EXAMINING THE EFFICACY OF A MULTITIERED INTERVENTION FOR AT-RISK READERS IN GRADE 1

ABSTRACT

This study reports the results of a cluster RCT evaluating the impact of Enhanced Core Reading Instruction on reading achievement of grade 1 at-risk readers. Forty-four elementary schools, blocked by district, were randomly assigned to condition. In both conditions, at-risk readers received 90 minutes of whole-group instruction (Tier 1) plus an additional 30 minutes of daily, small-group intervention (Tier 2). In the treatment condition, Tier 1 instruction included enhancements to the core program and Tier 2 intervention was highly aligned with the core program. In the comparison condition, Tier 1 instruction used the same core program as treatment schools in the district and Tier 2 intervention followed standard district protocol. Significant treatment effects were found on measures of phonemic decoding and oral reading fluency from fall to winter and word reading from fall to spring. Student- and classroom-level variables predicted student response to instruction differentially by condition.

Jean Louise M. Smith
Nancy J. Nelson
UNIVERSITY OF OREGON

Keith Smolkowski
OREGON RESEARCH INSTITUTE

Scott K. Baker
SOUTHERN METHODIST UNIVERSITY

Hank Fien
UNIVERSITY OF OREGON

Derek Kosty
OREGON RESEARCH INSTITUTE

MULTITIERED systems of reading support represent a comprehensive approach to early intervention designed to prevent reading difficulties and increase the reading achievement of all students (Baker, Fien, & Baker, 2010; Fletcher & Vaughn, 2009; Gersten et al., 2009; Greenwood,

THE ELEMENTARY SCHOOL JOURNAL
Volume 116, Number 4. Published online May 5, 2016
 0013-5984/2016/11604-0002$10.00

Kratochwill, & Clements, 2008). These types of support systems have been part of several state-level implementation models in U.S. schools, including Oregon's Effective Behavior and Instructional Support Systems initiative (EBISS; http://www.ode.state.or.us/search/page/?id=1389) and Michigan's Integrated Behavior and Learning Support Initiative (MiBLSi; http://miblsi.cenmi.org/). Despite wide implementation, however, there have been few evaluations of the impact of these multitiered systems on student achievement (Burns, Appleton, & Stehouwer, 2005; Fien et al., 2015; Glover, 2010; Speece & Walker, 2007). The purpose of this study is to report the results of a large, cluster randomized controlled trial evaluating the impact of Enhanced Core Reading Instruction (ECRI) on the reading achievement of at-risk readers in grade 1 across a diverse sample of schools and to explore predictors of differential response to the intervention.

Overview of the ECRI Intervention

The phrase "Enhanced Core Reading Instruction" refers to the use of prioritized content and teaching routines designed to increase the quality of explicit instruction provided in Tier 1 settings, plus core-aligned small-group instruction in Tier 2 (Baker et al., 2010; Carnine, Silbert, & Kame'enui, 1997; Kame'enui, Carnine, Dixon, Simmons, & Coyne, 2002). A distinguishing feature of ECRI is the inclusion of both (*a*) Tier 1 enhanced core reading instruction and (*b*) a supplemental Tier 2 small-group intervention for students at risk for reading difficulty. Therefore, ECRI is a "multitiered intervention" as both Tier 1 and Tier 2 are designed to make the instructional content of the core reading program more explicit with much more instructional scaffolding included to be accessible to a broader range of students. The ECRI Tier 1 enhancements prioritized instructional delivery of content related to beginning reading skills, including phonemic awareness, phonics, word reading, reading fluency, vocabulary, and comprehension. Students who were identified as being at risk for reading difficulty based on fall initial reading skill were provided with the ECRI Tier 2 intervention that embedded structured opportunities to preview and practice foundational skills, including phonemic awareness, phonics, word reading, and reading fluency, aligned with Tier 1 instruction, in a 30-minute, small-group lesson. A full description of the ECRI Tier 1 and Tier 2 multitiered intervention is provided in the Method section.

