns, & Griffin, 1998). A key finding has been that explicit and systematic instruction focused on essential skills (e.g., phonemic awareness, alphabetic principle, reading in connected text, vocabulary and comprehension; National Reading Panel, 2000; National Research Council, 1998) can prevent and remediate early reading difficulties for many students (i.e., estimates range from 92 to 98%; Torgesen, 2000). Research also indicates that a range of early reading measures can identify students who need early intervention (Deno & Fuchs, 1987; Shinn, 1989). Identifying students who need intervention, and successfully teaching these students the early reading skills they need, forms the basis of a robust model of prevention and early intervention (Baker, Fien, & Baker, 2010; Kame'enui, Simmons, & Coyne, 2000).

However, these positive research findings have not found their way into reading practice in a way that has made a clear difference in terms of student impact. Patterns on the National Assessment of Educational Progress (NAEP, 2011) support this conclusion. For example, recent NAEP findings show that fourth grade reading scores have stagnated over the past few years, and only minor increases have occurred over the past two decades (e.g., in 1992, 38% of fourth grade students scored in the below basic category; in 2011, the figure was 33%).

One source of our inability to translate effective research practices into effective classroom practices may be the compartmentalized nature of reading intervention research. Separate studies in early reading intervention may identify strong effects in discrete reading areas such as phonological awareness and phonics, but teachers in schools must put these discrete pieces together into an overall system, which may be substantially more difficult than sequencing together the findings from separate research studies. There is a noticeable lack of research on how separate research-based findings should be integrated into a cohesive system of

[1]University of Oregon, Eugene, USA
[2]Oregon Research Center, Eugene, USA

Corresponding Author:
Hank Fien, University of Oregon, 5292 University of Oregon, Eugene, OR 97403, USA.
Email: ffien@uoregon.edu

reading instruction and support in public school settings (Gersten et al., 2009; National Joint Committee on Learning Disabilities, 2005). The purpose of this article is to present the results of an experiment on the impact of a multitiered reading intervention in first grade, Enhanced Core Reading Instruction (ECRI), that addressed all major reading components (i.e., phonemic awareness, phonics, fluency, vocabulary and comprehension instruction) and was designed to improve the system of instruction used in Tiers 1 and 2.

Multitiered Instruction as a Conceptual Frame to Operationalize Early Intervention Research

Multitiered systems of support represent an attempt to synthesize research on prevention and early intervention into a cohesive and comprehensive approach that is empirically grounded. Multitiered approaches provide reading instruction that increases in intensity within tiers and from one tier to the next (Gersten et al., 2009). Students' tier placement is based on universal screening, and students' movement between tiers is presumably based on progress monitoring data. Tier 1 instruction occurs in general education contexts and is conceptualized as core reading instruction. Tier 2 supports offer more intense instruction for students who need reteaching of concepts, intensified explicit instruction, and additional practice with teacher feedback. A hallmark of Tier 2 is that instruction is provided in small group formats. The most intense instruction is provided in Tier 3, which is characterized by highly explicit, systematic instruction, delivered in small group formats or individually.

Recently the Institute of Education Sciences (IES) released a practice guide on response to intervention (RtI) and multitiered reading instruction and intervention in the primary grades. This practice guide, as well as the 17 other IES practice guides, is intended to summarize research in critical areas of education practice and provide an organized set of recommendations for practitioners designed to guide implementation efforts. The practice guide provides five recommendations for implementing multiple tiers of support in school settings (Gersten et al., 2009). Each recommendation is rated according to the strength of scientific evidence supporting its effectiveness. Evidence ratings occur at three levels. A *strong* rating implies there is consistent and generalizable evidence linking the practice to improved outcomes. *A moderate* rating means either that there is strong causal evidence that may not be generalizable to the population addressed in the recommendation or that there is not strong evidence available for causal attributions. A *low* rating means that empirical evidence is meager and the recommendation is based primarily on the panel's expert opinion. Given the focus of the current study is on testing a multitiered framework of support in Tiers 1 and 2, we briefly describe the four practice guide recommendations for Tiers 1 and 2.

Tier 1. The practice guide includes two recommendations for Tier 1. The first is to screen all students for potential reading problems at the beginning and middle of the year and in between these assessments, and following the middle of the year assessments, to regularly monitor the progress of students at elevated risk for reading difficulty. In the early grades (K–2), appropriate target areas are letter naming fluency, phoneme segmentation, pseudoword reading, word identification, and oral reading fluency. The evidence for this recommendation was moderate, based on five correlational studies demonstrating these measures are associated with later reading performance (see Compton, Fuchs, Fuchs, & Bryant, 2006; McCardle, Scarborough, & Catts; 2001; O'Connor & Jenkins, 1999; Scarborough, 1998; Speece, Mills, Ritchey, & Hillman, 2003).

The second practice guide recommendation is to provide time for differentiated instruction within Tier 1 based on assessment data. The evidence for this recommendation was *low*, based on one study meeting What Works Clearinghouse standards with reservations (see Connor et al., 2009). In this study, 10 schools used student data to group students for instruction. Schools in the intervention condition received professional development on how to individualize instruction using recommendations and strategies provided by software program. Correlational results indicated that there was an association between fidelity and outcomes. Student growth was higher in classrooms where teachers implemented the program with greater fidelity (Gersten et al., 2009).

The concept of Tier 1 differentiation is important in the current study. An important purpose of differentiation is to make core reading content accessible to students at the lower end of the Tier 1 spectrum. In other words, differentiation can result in more effective instruction for low Tier 1 students. Another way to achieve the same differentiation principle is to expand the overall quality of Tier 1 materials and instruction so that all students at the lower end of the Tier 1 spectrum receive better Tier 1 instruction. In other words, these students do not have to struggle first to receive more intense Tier 1 instruction. A fundamental purpose of the ECRI intervention was to expand the overall quality of Tier 1 instruction to better meet the needs of students at the low end of the Tier 1 continuum.

Tier 2. Clearly, much more is known about effective early reading instruction in Tier 2 than in Tier 1 or Tier 3. The findings regarding Tier 2 show a high degree of consistency and strong consensus on the implications for districts and schools working to translate research to practice. The major conclusion of this research is that approximately 30 min of small group instruction each day (3 to 5 days per week) is highly effective for students who are struggling with learning to read. The way instruction is provided is carefully described in these studies, which demonstrate consistency

in how instruction should be delivered by the teacher (Gersten et al., 2009).

Virtually all recommendations of Tier 2 instruction focus on the fundamental and foundational importance of explicit instruction (Gersten et al., 2009). A number of components define explicit instruction across these Tier 2 interventions. Essential components are (a) teachers model the skills and knowledge they expect students to use and apply (Mathes et al., 2005), (b) teachers provide multiple practice opportunities for students during the lesson (Gunn, Smolkowski, Biglan, Black, & Blair, 2005; Vadasy, Sanders, & Peyton, 2005), (c) teachers correct and/or address student errors and misunderstandings immediately and systematically (Gunn et al., 2005), and (d) teachers pace lessons in a brisk manner so that student engagement is high and so that multiple areas of reading development (e.g., phonemic awareness, vocabulary, and word reading) can be addressed in the 30-min lesson.

The purpose of these research studies was to test the effectiveness of Tier 2 instruction and the vast majority of the studies do not state what the Tier 1 instructional program is, or what is occurring in Tier 1. The Tier 2 programs use separate materials, and the integration of Tiers 1 and 2 is not described. Conceptually, however, Tier 2 early reading instruction is supposed to directly supplement, not replace, Tier 1 instruction, and, in most conceptualizations of the three-tier model, Tier 2 is provided *in addition to* Tier 1 instruction (Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007). Thus, it is imperative to examine the effectiveness of Tier 2 in concert with Tier 1 implementation. In a recent review of 22 empirical studies, Hill, King, Lemons, and Partanen (2012) concluded that the "lack of data on Tier 1 instruction suggests that researchers who are evaluating the efficacy of Tier 2 elementary reading interventions are doing so outside of the context of RTI" (p. 122).

They based this conclusion on the fact that of the 22 studies conducted between 2005 and 2011, not a single study reported fidelity of implementation for core reading instruction (e.g., Case et al., 2010; Fuchs, Compton, Fuchs, Bryant, & Davis, 2008; Mathes et al., 2005). There were four exceptions in cases where researchers had implemented a supplemental Tier 1 intervention in the context of core reading instruction and consequently reported fidelity of implementation for the supplemental intervention only, and not for the general core reading portion (Loftus, Coyne, McCoach, Zipoli, & Pullen, 2010; McMaster, Fuchs, & Compton, 2005; Pullen, Tuckwiller, Konold, Maynard, & Coyne, 2010; Schuele et al., 2008). For example, McMaster et al. (2005) trained teachers to implement peer-assisted learning strategies (PALS) into the teachers' Tier 1 core reading program and reported fidelity data for the PALS portion of the intervention, but not for the remaining portion of the Tier 1 core reading time.

Studies of Tier 2 in the Context of Multitiered Implementation

Despite the popularity of RtI and multitiered intervention models, there are few experimental research studies on their effectiveness (Gersten et al., 2009; Hill et al., 2012; IES, 2008). Although there are numerous studies that examine the efficacy of discrete Tier 2 interventions, *there is a lack of a research base on the effect of multiple tiers of support on student outcomes* or studies that even provide data on what is occurring in Tier 1 (Hill et al., 2012). However, there are a few notable exceptions that have clearly established a connection between the nature of intervention delivered in Tier 2 and instruction delivered in the Tier 1 core reading program (Chambers et al., 2011; Kerins, Trotter, & Schoenbrodt, 2011; O'Connor, Harty, & Fulmer, 2005; Scanlon, Gelzheiser, Vellutino, Schatschneider, & Sweeney, 2008).