Evidence for the Utility of Multitiered Systems of Reading Support

Multitiered systems of reading support most often include three tiers of support that increase in intensity to match the magnitude of student need (Baker et al., 2010; Chard et al., 2008; Fletcher & Vaughn, 2009; Gersten et al., 2009). Tier 1 occurs in general education contexts as core instruction and is intended to (*a*) enable the vast majority of students to meet end-of-year goals and (*b*) provide students at moderate risk for academic difficulty with the first of two essential tiers of support (Baker et al., 2010). Tier 2 interventions provide an avenue to deliver intensified instruction for students who need additional support (e.g., reteaching of concepts,

intensified explicit instruction, additional practice with teacher feedback) to attain academic goals. Tier 3 is characterized by highly explicit, systematic instruction, delivered to students in small-group formats or individually. Multitiered systems require the coordination of multiple components, including evidence-based instruction and intervention, formative assessment that includes universal screening and progress monitoring, professional development targeting evidence-based practices, and ongoing evaluation of the effectiveness of the system for continuous improvement (Chard et al., 2008; Harn, Chard, Biancarosa, & Kame'enui, 2011). Because multitiered systems of reading support infuse evidence-based practices to strategically support the needs of all students, they have significant practical value to educational systems. However, specific methods and implementation strategies to integrate singular components into a cohesive system of support are not well established.

Although there is abundant research support for individual components of multitiered systems of reading support, few studies have examined the impact of multitiered systems of reading support on outcomes of at-risk readers (Hill, King, Lemons, & Partanen, 2012). Some recent studies have examined Tier 2 intervention and simultaneously described the nature of Tier 1 instruction (e.g., Mathes et al., 2005; O'Connor, Harty, & Fulmer, 2005), but only a few studies have evaluated the combination of Tier 1 and Tier 2 support (see Chambers et al., 2011; Scanlon, Gelzheiser, Vellutino, Schatschneider, and Sweeney, 2008; Wonder-McDowell, Reutzel, & Smith, 2011). For example, Scanlon et al. (2008) observed a reduction in the percent of kindergarten students at risk for reading difficulties at the end of the school year based on whether they participated in a Tier 2 intervention in kindergarten or if their Tier 1 teachers participated in a professional development intervention, or both, and compared these students to those who were not yet exposed to these conditions. Students in the combined professional development and intervention condition performed significantly better on spelling and word recognition at the beginning of grade 1 compared to students in the reading-intervention-only condition. Chambers et al. (2011) contrasted a Tier 2, computer-assisted, small-group tutoring intervention with one-to-one tutoring in reading among students also receiving the Success for All Tier 1 core reading program. The computer-assisted Tier 2 program, which was more closely aligned with the Tier 1 program, was more effective for increasing grade 1 reading achievement. Similar to Chambers et al. (2011), Wonder-McDowell et al. (2011) demonstrated consistent benefits in grade 2 students' decoding, fluency, and comprehension outcomes by aligning the sequence and pacing of skills, strategies, and concepts taught in Tier 2 intervention with the sequence and pacing of skills, strategies, and concepts taught in (Tier 1) core reading instruction. The results of these studies support efforts to improve the quality of Tier 1 instruction and align Tier 2 interventions with Tier 1 objectives in the context of a multitiered approach to reading instruction.

The ECRI study differs from these studies in three important ways. First, in the ECRI study, schools are the unit of random assignment. This is important because most multilevel intervention models are conceptualized as a school-level approach. Second, in the ECRI treatment condition, the experimental manipulation of treatment material occurred in both Tier 1 and Tier 2, which allows for an examination of impact on at-risk readers under conditions that take into account the integrated

nature of Tier 1 and Tier 2 in a multitiered approach to intervention support. This approach contrasts with other studies that have evaluated the combined effects of Tier 1 and Tier 2 by holding Tier 1 instruction constant, which thereby tests the additive effects of Tier 2. Third, studies that have examined the combined impact of Tier 1 and Tier 2 have rarely reported effect sizes on student outcomes, limiting our ability to ascertain the relative impact of the intervention.

The current study examines the impact of a Tier 2 intervention that is closely aligned to the daily and weekly lesson objectives of an enhanced Tier 1 core program, compared to standard-practice multitiered instruction for students at risk of reading difficulties. Previous research on the ECRI multitiered intervention introduces positive results for the effect of the intervention on teacher and student outcomes. Nelson-Walker et al. (2013) demonstrated that ECRI had a positive impact on teachers' use of explicit instructional strategies ($g = 1.31$) and instructional intensity (e.g., increases in group practice opportunities; $g = 1.63$) in the Tier 1 component of the multitiered intervention. In terms of effects for at-risk students, Fien et al. (2015) presented preliminary evidence for a positive impact of participation in ECRI on fall-to-winter gains in decoding and oral reading fluency (effect sizes ranging from .34 to .42), and nonsignificant effects with trends favoring the treatment group, on fall-to-spring gains in oral reading fluency and total reading achievement (effect sizes of .30 and .43, respectively). Further, Baker, Smolkowski, Chaparro, Smith, and Fien (2015) used regression discontinuity analysis to compare differences between students who received Tier 2 intervention in addition to Tier 1 and comparable peers who received only Tier 1. There was a significant break in the regression line at the cut score indicating that students close to the cut score who received Tier 1 and Tier 2 performed better on reading achievement measures than students close to the cut score who received only Tier 1. The present article extends this research by comparing reading outcomes for students in schools that received the ECRI multitiered intervention to students in comparison schools that implemented their standard-practice, multitiered system approach.

Predictors of Differential Response to the ECRI Intervention

In the present study, we were also interested in exploring predictors of differential response to the ECRI intervention. Tests of moderation, or predictor × condition interactions, determine for whom or in which contexts the intervention works best (Jaccard & Turrisi, 2003; Judd & Kenny, 1981; Yoder & Compton, 2004). We selected student-level and classroom-level predictors of differential response to the ECRI intervention. Previous experiments in beginning reading have demonstrated that preskills are strong predictors of differential response (see Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009; Fletcher et al., 2011; Gunn, Smolkowski, & Vadasy, 2011; Mathes et al., 2005). Thus, we explored whether initial levels of early literacy skills, such as phonemic decoding, oral reading fluency, vocabulary knowledge, and overall beginning reading skill, interacted with participation in the ECRI treatment condition, with its strong emphasis on foundational reading skills, to differentially benefit reading outcomes depending on students' skill at intervention entry. Further, based on evidence that certain classroom-level variables are associated

with improvements in student achievement (see Moats, 2009; Nye, Hedges, & Konstantopoulos, 2000; Nelson, Benner, & Gonzalez, 2003), we were interested in exploring whether contextual variables, including quality of explicit instruction and rate of group practice, level of risk of the class (as indicated by number of at-risk readers per class), classroom management and instructional support, years teaching, teacher knowledge, and minutes of total reading instruction, influenced the efficacy of the ECRI intervention.

Purpose

The aims of this study were (*a*) to evaluate the effect of enhanced Tier 1 instruction in combination with a Tier 2 intervention that is closely aligned to the daily and weekly lesson objectives of the core program, relative to districts' standard implementation of multitiered intervention, on the reading achievement of grade 1 students at risk for reading difficulties, and (*b*) to explore predictors of differential response to the ECRI multitiered intervention. We hypothesized that at-risk readers receiving ECRI treatment would outperform students in the comparison condition on a range of reading outcomes. To meet our second aim, we explored student and contextual characteristics (moderators) that we hypothesized would predict differential response to ECRI to better understand who benefits most from the intervention.

Method

This study was designed to examine the efficacy of the Enhanced Core Reading Instruction (ECRI) multitiered intervention in grade 1 classrooms using a cluster randomized controlled trial that nested students and teachers within schools. In this study, we report results from two waves of schools ($N = 44$) that implemented the ECRI intervention for one year.

Recruitment and Assignment Procedures

In the spring of the year prior to intervention implementation, school districts implementing a multitiered service-delivery model for reading instruction were recruited to participate in the ECRI project. Eligible schools (*a*) used a published, comprehensive core reading program identified and adopted through standard district procedures during a 90-minute reading block for Tier 1 and (*b*) provided students identified for Tier 2 with an additional 30 minutes of small-group instruction per day.

In Wave 1, we recruited 22 schools in three Oregon school districts to participate. Four of these schools elected to not participate in the study due to changes in school leadership between recruitment and the beginning of the study. The remaining 18 schools were randomly assigned to the treatment or comparison conditions. We blocked on district in condition assignment to control for core curricula and other important factors, increasing the likelihood of including similar schools

in each condition. One district had only one participating school, and that school joined another district for randomization. After random assignment, two schools (one treatment and one comparison school) left the project, leaving 16 schools in Wave 1. In Wave 2, we recruited 20 schools in three districts in Oregon and eight schools in three districts in Massachusetts to participate, and conducted the random assignment process as we did with Wave 1 schools. All recruited Wave 2 schools participated in the study; thus, the combined wave sample included 44 schools—22 treatment schools and 22 comparison schools.