Two recent studies warrant attention and are most relevant to the current study in that they (a) employed clustered randomized controlled trials (i.e., randomly assigned schools to condition) and (b) directly or indirectly examined the relative effectiveness of variations of Tier 1 and Tier 2 intervention approaches on student learning outcomes. In one study, Chambers et al. (2011) examined the relative effects of a newly developed Tier 2 computer-assisted program and their standard one-to-one tutoring process. Both Tier 2 interventions were developed by the researchers to supplement and be integrated with their Success for All (SFA) Tier 1 core reading program; however, the authors reported that the computer assisted program was more aligned with their SFA core reading program. Although the authors found significant differences favoring the Tier 2 computer condition on Woodcock Word ID, Word Attack, and Passage Comprehension for the first grade participants and no significant differences for the second grade participants, they concluded that they could serve more at risk students in the highly aligned, integrated Tier 1 and Tier 2 approach that included small group computer-assisted supports. The authors also reasoned that the computer-assisted program was more effective (in first grade) or as effective (in second grade) as the one-on-one tutoring because it was more closely aligned with the Tier 1 SFA core reading program.

In another study, Scanlon et al. (2008) examined the relative effectiveness of three service delivery approaches for schools to provide two tiers of intervention for preventing early reading difficulties for at risk kindergarten students. Specifically, they compared a professional development only (PDO) condition, an intervention only (IO) condition, and a combined professional development and intervention (CPDI) condition. The PDO and CPDI conditions included a 3-day workshop on enhancing core reading instruction through the interactive strategies

approach (ISA). The ISA is not a program; rather, ISA is an approach that teaches teachers to best differentiate student supports by better understanding their students' instructional needs and maximizing instruction through precise student grouping. It also enables teachers to provide better instruction through a variety of techniques and activities (Scanlon et. al., 2008).

The IO condition provided small group instruction in groups of 3 students or fewer by research staff twice a week. The intervention was designed as a direct way for teachers to provide targeted intervention to at-risk readers utilizing principles of the ISA. Thus, it was not a packaged Tier 2 intervention, but a set of techniques that interventionists would implement that included "reading books, learning about letters and letter sounds, phonemic awareness, and writing" (Scanlon et al., 2008, p. 352). The CPDI condition incorporated the professional development from the first condition and the direct intervention components of the second condition; therefore, it arguably represented a highly aligned Tier 1 and Tier 2 model, given ISA approaches supported each instructional tier. In terms of student achievement, the authors hypothesized that the CPDI condition would be the most effective to improve student learning relative to the PDO or IO conditions. However, results indicated no significant differences across the three conditions for differentially reducing the percentage of students at risk for reading difficulties at the end of the school year.

Thus, across the Chambers et al. (2011) and the Scanlon et al. (2008) studies, results were mixed in terms of accelerating student learning through integrated and aligned Tier 1 and Tier 2 intervention approaches.

Overview of ECRI Multitiered Intervention

The multitiered instruction and intervention model implemented in the ECRI intervention incorporates many of the recommendations in the RtI practice guide, including (a) screening all students for reading difficulties at the beginning of the school year, (b) providing time in Tier 1 for differentiated reading instruction for all students based on assessment of student's current reading level, (c) providing intense and systematic instruction in Tier 2 on up to three foundational reading skills (i.e., targeted instruction; Gersten et al., 2009), and (d) monitoring the progress of Tier 2 students at least once per month (Gersten et al., 2009). However, there are key differences between ECRI and current models of RtI. First, ECRI increased the level of explicitness of Tier 1 instruction by redesigning the core reading program, following principles of direct instruction (Baker et al., 2010; Coyne, Kame'enui, & Carnine, 2011; Kame'enui, Simmons, Coyne, & Harn, 2003). Second, increased specificity in Tier 1 was achieved by implementing daily *lesson maps* and *intervention routines* that articulated the learning objectives and the procedures teachers would use to address these objectives (Western Regional Reading First Technical Assistance Center [WRRFTAC], 2005a, 2005b). Third, ECRI increased the alignment between Tier 1 instruction and Tier 2 intervention by implementing daily, 30-min Tier 2 lessons *using the same scope and sequence of skills from the core program*. These three principles are described in more detail below.

Increased explicitness. Although many schools implementing multitiered instruction models use core reading programs to guide Tier 1 instruction and support, we hypothesized that many of these programs do not provide the level of explicit instruction or instructional scaffolds students at risk of reading difficulty need to be successful (Baker et al., 2010). Explicit teacher modeling is frequently absent, and, if present, the models are vague and inconsistent. The practice opportunities necessary for deep understanding are limited in number and do not provide enough guided support for struggling readers. The practice opportunities are often restricted in range so that students end up practicing what they have learned under a narrow set of examples (Baker et al., 2010; Nelson-Walker et al., 2013). The ECRI enhancements were designed to make core instruction more explicit and systematic (Baker et al., 2010). Specifically, core reading programs were restructured to include the following explicit instructional elements: (a) learning objectives for the reading block were made more clear for the students, (b) teacher models were made more conspicuous through visual models, verbal directions, and fuller and clearer explanations, (c) explicit connections were systematically incorporated between new and previously learned concepts, (d) the amount of guided and independent practice opportunities were greatly increased and were calibrated to facilitate deliberate and highly accurate practice conditions for students, and (e) previously learned material was revisited and reviewed systematically (Carnine & Kameenui, 1992; Coyne et al., 2011).

The rationale for increasing the amount of explicit instruction in the context of Tier 1 is that the majority of at-risk students' reading instruction occurs in Tier 1 settings. For example, the RtI practice guide recommends 30 min of intervention *in addition* to Tier 1 instruction. Given that Tier 1 instruction may be 60 to 90 min per day, it stands to reason that this instruction should be aligned with the needs of struggling students. The ECRI Tier 1 enhancements maintained the core program materials, however the enhancements supplemented these materials when necessary and altered instruction by focusing on these materials by incorporating instructional design and delivery features that have been shown to be efficacious for struggling readers (Coyne et al., 2011).

Increased specificity. Our second principle was to increase the specificity for instruction and intervention routines. In Tier 1, we did this in two ways. *Lesson maps* were used to enhance

the design of existing programs that have already been created (e.g., WRRFTAC, 2005a). Lessons maps highlighted the most essential components of a lesson, provided time and grouping guidance for each component, and recommended deleting or deemphasizing parts of a lesson that are not essential (e.g., an arts and craft activity; Baker et al., 2010).

Second, *instruction templates* enhance the instructional design and delivery features of the core program by reorganizing lessons to increase explicitness and instruction intensity (WRRFTAC, 2005b). The *templates* provide (a) clear, step-by-step directions for how to model key reading skills and concepts and (b) guided and structured opportunities for students to practice foundational reading skills and to engage in interactive text-based discussions (Baker et al., 2010). Essentially, the *instruction templates* organize the instructional design elements of a lesson to be more explicit and to be more intensive for students that are reading below grade level. In addition, the high degree of specificity allows instructional leaders (i.e., reading coaches) to enter a classroom and quickly ascertain if instruction is begin delivered with fidelity.

Increased alignment. The final principle is strong alignment of content and instruction in Tiers 1 and 2 (Baker et al., 2010), to provide a cohesive instruction format within and across tiers. Gersten et al. (2009) recommend that schools should adopt supplemental Tier 2 interventions that are "compatible with their school's core reading program" (p. 20) but suggest that "alignment is not as critical as ensuring that instruction is systematic and explicit" (p. 23). We believe both goals are attainable and in concert should lead to student outcomes that are greater than either recommendation implemented in isolation (i.e., systematic and explicit instruction vs. instruction in Tiers 1 and 2 that is highly aligned; Baker et al., 2010). The ECRI intervention incorporates both of these features.

In the current study we examine the effectiveness of a Tier 2 intervention that is highly aligned with Tier 1 instruction. In fact, *we designed the small group lesson based on the scope and sequence of skill instruction from the Tier 1 core reading program*. We hypothesized that learners would benefit from the efficiencies gained by teaching fewer skills more thoroughly across both Tier 1 and Tier 2 (Baker et al., 2010). This prediction is based on strong theory and research demonstrating that deep and redundant learner practice leads to sustained cognitive changes, such as permanent lexical representations (Bransford, Brown, & Corking, 1999; Perfetti, 2007). For example, in an aligned intervention, small group instruction might include preteaching a new sight word in Tier 2 that will be taught the following day in Tier 1. We predict that this will be more beneficial for students at risk for reading difficulty compared to an intervention that teaches different content, with different instructional language and routines in Tier 2.

Purpose of the Study and Research Hypotheses

Due to the lack of systematic study of how multiple tiers of support function in concert, IES initiated the funding priority for the development and evaluation of systemic interventions, such as RtI or multitiered intervention models (IES, 2008). The purpose of this study was to evaluate the efficacy of a highly specified, multitiered reading intervention to improve the reading achievement of first grade at-risk readers relative to a business-as-usual multitiered condition. The project was funded through the IES Systemic Intervention topic area.

Our hypothesis is that at-risk readers attending intervention schools and receiving the multitiered ECRI Intervention would outperform students in control schools on a range of reading outcomes. We expected larger gains for at-risk readers on proximal measures, relative to distal measures, because our proximal measures were more closely aligned to the instructional focus of the intervention. In addition, we expected larger gains on foundational reading skills relative to comprehension measures because the major focus of the Tier 2 portion of the intervention focused on decoding, and accurate and fluent reading of connected text. It is important to note that we did not test the efficacy of an RtI model; rather, we tested the efficacy of a multitiered instruction and intervention model. Although we used screening data to place students in their respective tiers, and we did collect progress monitoring data to analyze student responsiveness, we did not use responsiveness data to regroup students or move students across tiers for the entire school year. Our rationale was that schools should not use responsiveness data to regroup students until after the school can demonstrate high-quality implementation of Tier 1 and Tier 2 intervention. In the subsequent year of the study, we conducted a follow-up study in which treatment schools utilized responsiveness data to guide intervention decision making that including regrouping of students across and within tiers.