Tier assignment. Students enrolled in grade 1 classrooms ($n = 3{,}466$; treatment schools, $n = 1{,}703$; comparison schools, $n = 1{,}763$) were recruited to participate in the project. Of these students, we obtained parental consent for 3,377 students (97%; 1,655 in treatment schools and 1,712 in comparison schools). Fall scores on the reading portion of the Stanford Achievement Test, 10th edition (SAT-10) were used to assign students to Tier 1, Tier 2, or Tier 3. Students who scored at or above the 10th percentile and at or below the 30th percentile were assigned to Tier 2. Students who scored above the 30th percentile on the SAT-10 were assigned to Tier 1; students below the 10th percentile were assigned to Tier 3. Norms from the 2007 edition of the SAT-10 were used to determine percentiles.

Across the two waves, 3,022 students participated in the fall assessment (1,517 in the treatment condition and 1,505 in the comparison condition). Of the students who took the SAT-10 in the fall, 16 (8 treatment, 8 comparison) did not finish the assessment due to absences and are not included in the analytic sample. In total, 1,634 students were assigned to Tier 1 ($n = 818$ treatment, $n = 816$ comparison), 811 students were assigned to Tier 2 ($n = 394$ treatment, $n = 417$ comparison), and 561 students were assigned to Tier 3 ($n = 297$ treatment, $n = 264$ comparison).

Participants

In the 44 elementary schools (22 per condition), a total of 142 grade 1 reading teachers participated in this study—70 in the treatment condition and 72 in the comparison condition. Across conditions, participating teachers reported 13.49 total years of teaching experience, on average ($M = 12.18$ years for treatment teachers; $M = 14.73$ years for comparison teachers). Most teachers (97%) were female. The analytic sample comprised 811 students at risk for reading difficulty and assigned to Tier 2 (394 treatment, 417 comparison). Schools reported that 7.4% of these students received special education services (6.6% treatment, 8.2% comparison) and 19.7% were English learners (EL: 23.7% treatment, 15.8% comparison). According to the 2009–2010 National Center for Educational Statistics (NCES, 2009), the average percentage of grade 1 Hispanic students across participating schools was 18.9% (19.8% treatment, 18.0% comparison) and the average percentage of grade 1 African American students was 3.6% (4.3% treatment, 3.0% comparison). Also according to NCES, the average percentage of grade 1 students eligible to receive free and reduced-price lunch was 46.7% (48.6% treatment, 44.8% comparison). Independent sample *t*-tests at the school level revealed no statistically significant differences between conditions on demographic variables.

Implementation

Treatment and comparison teachers provided daily reading instruction during a 90-minute reading block in Tier 1 using one of several widely used, published core reading programs that was identified and adopted through standard district procedures. Because districts implemented core reading programs and group assignment was blocked by district, core program implementation was divided equally across conditions. Students identified for Tier 2 received an additional 30 minutes of small-group reading intervention. Tier 3 was not a focus of the project. Schools followed standard district procedures to implement Tier 3 intervention in both the treatment and comparison conditions. We know schools followed standard district procedures for Tier 3 intervention, but we do not know the extent to which Tier 3 aligned with core instruction.

Treatment condition. The ECRI intervention was designed to increase the quality of instructional interactions between teachers and students by prioritizing academic content in activities and increasing the explicitness of instruction. The intervention supported implementation through (*a*) Tier 1 enhanced core reading instruction, (*b*) Tier 2 small-group intervention that was highly aligned with Tier 1 instruction, and (*c*) professional development and coaching. For a description of each of these activities, see Fien et al. (2015).

In ECRI treatment classrooms, core reading instruction was enhanced by prioritizing critical reading content and implementing explicit teaching routines (Baker et al., 2010). Core program activities related to vocabulary, comprehension, reading fluency, phonics, and phonemic awareness were restructured to emphasize the following instructional elements: (*a*) clear learning objectives; (*b*) increased modeling of key content through visual models, verbal directions, and fully developed, clear explanations; (*c*) explicit connections between new and previous content; (*d*) increased opportunities for guided and independent practice; and (*e*) deliberate and carefully designed review of previous content (Carnine & Kame'enui, 1992; Coyne, Kame'enui, & Carnine, 2011). The instructional teaching routines supplied teachers with language for explicitly modeling content, providing frequent choral practice opportunities, and delivering immediate feedback (Baker et al., 2010).