Our second hypothesis is related to an important part of the instructional environment in which students make strong reading growth. This concerns the role of student reading practice, particularly in the context of instructional goals related to the development of fluent reading. Specifically, we expected that practice that occurs in the context of instruction that is deliberate and highly specified (the intervention condition) would lead to better reading gains than practice in the control condition (Ericsson, 2008). We also expected that intervention teachers would provide more opportunities for students to practice reading skills than control teachers. Thus, we expect an interaction between treatment condition and the rate of student practice opportunities, and the value of practice opportunities in the treatment condition to be stronger in terms of their association with student outcomes than practice

opportunities in the control condition. This hypothesis is tied to a rather straightforward proposition that practicing something (i.e., fluent reading) with lots of errors may be more harmful than similar amounts of practice that is more deliberate and error free.

Method

This study was part of a 4-year project designed to examine the efficacy of the ECRI multitiered reading intervention in first grade classrooms. Project ECRI was a cluster-randomized controlled trial that nested students and teachers within schools. In this study, we report results from the first wave of schools that participated in the project in the 2009–2010 academic year.

Participants

In the first wave of the project, 16 schools in three school districts were randomly assigned to the treatment or comparison condition, resulting in 8 schools in each condition. Within participating schools 42 first grade teachers participated in the study, including 23 teachers in the ECRI treatment condition and 19 in the comparison condition. Teachers averaged 14 years of teaching experience (*SD* = 8.71). Instructional assistants, employed by districts and trained by ECRI staff, provided small group instruction. A total of 267 students at risk for reading difficulty were included in this analysis (120 treatment; 147 comparison). Of these students, the schools reported that 5.2% received special education services (3.3% treatment; 6.8% comparison) and 12.0% were English language learners (12.5% treatment; 11.6% comparison). According to the 2009–2010 National Center for Educational Statistics (NCES), the average percentage of first grade Hispanic students across the schools was 21% (19% treatment; 24% comparison) and the average percentage of first grade African American students was 3% (1% treatment; 4% comparison). Also according to NCES, an average percentage of first grade students eligible to receive free- or reduced-price lunch was 46% (43% treatment; 49% comparison).

Recruitment and Assignment Procedures

In the spring of the year prior to participation in the project, school districts implementing a multitiered service delivery model for reading instruction were recruited to participate in the ECRI project. Principal investigators met with school district administrators, principals, and first grade teachers who expressed interest in participating to describe the project and answer questions about participation. To be eligible to participate, schools agreed to (a) use a published, comprehensive core reading program (identified and adopted through standard district procedures) during a 90-min reading block for Tier 1 and (b) provide students identified for Tier 2 with an additional 30 min of small group instruction.

In the first wave of the project, we recruited 22 schools in three Oregon school districts to participate in Project ECRI. Four of these schools elected to not participate in the study because of changes in school leadership between recruitment and the beginning of the study. The remaining 18 schools were randomly assigned to the treatment or comparison (standard practice) condition. We blocked schools on district before random assignment to control for core curricula and other important factors, ensuring similar schools in each condition. After random assignment, two schools participating in Wave 1 (one treatment and one comparison school) left the project, leaving 16 schools participating in this study.

Identification of At-Risk Readers

Students enrolled in first grade classrooms (n = 1,037 were recruited to participate in the project, 533 in the treatment condition and 504 in the comparison condition). Of these students, we obtained parental consent for 969 students, 500 in the treatment condition and 469 in the comparison condition. Parents and guardians were unaware of the school's assignment to condition at the time of consent. Fall scores on the reading portion of the *Stanford Achievement Test–10th Edition* (SAT10) were used to assign students to Tier 1, Tier 2, or Tier 3, although Tier 3 was not a focus in the project. Students that scored between from the 10th to the 30th percentiles on the SAT10 in the fall of first grade were assigned to Tier 2. Students that scored at or above the 31st percentile on the SAT10 or at the 9th percentile or below were assigned to Tier 1 and Tier 3, respectively. Of the students for whom we obtained parental consent, 870 students participated in the fall assessment (422 treatment, 448 comparison). In total, 360 students were assigned to Tier 1 (167 treatment, 193 comparison), 267 students were assigned to Tier 2 (120 treatment, 147 comparison), and 243 students were assigned to Tier 3 (135 treatment, 108 comparison).

Implementation

Treatment and comparison teachers conducted daily reading instruction using their district-adopted, published, comprehensive core program during a 90-min reading block for Tier 1. Across the project, districts were implementing Houghton Mifflin and Scott Foresman core reading programs. Within district, the same core program was used across treatment and comparison conditions. Students identified for Tier 2 received an additional 30 min of small group instruction, delivered by the classroom teacher or an instructional assistant.

Treatment condition. The ECRI intervention was designed to increase the quality of instructional interactions between teachers and students by supporting the implementation of high-quality explicit instruction. The intervention supported implementation through (a) Tier 1 ECRI, (b) Tier 2 small group instruction that was highly aligned with Tier 1 instruction, and (c) comprehensive professional development.

Tier 1: ECRI. The mean Tier 1 class was 21.5 students (*SD* = 5.1), with a mean of 22.4 students (*SD* = 4.5) in the treatment condition and a mean of 20.7 (*SD* = 5.5) students in the comparison condition. As previously described, we enhanced core instruction so that it was more explicit and systematic than would otherwise be provided using the core reading program alone (Baker et al., 2010). To enhance core instruction, as previously described, ECRI teachers used lesson maps that specified content and instructional templates that specified explicit teaching routines (WRRFTAC, 2005b). Teachers reported that Tier 1 classrooms spent an average of 57.5 (*SD* = 24.8) min in whole group instruction, 29.4 (*SD* = 21.7) min in small group instruction and 34.7 (*SD* = 22.7) min in independent work during Tier 1 instruction.

Tier 2: Small group instruction highly aligned with ECRI Tier 1. District-employed instructional assistants led ECRI small groups. Overall, 19 small group instructors participated in the study in the treatment condition. All of the small group instructors were female. Of the small group instructors in the treatment condition, 17 were instructional assistants, one was a classroom teacher and one was a Title I teacher. Groups met 30 min daily for approximately 26 weeks between pre- and posttesting. The location where Tier 2 intervention was provided differed across schools, but was most often held in a reading specialists' room outside of the Tier 1 classroom.

Tier 2 intervention materials were developed by the project team to be highly aligned with core reading program content. Small group instructors used the same instructional routines as were used in Tier 1 to ensure consistent instructional language and delivery across enhanced core and intervention instruction. Students receiving Tier 2 intervention were pretaught content in daily, 30-min lessons in small groups comprised of 3 to 5 students. For example, if the sound spelling /oi/ was being taught in the next day's core lesson, then small group instruction preceding that lesson would preteach sound spelling /oi/, along with sound combinations and words that would be taught during whole group instruction the next day. Although multitiered intervention was delivered using a standard protocol approach, teachers and small group instructors used data to group students and to refine instruction. For example, the small group instructor tracked group and individual errors during instruction and used a portion of the time devoted to Tier 2 small group intervention to reteach content as needed. Because of the focus on preteaching and increasing students' accuracy with core content, we hypothesized that our at-risk readers would be more engaged and successful during practice opportunities during the whole group instruction.

The content of small group lessons emphasized foundational reading skills including phonemic awareness, word reading, and fluency in reading connected text. Core programs introduced one to two new sounds per week in the beginning of the year and two to four sounds per week starting midyear. At the beginning of the year, core programs introduced easier to pronounce sounds and sounds that occur more often in shorter words. As the year progressed, the order of when blends, digraphs, long vowel sounds, and vowel teams were introduced varied by program. Because the Tier 2 intervention was aligned to each school's core program, the scope and sequence of foundational skills content varied slightly by core program. Each 30-min daily lesson included seven activities.

Irregular word reading (2 min). The lesson began with irregular reading that included common phonetically irregular (nondecodable; e.g., "the") and phonetically regular words for which all the sound spellings had not yet been taught (decodable; e.g., "them" when /th/ had not yet been introduced). Generally, six to eight new words were introduced each week (Days 1, 4, and 5) and eight words were reviewed daily. Words being introduced for the first time were taught by prompting the students to repeat the whole word, spell the word, and repeat the word again. Review words were read as a whole word.

Phonemic awareness (1 min). The lesson included eight words to be used with the phoneme blending routine and eight additional words to be used with the phoneme segmentation routine. Words used for blending and segmenting practice were either from the core programs phoneme blending or segmenting activities, or words that were added that contained the new sound spellings and previously taught sound spellings.

Sound spelling introduction and review (2 min). After phonemic awareness activities, the new sound spelling was practiced and previously taught sounds spellings were reviewed. Sound-spelling cards from the core program were used in the intervention program.

Blending and word reading (4 min). Blending included eight words to be used with a sound-by-sound blending routine to practice decoding words that included new sound spellings and previously taught sound spellings. Words used for blending practice were selected from the core program texts or from the core program blending practice activities.

Word reading included 16 words read as a whole word. Words practiced had been used for blending practice on a previous day. The words included new sound spellings and previously taught sound spellings.

Accuracy and fluency reading decodable text (12 min). Nearly half of the lesson time was then focused on reading decodable text for the purposes of applying decoding skills and building accuracy and fluency in connected text. Decodable readers from the core program were used as texts so that the decodable text corresponded to the phonic elements being introduced in the Tier 1 lesson. Interventionists provided word-by-word reading with students rereading the sentence as a whole. The interventionist monitors students' accuracy and provides corrective feedback. Students then individually reread the text for accuracy and fluency practice while the interventionist monitored and checked for individual accuracy and fluency mastery. An introductory routine was used during the first month of the intervention program that fully modeled the lesson. An intermediate routine was used during Months 2 through 9, which included less teacher modeling and wording/scaffolding. As students reading fluency reached end-of- year goals, they transitioned to advanced routine that provided less think time between words to enable students to read accurately and fluently in the text.

Encoding practice (5 min). The encoding practice involved explicit instruction to spell four to six words that included the sound spelling being introduced in the lesson and previously taught sound spellings. The words used for encoding practice were from a previous day's lesson for blending and regular word reading practice. Students practiced using phonemic segmentation to identify the sounds in the presented word and then wrote the letter sounds to form the correctly spelled word. After students wrote the word, the interventionist wrote the correct spelling on a white board and students corrected their mistakes, if necessary.

Reteaching of challenging words (3 min). The final activity of the lesson involved practicing one to eight sound spellings and words that students had pronounced or read incorrectly during the small group lesson. For example, the interventionist collected a list of errors made by the group or by individuals. At the end of the lesson, the interventionist displayed the sound spellings and words from this list and re-presented the sounds spellings and words following the explicit routines for additional practice.

Professional development and coaching. Teachers in the treatment condition participated in 3 days of professional development on the ECRI instructional model prior to the beginning of the school year and 2 days of follow-up activities in October, delivered by ECRI trainers (i.e., ECRI expert coaches and project staff). Professional development for teachers emphasized (a) instructional teaching routines; (b) overview of research on beginning reading content and skills including phonemic awareness, phonics, vocabulary, comprehension, and fluency in reading connected text; and (c) strategies for increasing student engagement in the lessons. Small group instructors received 2 days of professional development in the fall and 1 day in January. Professional development for small group instructors emphasized instructional teaching routines and strategies for increasing student engagement in lessons. All professional development sessions included instructional delivery demonstrations by the ECRI trainers and opportunities for participants to practice using instructional strategies with feedback from the ECRI trainers.

Teachers and small group instructors also received comprehensive coaching support from ECRI project staff through classroom and small group visits once per month and regular study group meetings facilitated by an ECRI expert coach. The high degree of specificity utilized in lesson plans and instructional templates allowed ECRI coaches to enter a classroom and quickly determine if instruction was being delivered as intended. Regular observations allowed coaches to provide feedback to teachers regarding their delivery of instruction related to the explicit instructional design principles associated with the intervention. ECRI coaches followed a coaching protocol to ensure standardization of coaching support across the treatment schools.

Comparison condition. Comparison schools used the standard, district-adopted core reading program for Tier 1 instruction. Comparison teachers reported that Tier 1 classrooms spent an average of 36.5 (SD = 19.9) min in whole group instruction, 44.3 (SD = 22.3) min in small group instruction and 28.0 (SD = 18.2) min in independent work. The research team provided control teachers the list of students that qualified for Tier 2 support and were asked to deliver Tier 2 support for the students for the entire school year. Comparison schools used a business-as-usual model to deliver Tier 2 small group instruction (i.e., they did not use the ECRI Tier 2 intervention materials). Teachers in the comparison condition reported that Tier 2 intervention materials included a range of supplemental and intervention materials, including published, standardized protocol intervention materials as well as teacher-developed materials.

In the comparison condition, the school district scheduled and provided standard professional development activities. In the comparison condition, 80% of teachers reported that someone in their school provided them with instructional support, guidance and coaching about teaching students how to read. Teachers that indicated they received support reported that they received an average of 6.0 (SD = 12.7) hr per month of this type of coaching. Teachers in the comparison condition reported receiving 1.7 (SD = 1.6)

Table 1. Fidelity of Implementation by Intervention Condition.

	Comparison (n = 24)		Intervention (n = 18)	
	M	*SD*	*M*	*SD*
1. Lesson maps used	.11	.16	.90	.16
2. Card templates used	.08	.13	.91	.16
3. Instruction in phonemic awareness	.17	.28	.85	.19
4. Instruction in alphabetic principle	.62	.26	.94	.13
5. Instruction in word reading	.69	.30	.90	.16
6. Instruction in reading connected text	.87	.17	.85	.21
7. Instruction in vocabulary	.50	.33	.60	.31
8. Instruction in comprehension strategy	.59	.25	.57	.33
9. Consistent with intentions and design		—	.83	.21
10. Activities in maps were executed		—	.76	.16
11. Lesson was complete		—	.73	.29
Explicit instruction[a]	.57	.16	.78	.16
ECRI specific fidelity summary score[b]		—	.83	.15

Note. Fidelity items were rated as 0 (*no*), 0.5 (*partially*), or 1 (*yes*).
[a]The explicit instruction score represents the mean across Items 3 through 8. [b]The Enhanced Core Reading Instruction (ECRI) specific fidelity represents the mean across Items 1, 2, 9, 10, and 11. Items and scale scores were averaged across the three observation occasions.

days of professional development directly related to reading instruction across the school year, on average.

Fidelity of implementation. Direct observations were conducted only in the general education classrooms in both the treatment and control conditions. Research staff provided ongoing data to the Tier 2 interventionists regarding implementation fidelity, though these data were not standardized or used in the analysis. Observations of implementation fidelity conducted by trained data collectors indicated nearly all treatment teachers used lesson maps (M = .90) and instructional templates (M = .91) during instruction, and instruction in treatment classrooms was consistent with the intent of the lesson maps and templates to improve the quality of explicit instruction (M = .83). As expected, observations conducted in comparison classrooms indicated comparison teachers rarely used lesson maps (M = .11) and instructional templates (M = .08) during instruction. Although treatment diffusion was minimal, some comparison teachers did have access to intervention materials. Table 1 presents a summary of fidelity data, indicating the extent to which (a) ECRI materials were used in treatment and comparison classrooms (Items 1–2), (b) instruction provided in each of six areas of reading was explicit (Items 3–8), and (c) instruction in treatment classrooms was consistent with the targets of the ECRI intervention (Items 9–11).

Student Assessment Measures and Procedures

Prior to testing students in the fall, data collectors attended 3 days of standardized assessment training. Across the winter and spring, data collectors attended 4 days of additional training. For individually administered student measures, assessment coordinators evaluated interrater agreement by shadow scoring with data collectors and providing feedback on test administration. Interrater agreement is reported for assessments in the paragraphs below.

Dynamic Indicators of Basic Early Literacy Skills. The *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) assessments were used to measure phonemic decoding skill (i.e., the alphabetic principle) and passage reading fluency. The Nonsense Word Fluency (NWF) and Oral Reading Fluency (ORF) measures were administered in the fall, winter, and spring. Trained data collectors administered DIBELS assessments in all districts. Average interrater agreement was 97.7% (range = 90%–100%) for NWF and 97.4% (range = 91%–100%) for ORF across the study.

NWF. NWF (Kaminski & Good, 1996) is an individually administered, 1-min, timed measure of student skill in reading consonant–vowel and consonant–vowel–consonant pseudowords. Students can either provide the sounds of the individual letters or the pseudoword as a unit. The score for Correct Letter Sounds (NWF-CLS) was obtained by counting the number of correct letter sounds students provided. The score for Words Read Correctly (NWF-WRC) was obtained by counting the number of nonwords students recoded correctly. NWF alternate-form reliability coefficients range from .67 to .80, and concurrent validity coefficients with readiness subtests of the *Woodcock–Johnson Psycho-Educational Test* ranged from .35 to .55 (Good & Kaminski, 2002).

ORF. DIBELS ORF (Good & Kaminski, 2002) is an individually administered, 1-min, timed measure of student skill in accurately and fluently reading connected text. The number of words read correctly in 1 min is the student's score on a single passage. To determine a student's benchmark score, the student is administered three grade level passages at a single benchmark assessment time point during the school year (beginning, middle, or end) and the median score is recorded. In the beginning grades, ORF alternate-form reliability coefficients range from .89 to .94, and test–retest reliability coefficients range from .92 to .97 (Good & Kaminski, 2002).

Woodcock Reading Mastery Tests–Revised. The *Woodcock Reading Mastery Tests–Revised* (WRMT; Woodcock, 1998) is a standardized, comprehensive battery of tests that measures multiple aspects of reading ability, including comprehension, word recognition, and word analysis. Two skill clusters comprised of four subtests were administered to students and used in this analysis: the Basic Skills Cluster (WRMT-BS), which includes the Word Identification and Word Attack subtests, and the Comprehension Cluster, which includes Word Comprehension and Passage Comprehension subtests. These two clusters compose the Total Reading (WRMT-TR; full scale), which includes the Word Identification, Word Attack, Word Comprehension, and Passage Comprehension subtests. The total standard score, derived from grade-based norms, was used in all analyses. As reported in the testing manual, the correlation between scores on the WRMT-TR and the *Woodcock–Johnson Psycho-Educational Battery*, Total Reading Full Scale is .88 in Grade 1. Internal consistency for the subtests ranges from .94 to .98 in Grade 1. We administered Form H of the WRMT in the fall and spring. Average interrater agreement between data collectors across the study was 97.8% (range = 93 – 100%).

SAT10. The SAT10 (Harcourt Educational Measurement, 2002) is a group administered, norm-referenced test of reading proficiency. Total scaled scores, derived from grade-based norms, were used in all analyses. The SESAT 2 was used in the fall and the Primary 1 was used in the spring of Grade 1. The test manual indicates that Kuder–Richardson reliability coefficients, a measure of internal consistency, are .94 for SESAT 2 and .97 for Primary 1. Also, SESAT 2 and Primary 1 Total Reading scores are moderately correlated with *Otis–Lennon School Ability Test, 8th Edition* total scores (r = .68 and .61, respectively). Trained data collectors administered the SAT10 in the fall and spring of first grade to all students participating in the project. Subtests administered included Sounds and Letters, Word Reading, and Sentence Reading (SESAT 2), or Word Study Skills, Word Reading, Sentence Reading, and Reading Comprehension (Primary 1). Total SAT10 testing time ranged from 110 to 155 min. Results of the fall administration were used to assign students to tiers, whereas spring results served as the primary reading outcome measure in the study.