The Tier 2 intervention was highly aligned with core reading program content. Students were identified for Tier 2 support based on fall screening (SAT-10) and participated in the intervention for the year (36 weeks). Students receiving Tier 2 intervention were pretaught content in daily, 30-minute lessons in small groups of three to five students. For the purposes of this study, progress-monitoring data were used to regroup students within Tier 2 support during the year, but not across tiers. Overall, 67 small-group instructors participated in the study in the treatment condition. Five of these instructors were Title 1 teachers or reading specialists and 62 were instructional assistants. The content of small-group lessons emphasized foundational reading skills including phonemic awareness, word reading, and fluency in reading connected text. The scope and sequence of foundational skills content varied slightly depending on the core program. Each 30-minute lesson included seven activities: irregular word reading, phonemic awareness, sound spelling introduction and review, blending and word reading, accuracy and fluency reading decodable text, encoding practice, and reteaching of challenging words.

Teachers in the treatment condition participated in three days of professional development focused on the ECRI instructional model prior to the beginning of the school year, and two days of follow-up activities in October. 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.

Small-group instructors received two days of professional development in the fall and one day in January. Professional development for small-group instructors emphasized instructional teaching routines and strategies for increasing student engagement in lessons. All professional development included teaching demonstrations by ECRI expert coaches and participant practice with coach feedback. Teachers and small-group instructors also received comprehensive coaching support through classroom and small-group visits conducted once per month and regular study-group meetings, facilitated by an ECRI coach.

Comparison condition. Comparison schools used the standard, adopted core reading program for Tier 1 instruction. Comparison teachers reported that Tier 1 classrooms spent an average of 50.0 ($SD = 29.3$) minutes in whole-group instruction, 36.7 ($SD = 24.2$) minutes in small-group instruction, and 29.2 ($SD = 15.7$) minutes in independent work. Comparison schools used the districts' standard practice to deliver Tier 2 small-group intervention (i.e., they did not use the ECRI Tier 2 intervention materials), which may have included regrouping students across tiers based on progress-monitoring data. Teachers in the comparison condition reported that Tier 2 intervention materials included a range of supplemental and intervention items, including published, standardized protocol intervention materials as well as teacher-developed materials. In the comparison condition, 67% reported that someone in their school provided them with instructional support, guidance, and coaching about teaching students to read.

Fidelity of implementation. Observations of implementation fidelity conducted by trained data collectors using a standardized protocol across conditions indicated that nearly all treatment teachers used ECRI intervention materials during instruction ($M = 0.89$, $SD = 0.16$), where 1.0 indicates a perfect score. As expected, observations conducted in comparison classrooms indicated that comparison teachers rarely used ECRI intervention materials during instruction across the year ($M = 0.08$, $SD = 0.14$). Although treatment diffusion across the year was minimal, some comparison teachers did have access to intervention materials. In addition, in treatment classrooms, the mean score for quality of explicit instruction was 0.77 ($SD = 0.15$), while it was 0.51 ($SD = 0.17$) in comparison classrooms. Thus, teachers in ECRI treatment classrooms were rated as providing higher-quality instruction, on average, compared to teachers in comparison classrooms (Nelson-Walker et al., 2013). In the ECRI treatment classrooms, overall fidelity of implementation was 0.83 ($SD = 0.15$).

Data Collection

Outcome measures included curriculum-based assessments of foundational reading skills and comprehensive measures of reading achievement. Districts pro-

vided student-level demographic data, including special education status, gender, and limited English proficiency (LEP) status each spring. Class size, school size, and number of at-risk readers per grade 1 classroom were derived from tier assignment files and based on student participation in the fall screening assessment. Classroom teachers reported the average number of daily instructional minutes for Tier 2 instruction.

Student assessments. Prior to fall assessment, assessors attended three days of training targeting administration and scoring procedures. Across the winter and spring, assessors attended four days of additional assessment training. For individually administered student measures, assessment coordinators evaluated interrater agreement, reported below, by shadow scoring with assessors and providing feedback on test administration.