Observations

Classroom Observations of Student Teacher Interactions. The classroom observation protocol used in the study was a modified version of the *Classroom Observations of Student–Teacher Interactions* (COSTI; Smolkowski & Gunn, 2012), which is used to record contextual information and quantify student–teacher interactions during instructional activities. In the modified version, documented context codes include group size (i.e., whole group, small group, and number of students), content codes (i.e., phonemic awareness, alphabetic principle, word reading, reading connected text, vocabulary, comprehension, nonreading content, and other), curriculum materials used, activity start time, and activity stop time. Similar to the original COSTI, observers document instances of teacher demonstrations, or models, student practice, and student errors. However, the modified COSTI is used to document three types of independent practice: individual responses (i.e., a response by a single student initiated by the teacher), group responses (i.e., response by two or more students initiated by the teacher), and covert responses (i.e., a nonverbal response by one or more students where the accuracy of the response cannot be verified). The modified COSTI is also used to record two types of teacher feedback: Observers recorded immediate confirmatory and corrective academic feedback delivered by the teacher in response to student practice opportunities. Frequencies of teacher and student behaviors are coded within blocks of time, allowing calculation of a rate of student–teacher interactions for each observation, irrespective of content code. Rates reported in this study represent the number of behaviors observed per minute during a single observation, averaged across observations conducted during the year. Research on the use of the COSTI indicates the rate of student–teacher interactions can be used to predict early reading achievement (Smolkowski & Gunn, 2012).

Observation training and reliability. Observations were conducted during core reading instruction in all ECRI treatment and comparison classrooms in winter and spring. Unless schools reported classrooms were providing less than 90 min of core reading instruction, all observations were scheduled for at least 90 min. On average, reading instruction was observed for 99.86 min in ECRI treatment classrooms (range = 80.50–146.50) and 91.17 min in comparison classrooms (range = 74.50–114.50). In total, 42 first grade classrooms located in three school districts were observed two times during the school year using the COSTI instrument.

At the beginning of each observation round, independent observers received 3 days of training to use the observation protocol and were required to complete a training reliability checkout to document interrater agreement with the observation coordinator. Interrater agreement was calculated as the number of coding agreements divided by the total number of behaviors coded. Average training reliability across the year was .74. Observers were also required to complete a field reliability checkout and meet an interrater agreement standard of .80 before observing independently at each observation round. Average field reliability across the year was .94. To examine maintenance of reliability throughout each observation round, at least 20% of all observations were paired. For the paired observation, we estimated the variation in scores within and between observation occurrences using an unconditional multilevel model and computed the intraclass correlation coefficient (ICC) as the proportion of between reliability pairing variance. Based on 31 occasions in which two observers collected COSTI data simultaneously, moderate to high interobserver reliability was obtained for documenting rates of teacher models, group practice, individual practice, covert practice, errors, and feedback (ICCs = .74, .94, .62, .70, .54, and .79, respectively; Landis & Koch, 1977).

Statistical Analysis

We assessed intervention effects on each of the primary outcomes with a mixed-model (multilevel) time by condition analysis (Murray et al., 2006). This tests differences between conditions on change in outcomes from the fall of first grade (T_1) to the winter or spring of first grade (T_2). The specific model tests time coded 0 at T_1 and 1 at T_2, condition coded 0 for control and 1 for ECRI, and the interaction between the two. With only 16 schools, tests of time by condition used 14 degrees of freedom. The analyses included only students at risk for reading difficulty. To address attrition, we used a similar multilevel model.

This analysis approach included all available data—whether or not students' scores were present at both time points—to estimate differences between assessment times and between conditions. The nested time by condition analysis accounts the intraclass correlation associated with multiple students nested within the same schools. As a test of net differences, it also provides an unbiased and straightforward interpretation of the results (Cribbie & Jamieson, 2000; Fitzmaurice, Laird, & Ware, 2004).

Model estimation. We fit models to our data with SAS PROC MIXED Version 9.1 (SAS Institute, 2009) using restricted maximum likelihood, generally recommended for multilevel models (Hox, 2002). From each model, we estimated fixed effects and variance components. Maximum likelihood estimation for the time by condition analysis allows the use of all available data and provides potentially less biased results even in the face of substantial attrition, provided the missing data were missing at random (Schafer & Graham, 2002). In the present study, we did not believe that attrition or other missing data represented a meaningful departure from the missing at random assumption, meaning that missing data did not likely depend on unobserved determinants of the outcomes of interest (Little & Rubin, 2002). For example, some students were ill at the time of the posttest assessments and others transferred to a new school because their family moved.

The models assume independent and normally distributed observations. We addressed the first, more important assumption (van Belle, 2008) by explicitly modeling the multilevel nature of the data. Regression methods have also been found quite robust to violations of normality and outliers have a limited influence on the results in a variety of multilevel modeling scenarios (Bloom, Bos, & Lee, 1999; Donner & Klar, 1996; Fitzmaurice et al., 2004; Maas & Hox, 2004a, 2004b; Murray et al., 2006). This feature of multilevel models also eases concerns about the use of different scoring methods used for different measures in the analyses (e.g., raw scores, scaled scores, standard scores, etc., available for the WRMT and SAT10).

Effect sizes. To ease interpretation, we computed an effect size, Hedges's *g* (Hedges, 1981), for each fixed effect. Hedges's *g*, recommended by the What Works Clearinghouse (WWC, 2008), represents an individual-level effect size comparable to Cohen's *d* (Cohen, 1988), except that Cohen's *d* uses the sample standard deviation whereas Hedges's *g* uses the population standard deviation (Rosnow & Rosenthal, 2008).

Results

Table 2 presents means, standard deviations, and sample sizes for each measure by assessment time and condition. Next, we tested the sample for attrition effects, and differential attrition in particular. Finally, we addressed our research hypotheses with the nested time by condition model previously defined. Complete model results can be found in Table 3.

Attrition

Participant attrition, also called experimental mortality, can pose a threat to both external and internal validity of a study (Barry, 2005; Shadish, Cook, & Campbell, 2002). Student attrition was defined as students with data at T_1 but missing data at T_2.

Scores for the DIBELS measures of NWF-CLS, NWF-WRC, and ORF and for WRMT were missing for 10.5% of students, with 10.2% of the students missing data in control schools and 10.8% missing in intervention schools, and the

Table 2. Descriptive Statistics for All Measures by Time and Condition.

	Fall		Winter		Spring	
Measure	ECRI	Control	ECRI	Control	ECRI	Control
Nonsense Word Fluency, Correct Letter Sounds						
M	33.5	35.9	58.02	54.8	75.2	73.6
SD	13.2	15.8	22.3	20.1	29.1	29.0
n	116	143	115	138	107	132
Nonsense Word Fluency, Words Read Correctly						
M	7.7	7.4	14.4	10.1	19.5	14.9
SD	4.8	5.0	10.3	8.9	12.6	12.9
n	116	143	115	138	107	132
Oral Reading Fluency						
M	10.1	11.2	30.6	26.7	60.2	56.5
SD	6.8	7.5	18.1	14.7	22.2	23.8
n	116	143	115	138	107	132
SAT10 Total Reading scaled score						
M	467.0	468.4	—	—	553.0	543.7
SD	11.0	11.2	—	—	32.8	33.5
n	120	147	—	—	113	132
WRMT Basic Skills Cluster standard score (grade norm)						
M	105.4	105.0	—	—	112.8	109.3
SD	6.6	7.0	—	—	7.8	8.3
n	111	142	—	—	107	132
WRMT Total Reading Cluster standard score (grade norm)						
M	99.7	99.9	—	—	109.2	106.7
SD	7.4	7.5	—	—	7.4	7.7
n	111	142	—	—	107	132

Note. ECRI = Enhanced Core Reading Instruction; SAT10 = *Stanford Achievement Test–10th Edition*; WRMT = *Woodcock Reading Mastery Tests–Revised*.

proportion missing did not depend on condition for any of these measures ($\chi^2 = 0.03$, $df = 1$, $p = .8674$). For the SAT10, we experienced 8.2% attrition at T_2, with 10.2% of the students missing data in control schools and 5.8% missing in intervention schools ($\chi^2 = 1.67$, $df = 1$, $p = .1964$). Thus, attrition rates did not differ between conditions.

Although differential rates of attrition are undesirable, differential scores by condition on literacy measures present a far greater threat. We conducted analyses to test whether student T_1 scores on NWF, ORF, SAT10, WRMT-BS, and WRMT-TR differed by attrition, condition, or the interaction. The analysis used a mixed-model ANOVA, which nests students' T_1 scores within schools and condition. We found no statistically significant interactions between attrition and condition.

Impact on DIBELS

For students at risk for reading difficulty, we expected students exposed to ECRI to greater gains on NWF-CLS, NWF-WRC, and ORF from fall to winter and fall to spring of first grade. Table 2 presents the means and standard deviations for DIBELS measures collected in the fall, winter, and spring.

The time by condition analysis demonstrated that students at risk for reading difficulties in ECRI schools made statistically significant gains from fall to winter over those in control schools on NWF-WRC ($t = 2.55$, $df = 14$, $p = .0230$) and ORF ($t = 2.24$, $df = 14$, $p = .0415$). We found no differences between conditions, however, in gains from winter to spring on NWF-CLS, NWF-WRC, and ORF ($p = .7750$, .5982, and .5971, respectively; results not tabled). Students in ECRI schools, when compared to controls, made marginally significant gains on NWF-WRC ($t = 1.81$, $df = 14$, $p = .0919$) fall to spring but not on NWF-CLS or ORF.

Impact on SAT10 and WRMT

We next tested whether students at risk for reading difficulty performed better in ECRI schools than control schools on fall to spring gains on the SAT10 and WRMT. The analyses did not result in statistically significant findings for SAT10, WRMT-BS, or WRMT-TR. We found a marginally significant gain, however, for WRMT-BS ($t = 2.00$, $df = 14$, $p = .0651$). Although we had not hypothesized that the ECRI intervention would produce differences between conditions in comprehension, due to the focus on foundational reading skills, we tested student scores on the WRMT-RC. The difference between condition was in the expected direction but not statistically significant for WRMT-RC ($p = .3419$, $g = .28$).

Practice Opportunities and Condition

Finally, we examined whether relationship between condition and literacy outcomes was altered by the rate of independent group practice opportunities. We found an increasing difference between conditions in gains on ORF from fall to winter as the rate of group practice increased across classrooms ($t = 2.55$, $df = 14$, $p = .0230$). Intervention schools performed statistically significantly better than comparison schools when their teachers offered group practice at a rate of a rate of approximately 1.8 opportunities per minute or better, as shown in Figure 1.

Figure 1 depicts the interaction between treatment condition and independent group practice opportunities for the ORF dependent variable. The vertical axis shows the difference between conditions, intervention minus control, and the horizontal axis depicts the rate of independent group practice opportunities. The figure, then, shows the difference between conditions across the range of classrooms characterized by the rate of group opportunities for practice. The increasing slope of the line shows that as classrooms

Table 3. Estimates from Mixed-Model Time by Condition Analysis to Test Condition Effects for ECRI Schools Compared to Control Schools.

	Fall to Winter			Fall to Spring					
Effect or Statistic	NWF-CLS	NWF-WRC	ORF	NWF-CLS	NWF-WRC	ORF	SAT10	WRMT-BS	WRMT-TR
Fixed effects									
Intercept	35.90*** (3.02)	7.58*** (1.06)	11.36*** (1.44)	36.18*** (3.61)	7.58*** (1.35)	11.52*** (2.38)	468.47*** (4.34)	104.89*** (1.23)	99.74*** (1.37)
Time	18.57*** (2.28)	2.82* (1.04)	14.96*** (1.68)	36.12*** (4.31)	7.23** (1.82)	44.06*** (3.32)	73.02*** (5.62)	3.99** (1.25)	6.47*** (1.39)
Condition	–3.53 (4.41)	–0.03 (1.56)	–1.32 (2.14)	–3.70 (5.28)	–0.07 (1.99)	–1.68 (3.50)	–2.61 (6.31)	0.25 (1.80)	–0.27 (1.99)
Time × condition	5.61 (3.36)	3.93* (1.54)	5.53* (2.46)	6.36 (6.30)	4.83† (2.67)	6.92 (4.85)	14.19 (8.17)	3.65† (1.82)	3.30 (2.01)
Variances									
Residual	148.77*** (13.55)	35.60*** (3.24)	75.72*** (6.95)	301.93*** (28.14)	64.39*** (5.95)	174.92*** (16.42)	373.95*** (34.02)	22.91*** (2.20)	18.44*** (1.78)
Student	134.41*** (20.13)	16.70*** (3.53)	74.77*** (10.77)	147.29*** (30.70)	17.24** (5.39)	81.81*** (17.74)	114.60*** (31.91)	24.94*** (3.61)	27.98*** (3.59)
School intercept	44.41† (25.43)	3.67 (2.88)	1.28 (5.07)	22.72 (30.51)	0.62 (4.05)	–2.55 (13.13)	19.33 (42.99)	4.51 (3.86)	5.69 (4.77)
School gains	11.77 (7.55)	2.21 (1.66)	6.62 (4.62)	54.90† (28.89)	9.19† (5.08)	32.86† (18.22)	102.83* (50.08)	4.73† (2.45)	6.48* (3.06)
ICC	.073	.059	.080	.154	.125	.158	.216	.171	.260
Hedges's g									
Time × condition	.265	.415	.339	.219	.379	.300	.428	.452	.437

Note. Table entries are parameter estimates with standard errors in parentheses. Time was coded 0 for T_1 and 1 for T_2. Condition was coded 0 for control and 1 for ECRI. All tests fixed effects used 14 *df*. BS = Basic Skills Cluster; CLS = Correct Letter Sounds; ECRI = Enhanced Core Reading Instruction; ICC = intraclass correlation coefficient; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; SAT10 = *Stanford Achievement Test–10th Edition*; TR = Total Reading; WRC = Words Read Correctly; WRMT = *Woodcock Reading Mastery Tests–Revised*.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

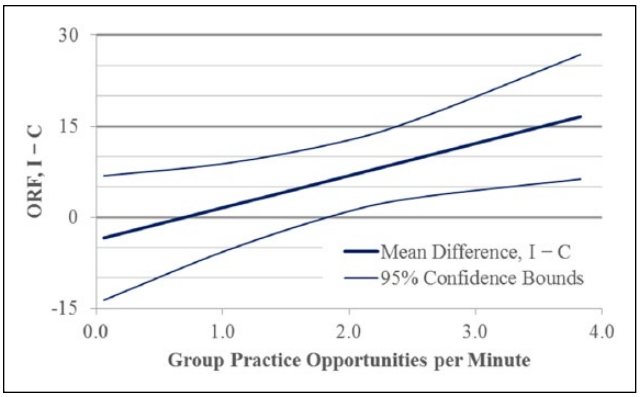

Figure 1. The interaction between treatment condition and independent group practice opportunities for Oral Reading Fluency (ORF). The vertical axis represents the difference between conditions (intervention, I, minus control, C), and the horizontal axis depicts the rate of independent group practice opportunities. The distribution of the rate of group practice opportunities included 0.1 per minute at the 5th, 1.0 at the 25th, 1.9 at the 50th, 2.6 at the 75th, and 3.8 at the 95th percentile.

increase the rate of practice they provide to students, the ECRI intervention produces greater differences. The two lighter lights show the 95% confidence bounds on the difference between conditions. At the lower end of the range, in terms of rate of practice, the 95% confidence bounds include zero (no difference between conditions), which implies that two conditions do not differ. Once the rate of practice reaches 1.8 attempts per minute, the lower of the 95% confidence bounds crosses the zero, which implies that the ECRI classrooms perform better than controls at that point. The analyses did not produce statistically significant interactions for the remaining dependent variables.

Discussion

The primary objective of the study was to evaluate the efficacy of a highly specified, tightly aligned multitiered reading intervention to improve the reading achievement of first grade at-risk readers relative to a business-as-usual RtI condition. We examined student gains for proximal and distal outcomes from fall to winter, winter to spring, and fall to spring. For students at risk for reading difficulty, we expected greater gains for ECRI intervention students on the both proximal and distal measures, as compared to students in the comparison condition. We also expected students who participated in the ECRI intervention to demonstrate greater gains on proximal outcomes, than for distal outcomes. Finally, we also examined whether the relationship between condition and literacy outcomes was altered by the rate of group practice opportunities. We hypothesized that the highly controlled and deliberate practice (Ericsson, 2008) in the treatment condition would lead to stronger fluency gains relative to less controlled practice in the comparison condition. In the following sections we summarize the results of the study, present implications, describe limitations of the study, and conclude with some next steps for future research.

The nature of the counterfactual in this study provides an interesting frame to situate the findings of this study and should guide the interpretation of the significant impacts on proximal measures as well as the nonsignificant, yet substantive effect sizes for distal outcomes. All of the schools recruited for the study were implementing a multitiered approach to reading instruction. Each district and school was identified to participate due to their ostensibly strong implementation of multitiered instruction. All of the schools had a 90-min reading block in which they implemented a core reading program, used a screening measure to identify students for Tier 2 support, and had a 30-min small group Tier 2 intervention block. In addition, due to blocking on district (and functionally on core reading program), treatment and control schools within district *were using the same core reading program*. Thus, the nature of the randomized controlled trial tested the effect of a multitiered intervention *that was highly specified, with increased explicitness and intensity, and with highly aligned and cohesive Tier 1 instruction and sustained Tier 2 intervention support relative to a control condition that was implementing the defining features of multitiered instruction and intervention*. Thus, the intervention effects were not derived from a study with a no treatment control, rather the control group was implementing a research-based, tiered system of support for their first grade students. The key difference was the degree of specificity, explicitness, alignment, and sustained Tier 2 support that composed the ECRI intervention.

Results Summary

First, we hypothesized that the at-risk readers in the ECRI intervention would demonstrate stronger gains on all outcomes, when compared to at-risk readers in the business-as-usual multitiered comparison condition. Compared to control students, we also expected stronger gains for at-risk readers in the intervention group on proximal measures relative to distal measures. Analysis of proximal measures revealed a statistically significant difference favoring intervention students on NWF-WRC and ORF fall to winter, with effect sizes of $g = +.42$, and $+.34$, respectively. There were no significant effects for NWF-WRC, NWF-CLS or for ORF fall to spring; however, there were substantive and potentially positive effect size differences favoring the treatment condition on NWF WRC and ORF fall to spring ($g = +.38$ and $+.30$, respectively). The WWC describes nonsignificant findings with effect sizes larger than $g = +.25$ as substantive and rate the effectiveness of a particular intervention as "potentially positive."

The magnitude of the effects on decoding outcomes is comparable to those found in previous studies of effective Tier 2 interventions (Ebaugh, 2000; Gunn, Biglan, Smolkowski, & Ary, 2000; Jenkins, Peyton, Sanders, & Vadasy, 2004; Lennon & Slesinski, 1999; Vadasy et al., 2005). However, the significant and substantial effect on fall to winter ORF is fairly unique and, to our knowledge, has been demonstrated in only one previous study (Gunn, Smolkowski, Biglan, & Black, 2002). It is also interesting to consider the attenuating of the fluency effect from the fall to winter to the winter to spring, and accordingly on the fall to spring analysis. Although we know the effect clearly attenuated over the course of the year, the lack of significant effect (though substantive effect size) could be the result of lower statistical power. In other words, we may have been adequately powered to detect an minimal detectable effect size as low as .42, but were not adequately powered to detect an effect size as low as .3 on the fluency construct. Another interpretation is that the efficacy of the intervention lessened as the year progressed. We believe this is a reasonable conclusion. In other words, the fall to winter finding may represent an ephemeral effect. However, an alternative explanation, and one partially supported by research, has to do with the developmental nature of fluency development. There are several recent studies that have demonstrated nonlinear growth on ORF measures, with the greatest amount of growth occurring in the fall season (Baker et al., 2008; Christ, Silberglitt, Yeo, & Cormier, 2010). This is consistent with the literature, which shows that students generally make greater gains on measures such as ORF in the fall of each grade (e.g., Fuchs et al., 1993; Hasbrouck & Tindal, 1992, 2006). Thus, it may be reasoned that the ECRI intervention was not necessarily less effective in the spring season; rather, ORF growth is typically stunted in the spring season. Even so, the question of nonsignificance remains (i.e., low power versus ephemeral effect), and further replication is warranted.