SAT-10. The SAT-10 (Harcourt Educational Measurement, 2002) is a group-administered, norm-referenced test of reading proficiency, which took from 110 to 155 minutes to administer. Trained data collectors administered the SESAT 2 in the fall (September) and the Primary 1 in the spring (April/May) of grade 1 to all participating students. SESAT 2 subtests administered included Sounds and Letters, Word Reading, and Sentence Reading; Primary 1 subtests administered included Word Study Skills, Word Reading, Sentence Reading, and Reading Comprehension. The test manual indicates that Kuder-Richardson reliability coefficients, a measure of internal consistency, are .94 for SESAT 2 and .97 for Primary 1. SESAT 2 and Primary 1 Total Reading scores are moderately correlated with Otis-Lennon School Ability Test ($r = .68$ and .61, respectively). For the analyses, we used scaled scores, derived from grade-based norms, for the Total, Word Reading, and Sentence Reading scores. Sample quartiles (25th, 50th, and 75th percentiles) were 530, 546, and 564 for the SAT-10 Total; 505, 530, and 558 for Word Reading; and 526, 553, and 582 for Sentence Reading.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). DIBELS assessments were used to measure phonemic decoding skill and passage-reading fluency. Nonsense Word Fluency (NWF) and Oral Reading Fluency (ORF) measures were administered in the fall (October), winter (January), and spring (April/May). Trained data collectors administered DIBELS assessments in all districts. Average interrater agreement between data collectors was 97.6% (range = 92%–100%) for NWF across the study; average interrater agreement for ORF was 98.2% (range = 90%–100%) across the study.

NWF. NWF (Kaminski & Good, 1996) is an individually administered, one-minute, timed measure of student skill in reading consonant-vowel and consonant-vowel-consonant pseudo-words. The score for Words Recoded Completely and Correctly (NWF-WRC) was obtained by counting the number of nonwords students recoded accurately.[1] Alternate-form reliability ranged from .67 to .80, and concurrent validity 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, one-minute, timed measure of student skill in accurately and fluently reading connected text. The number of words read correctly in one minute is the student's score on a single passage. To determine a student's benchmark score, the assessor administers three grade-level passages at a single time point during the school year (be-

ginning, middle, or end) and records the median score. In the beginning grades, 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 Test—revised (WRMT). The WRMT (Woodcock, 1998) is a standardized, comprehensive battery of tests that measure multiple aspects of reading ability, including comprehension, word recognition, and word analysis. We administered Form H for two subtests in the fall (October) and spring (May): Word Identification and Word Attack. As reported in the testing manual, the correlation between Total Reading scores on the WRMT and the Woodcock-Johnson Psycho-Educational Battery is .88 at grade 1. Internal consistency for the subtests ranges from .94 to .98 in grade 1. Average interrater agreement between data collectors across the study was 97.5% (range = 93%–100%).

Peabody Picture Vocabulary Test, fourth edition (PPVT-4). The PPVT-4 (Dunn & Dunn, 2007) is an individually administered, norm-referenced test of receptive vocabulary. Each test item consists of four pictures arranged on the page, and students select the picture that best illustrates the word provided by the examiner. We administered Form A of the assessment in the fall (October). As reported in the technical manual, test-retest reliability is .92 for ages 5–6 and .93 for ages 7–10, and concurrent correlations with the Expressive Vocabulary Test (Williams, 2007) are .84 for ages 5–6 and .82 for ages 7–10. Average interrater agreement between data collectors across the study was 99.6% (range = 90%–100%).

Comprehensive Test of Phonological Processing (CTOPP). The CTOPP (Wagner, Torgesen, & Rachotte, 1999) is a measure of phonological awareness, phonological memory, and rapid naming. We administered the subtests associated with the Rapid Naming (RN) and Phonological Memory (PM) Composite scores, using the form for students ages 5–6 in the fall (October) of grade 1. As reported in the testing manual, test-retest reliability for the RN Composite is reported as .70 for ages 5–7 (Wagner et al., 1999). The PM and RN Composites in grade 1 demonstrate moderate to strong correlations with the WRMT in grade 2 (.52 and .70, respectively; Wagner et al., 1999). Average interrater agreement between data collectors across the study was 98.5% (range = 94%–100%).