Analysis of distal measures revealed nonsignificant effects for WRMT-BS, WRMT-RC, SAT10, and WRMT-TR. However, the magnitude of the effect size on each measure is substantive and represents a "potentially positive" effect. On the WRMT-BS the impact was commensurate with the effect on foundational reading proximal measures (g = +.45). The potentially positive effect on the Reading Comprehension Cluster was smaller, but still meets the WWC criteria (g = +.28). Of note was the trend favoring the intervention group on the SAT10 and WRMT-TR scaled scores. Statistical tests indicate robust effect size differences between conditions, albeit nonsignificant, on these two distal indicators of general reading achievement, one a group administered (SAT10; g = +.43) and the other an individually administered measure (WRMT-TR; g = +.44). Although these results are promising, these estimates are not statistically significant, and should be interpreted accordingly, with extreme caution, and are preliminary in nature.

Unfortunately, our study did little to disentangle the mixed results demonstrated to date on the importance (or lack thereof) of curricular alignment across Tiers 1 and 2. Our findings are generally similar to the findings in the Chambers et al. (2011) study such that a multitiered intervention that has strong alignment between Tiers 1 and 2 appears to contribute to stronger results for at-risk readers. However, our findings contrast to the Scanlon et al. (2008) study in that in their study there were no significant differences between the condition that had a strong alignment between Tiers 1 and 2 (in the combined professional development and intervention condition) and the other conditions that did not have strong alignment between Tiers 1 and 2.

It is interesting to note that the approach that Scanlon et al. (2008) implemented substantially differed from the ECRI approach in terms of level of specificity provided to teachers and interventionists. The ECRI intervention includes carefully controlled teacher wording, a set scope and sequence of skills and concept instruction, carefully planned practice opportunities for teachers to deliver students, and specific protocols for providing corrective feedback to students. The ISA approach alternatively relies on professional development and recommendations for teachers to use a variety of techniques and strategies to work with students. There is no teacher manual or standard protocol top follow. Given the ECRI intervention increased level of specificity was utilized to increase the quantity and quality of explicit instruction, it is also likely that the ISA approach varied widely across teachers and interventionists in terms of the level of explicitness and given the lack of a standard protocol or program to follow with the ISA intervention. Thus, it appears that the combination of curriculum alignment, specificity and explicitness may have been a distinguishing factor that contributed to the range of positive and potentially positive outcomes in the current study. There is a possibility that the mixed results in the research literature on curricular alignment may be somewhat related to what is actually being aligned across tiers (i.e., degree of explicit instruction, design and delivery principles, pedagogy, content, etc.), and the amount of specificity provided to interventionists.

Our second hypothesis was related to the role of student practice in the development of fluent reading skill. Specifically, we hypothesized that the highly controlled and deliberate practice in the treatment condition (e.g., ensuring student were practicing building fluency in appropriate texts, including immediate corrective feedback) would lead to stronger fluency gains relative to less controlled practice in the comparison condition. Accordingly, we expected an interaction between treatment condition and the rate of student practice opportunities. We found interactions between

condition and rate of practice opportunities for ORF gains fall to winter. In other words, when teachers offered group practice at a rate of 1.8 opportunities per minute or higher, intervention schools performed better than comparison schools.

This interaction was not the result of the intervention classrooms scoring greater ORF scores as you moved up the rate of practice distribution, but as a result of the control classrooms demonstrating a greater decrease in ORF scores as you increased the rate of practice distribution. In other words, the more practice that control classroom conducted, the lower their ORF scores. This finding aligns with our hypothesis that practice that is not highly structured (i.e., students are not placed in appropriate text to practice fluency building and teachers are not providing immediate corrective feedback) may result in poorer learning outcomes. It is also consistent with a recently published study that indicated teachers in the ECRI intervention provided more error-free reading practice to students compared to students in the control condition (Nelson-Walker et al., 2013).

Study Implications

The findings from the current study have potential implications for a number of education stakeholder groups, though we temper our interpretation until further replication studies are conducted. We contend that the current study offers *preliminary evidence* that school and district decision makers responsible for the implementation of multitiered instruction and intervention models might consider three elements when choosing a model to support their struggling learners: the level of specificity provided to teachers and interventionists to implement the model, the level of explicitness and intervention intensity provided to students, and finally the degree of alignment between Tier 1 instruction and Tier 2 intervention materials. In the context of the current study, we manipulated the design and delivery of Tier 1 instruction to be more accessible for students that were below grade level for reading achievement. As a result, treatment classroom teachers and interventionists provided more explicit instruction, incorporated more frequent and deliberate practice opportunities, and provided more precise and immediate corrective feedback compared to control teachers (Nelson-Walker et al., 2013). However, as we revisit in the limitations section, we cannot for certain make any statement about any particular component we believe to be essential, rather we can say that when these components are brought together, these components can accelerate student learning relative to the business-as-usual multitiered control group.

Increasing the likelihood of teachers implementing research-based strategies in authentic school settings is a major goal of education leaders. Likewise, decreasing the variability of instruction practices and increasing fidelity of implementation to models of instruction and intervention is particularly difficult (Gersten, Chard, & Baker, 2000; Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). To address these issues in the context of ECRI, we developed highly specified lesson plans and teaching routines to support standard implementation of instruction and intervention materials. Our goal was to increase the level of specificity to ensure that teachers provided students with explicit and, when appropriate, intensive instructional supports (i.e., in the context of both Tier 1 *and* Tier 2). These routines provided clear expectations to teachers for what content to cover during instruction and intervention lessons and highly specified guidance for explicit and engaging teacher- student interactions. Akin to the Checklist Manifesto (Gawande, 2009), the goal of the specified routines was to increase the degree to which practitioners implement evidence-based practices with fidelity and integrity.

The approach of using highly specified instruction and intervention routines can also be used as a tool for coaches and school leaders to define and measure implementation fidelity and to provide subsequent implementation goals for teachers. It is important to note that school based personnel (rather than researchers) delivered both the Tier 1 portion and the Tier 2 portions of the model. Having school personnel as implementers, notably a unique feature of this study, increases the external validity of the study's results.

The study findings also have potential implications for publishers and developers of core reading programs and tier 2 interventions. First, in our opinion, the degree of specificity and guidance provided to teachers for delivering explicit instruction in current reading programs is lacking. Many programs do not provide enough explicit, scaffolded instruction or practice opportunities for learners at risk of reading difficulty (Gersten, 1999). Second, core program and intervention developers and publishers should strive to align instruction and intervention materials to ensure struggling students are delivered a robust and coherent tiered support plan. Most publishers of core programs also offer Tier 2, small group intervention materials. However, often different authors with competing pedagogical approaches develop them or they are very superficial in the level of intervention intensity (Baker et al., 2010). Even when the programs are philosophically aligned, the misalignment of scope and sequence of skill introduction, structured practice opportunities, and review cycles between Tier 1 and Tier 2 materials can compromise a coherent reading plan within and across each day of instruction for struggling learners.

Limitations and Future Directions for Research

One of the primary limitations of the ECRI study is that we are not able to determine the relative contribution of different features or components of the multitiered intervention on student achievement. Given the scope of the intervention,

the nature of the research design, and the funding priorities set forth in the IES Special Education Systemic Intervention grant program, we can determine and discuss only the net effect of the whole ECRI system on student achievement. This does not minimize the value or significance of the findings of the current study; however, it does raise a number of questions related to the relative value for each of the components of the system. For example, did the professional development and in-depth classroom coaching have a similar impact on student achievement as the instructional design templates that made classroom instruction more explicit and systematic? In addition, how much of a boost in Tier 2 student gains were as a result of the daily small group Tier 2 intervention? We cannot answer these and many other important questions in the context of the current study. However, follow-up component analyses or planned variation studies could help determine the relative merit of different components in the multitiered intervention.

A second limitation is related to our relatively modest sample size ($N = 16$ schools) and larger than expected school level ICC's for our outcome variables (Hedges & Hedberg, 2007). This particular point has bearing on our nonsignificant outcomes with relatively large effect sizes (i.e., range $g = +.27$ to $+.45$). A subsequent study with more power (i.e., more schools) could help us determine if these no significant effects are truly substantive in nature or merely fleeting. In addition, the schools were sampled from one state in the Pacific Northwest that has received a lot of statewide support for RtI and multitiered models. We are not sure if the positive effects generated in the current study would replicate in states where there is more or less support. A final limitation is the lack of fidelity of implementation data reported in the Tier 2 portion of the study. Although fidelity data were reported in the Tier 1 portion, a subsequent study should document the extent to which tutors delivered the intervention as intended. This limitation is also related to our inability to precisely describe what occurred in Tier 2 control settings. To address these three limitations we are currently conducting a larger replication study with three times the amount of schools and in more diverse geographical settings.

Conclusion

In summary, we found significant effects on three of the outcome variables in the area of word reading, decoding, and fluency for the ECRI intervention. The intervention also had substantive effects on reading comprehension and general reading achievement. These findings are particularly compelling given the nature of the counterfactual in that control schools were implementing the same core reading programs and a similar tiered approach to reading instruction. However, further studies that assess the relative value of intervention components and planned variations of intervention components are necessary. In addition, an examination of the potential for the replication of intervention effects in more diverse geographical settings is warranted.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A120071 to the University of Oregon. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education. Some of the authors of this article have a financial interest in the intervention materials described in the study.

References

Baker, S. K., Fien, H., & Baker, D. L. (2010). Robust reading instruction in the early grades: Conceptual and practical issues in the integration and evaluation of tier 1 and tier 2 instructional supports. *Focus on Exceptional Children*, *42*, 1–20.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Beck, C. T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*, *37*, 18–37.

Barry, A. E. (2005). How attrition impacts the internal and external validity of longitudinal research. *Journal of School Health*, *75*, 267–270. doi:10.1111/j.1746-1561.2005.tb06687.x

Bloom, H. S., Bos, J. M., & Lee, S.-W. (1999). Using cluster random assignment to measure program impacts. *Evaluation Review*, *23*, 445–469. doi:10.1177/0193841x9902300405

Bransford, J., Brown, A., & Corking, R. (1999). *How people learn: Brain, mind, experience and school*. Washington, DC: National Research Council.

Carnine, D., & Kameenui, E. (1992). *Higher order thinking: Designing curriculum for mainstreamed students* Austin, TX: PRO-ED.

Case, L. P., Speece, D. L., Silverman, R., Ritchey, K. D., Schatschneider, C., Cooper, D. H., . . . Jacobs, D. (2010). Validation of a supplemental reading intervention for first-grade children. *Journal of Learning Disabilities*, *43*, 402–417. doi:10.1177/0022219409355475

Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P., Logan, M. K., & Gifford, R. (2011). Small-group, computer-assisted tutoring to improve reading outcomes for struggling first and second grades. *Elementary School Journal*, *111*, 625–640. doi:10.1086/659035

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review*, *39*, 447–462.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Compton, D. L., Fuchs, D., Fuchs, L., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures *Journal of Educational Psychology*, *98*, 394–409. doi:10.1037/0022-0663.98.2.394

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child × instruction interactions on first graders' literacy development. *Child Development*, *80*, 77–100. doi:10.1111/j.1467-8624.2008.01247.x

Coyne, M. D., Kame'enui, E. J., & Carnine, D. W. (2011). *Effective teaching strategies that accommodate diverse learners*. Upper Saddle River, NJ: Merrill.

Cribbie, R. A., & Jamieson, J. (2000). Structural equation models and the regression bias for measuring correlates of change. *Educational and Psychological Measurement*, *60*, 893–907. doi:10.1177/00131640021970970

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children*, *19*, 1–16.

Donner, A., & Klar, N. (1996). Statistical considerations in the design and analysis of community intervention trials. *Journal of Clinical Epidemiology*, *49*, 435–439. doi:10.1016/0895-4356(95)00511-0

Ebaugh, J. C. (2000). The effects of fluency instruction on the literacy development of at-risk first graders (Doctoral dissertation, Fordham University, 2000). *Dissertation Abstracts International*, *61*(06A), 0072.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, *15*, 988–994. doi:10.1111/j.1553-2712.2008.00227.x

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Hoboken, NJ: John Wiley.

Fuchs, D., Compton, D., Fuchs, L., Bryant, J., & Davis, G. (2008). Making "secondary intervention" work in a three-tier responsiveness-to intervention model: Findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading and Writing*, *21*, 413–436. doi:10.1007/s11145-007-9083-9

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*, 27–48.

Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York, NY: Metropolitan Books.

Gersten, R. (1999). Lost opportunities: Challenges confronting four teachers of English-language learners. *Elementary School Journal*, *100*, 37–56.

Gersten, R. M., Chard, D., & Baker, S. K. (2000). Factors enhancing sustained use of research-based instructional practices: A historical perspective on relevant research. *Journal of Learning Disabilities*, *33*, 444–457.

Gersten, R. M., 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: A practice guide* (NCE 2009-4045). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Good, R. H., & Kaminski, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Education Achievement.

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice*, *15*, 198–205. doi:10.1207/sldrp1504_4

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. *Journal of Special Education*, *34*, 90–104. doi:10.1177/002246690003400204

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. *Journal of Special Education*, *36*, 69–79. doi:10.1177/00224669020360020201

Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and non-Hispanic students. *Journal of Special Education*, *39*, 66–85. doi:10.1177/00224669050390020301

Harcourt Educational Measurement. (2002). *Stanford Achievement Test*. San Antonio, TX: Author.

Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2-5. *Teaching Exceptional Children, 24*(3), 41–44.

Hasbrouck, J., & Tindal, G. A. (2006) Oral reading fluency norms: A valuable assessment tool for reading teachers. The Reading Teacher, *59*(7), 636–644.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107–128. doi:10.3102/10769986006002107

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*, 60–87. doi:10.3102/0162373707299706

Hill, D. R., King, S. A., Lemons, C. J., & Partanen, J. N. (2012). Fidelity of implementation and instructional alignment in response to intervention research. *Learning Disabilities Research & Practice*, *27*, 116–124. doi:10.1111/j.1540-5826.2012.00357.x

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

Institute of Education Sciences. (2008). *RFA: Request for applications: Special education research grants. CFDA Number: 84.324A*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Jenkins, J. R., Peyton, J., Sanders, E. A., & Vadasy, P. (2004). Effects of reading decodable texts in supplemental first-grade tutoring. *Scientific Studies of Reading*, *8*, 53–85. doi:10.1207/s1532799xssr0801_4

Kame'enui, E. J., Simmons, D. C., & Coyne, M. D. D. (2000). Schools as host environments: Toward a school-wide

reading improvement model. *Annals of Dyslexia*, *50*, 31–52. doi:10.1007/s11881-000-0016-4

Kame'enui, E. J., Simmons, D., Coyne, M., & Harn, B. (2003). *Institute on beginning reading: Day 1*. Eugene, OR: IDEA.

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

Kerins, M. R., Trotter, D., & Schoenbrodt, L. (2010). Effects of tier 2 intervention on literacy measures: Lessons learned. *Child Language Teaching and Therapy*, *26*, 287–302.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

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.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: John Wiley.

Loftus, S. M., Coyne, M. D., McCoach, D. B., Zipoli, R., & Pullen, P. C. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice*, *25*, 124–136. doi:10.1111/j.1540-5826.2010.00310.x

Maas, C. J. M., & Hox, J. J. (2004a). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, *46*, 427–440. doi:10.1016/j.csda.2003.08.006

Maas, C. J. M., & Hox, J. J. (2004b). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *58*, 127–137. doi:10.1046/j.0039-0402.2003.00252.x

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, *40*, 148–182. doi:10.1598/rrq.40.2.2

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research & Practice*, *16*, 230–239. doi:10.1111/0938-8982.00023

McMaster, K. L., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*, 445–463.

Murray, D. M., Hannan, P. J., Pals, S. P., McCowen, R. G., Baker, W. L., & Blitstein, J. L. (2006). A comparison of permutation and mixed-model regression methods for the analysis of simulated data in the context of a group-randomized trial. *Statistics in Medicine*, *25*, 375–388. doi:10.1002/sim.2233

National Assessment of Educational Progress. (2011). *The nation's report card: Reading 2011* (NCES 2012457). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

National Joint Committee on Learning Disabilities. (2005). *Responsiveness to intervention and learning disabilities: A report prepared by the National Joint Committee on Learning Disabilities representing eleven national and international organizations*. Retrieved from http://www.ldonline.org/about/partners/njcld

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Pub. No. 00-4769). Washington, DC: National Institute of Child Health and Human Development. Retrieved from http://www.nichd.nih.gov/publications

National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academies Press.

Nelson-Walker, N. J., Fien, H., Kosty, D. B., Smolkowski, K., Smith, J. L. M., & Baker, S. K. (2013). Evaluating the effects of a systematic intervention on first-grade teachers' explicit reading instruction. *Learning Disabilities Quarterly*, *36*, 215–230.

O'Connor, R. E., Harty, K. R., & Fulmer, D. (2005). Tiers of intervention in kindergarten through third grade. *Journal of Learning Disabilities*, *38*, 532–538. doi:10.1177/00222194050380060901

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, *3*, 159–197. doi:10.1207/s1532799xssr0302_4

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383. doi:10.1080/10888430701530730

Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice*, *25*, 110–123. doi:10.1111/j.1540-5826.2010.00309.x

Rosnow, R., & Rosenthal, R. (2008). Assessing the effect size of outcome research. In A. Nezu & C. Nezu (Eds.), *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions* (pp. 379–401). New York, NY: Oxford University Press.

SAS Institute. (2009). *SAS/STAT 9.2 user's guide*. Cary, NC: Author.

Scanlon, D., Gelzheiser, L., Vellutino, F., Schatschneider, C., & Sweeney, J. (2008). Reducing the incidence of early reading difficulties: Professional development for classroom teachers versus direct interventions for children. *Learning and Individual Differences*, *18*, 346–359. doi:10.1016/j.lindif.2008.05.002

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–119). Timonium, MD: York Press.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037//1082-989X.7.2.147

Schuele, C. M., Justice, L. M., Cabell, S. Q., Knighton, K., Kingery, B., & Lee, M. W. (2008). Field-based evaluation of two-tiered instruction for enhancing kindergarten phonological awareness. *Early Education and Development*, *19*, 726–752. doi:10.1080/10409280802375299

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York, NY: Guilford.

Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the Classroom Observations of Student–Teacher Interactions (COSTI) for kindergarten reading instruction. *Early Childhood Research Quarterly*, *27*, 316–328.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academies Press.

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education*, *36*, 223–233. doi:10.1177/002246690303600403

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, *15*, 55–64. doi:10.1207/SLDRP1501_6

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring. *Journal of Learning Disabilities*, *38*, 364–380. doi:10.1177/00222194050380041401

van Belle, G. (2008). *Statistical rules of thumb* (2nd ed.). Hoboken, NJ: John Wiley.

Vaughn, S., Wanzek, J., Woodruff, A. L., & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Haager, J. Klingner, & S. Vaughn (Eds.), *Evidence-based reading practices for response to intervention* (pp. 11–27). Baltimore, MD: Brookes.

Western Regional Reading First Technical Assistance Center. (2005a). *Grade 1 lesson maps to accompany professional development for use of Houghton Mifflin reading in schools with high risk students*. Eugene, OR: Author.

Western Regional Reading First Technical Assistance Center. (2005b). *Instructional routine templates 1–17*. Eugene, OR: Author.

What Works Clearinghouse. (2008). *What Works Clearinghouse: Procedures and standards handbook* (Version 2.0). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests–Revised*. Circle Pines, MN: American Guidance Service.