Beliefs about learning to read. We adapted two existing measures (Motivations for Reading Scale: Baker & Scher, 2002; Implicit Theories measures: Heyman & Dweck, 1998) to assess students' motivation for reading and their thinking about traits related to academic performance. The existing measures were abbreviated and revised to be appropriate for use in grade 1, combined into a single measure, and changed from self-response to an interview. The interview included 16 questions that probed enjoyment, value of reading, and perceived competence (e.g., "Do you like to read?" and "Do you think people can learn new things from books?") and six questions that addressed students' thinking about implicit theories related to behavior and academic ability (e.g., "Imagine a [boy or girl] in your class who is not good at reading. If [he or she] works extra hard with [his or her] teacher and does [his or her] homework, could [he or she] be good at reading by the end of the school year?" Analyses used five subscales of the revised measure: (*a*) reading enjoyment, (*b*) reading value, (*c*) perceived reading competence, (*d*) implicit theory of behavior, and (*e*) implicit theory of academic ability. Average interrater agreement across the study was 99.7% (range = 95%–100%).

Teacher surveys. Teachers completed an online survey in the spring (May) to assess their knowledge of beginning reading instruction and espoused instructional practices in reading across the school year. Completion rate for the surveys was 85% ($n = 121$ teachers).

Teacher Knowledge Survey (TKS). The TKS (see Teaching Reading Essentials survey; Moats, 2006) assesses teacher knowledge of reading concepts (e.g., "How many spoken syllables are in the word *rhythm*?") and reading instructional practices (e.g., "True or false: In whole word blending, all the phonemes of a word are articulated in order and then blended in sequence."). The survey contains 17 multiple-choice questions and 18 true-false questions. Two items from the original TKS were adapted for ECRI administration to explore teacher knowledge related to key features of the intervention. We used the number of correct responses for analysis.

Instructional Practices Survey (IPS). The IPS was used to gather information about teacher demographics and experience. On the IPS, teachers are also asked to self-report the number of instructional minutes provided in reading, formats used for instruction (e.g., grouping, "walk to read" models), curriculum implementation, use of data-based decision making, participation in professional development, and coaching and feedback received. The full survey contains 15 items and uses a variety of formats (e.g., multiple choice, yes-no, open-ended). A total score was not obtained for the survey; individual item responses were used to describe the study sample and denote fidelity of implementation.

Classroom observations. Observations were conducted during core reading instruction in all treatment and comparison classrooms in the fall (November), winter (February), and spring (April), using three observation instruments: the Classroom Observations of Student Teacher Interactions (COSTI; Smolkowski & Gunn, 2012), Ratings of Classroom Management and Instructional Support (RCMIS; Doabler & Nelson-Walker, 2009), and the Quality of Explicit Instruction (QEI; Nelson-Walker et al., 2013) scale. Unless schools reported classrooms were providing less than 90 minutes of core reading instruction, all observations were scheduled for at least 90 minutes. On average, reading instruction was observed for 85.3 minutes in ECRI treatment classrooms ($SD = 27.4$, range = 15.0–139.5) and 83.1 minutes in comparison classrooms ($SD = 20.3$, range = 22.0–114.5). In total, 142 grade 1 classrooms located in nine school districts were observed three times during the school year.

COSTI. An adapted version of the COSTI (Smolkowski & Gunn, 2012) was used in the study to record contextual information and quantify student-teacher interactions during instructional activities. Similar to the original COSTI instrument, the modified COSTI can be used to record teacher demonstrations, independent student practice, student errors, and teacher corrections. The modified COSTI, however, distinguishes three types of teacher-initiated independent practice: individual responses (i.e., a single student's response), group responses (i.e., a response by two or more students), and covert responses (i.e., nonverbal responses by one or more students where the accuracy of the response could not be verified). The modified COSTI also distinguishes confirmatory from corrective teacher feedback. Research on the use of the COSTI indicates it is a valid and reliable measure of teacher-initiated behavior in early reading (Nelson-Walker et al., 2013; Smolkowski & Gunn, 2012).

RCMIS. The RCMIS (Doabler & Nelson-Walker, 2009) contains 11 four-point, Likert-type items that are used to rate features of instructional quality: classroom management, delivery of instruction, and the learning environment (Doabler et al., 2015; Nelson-Walker, 2010). Predictive validity for reading and mathematics ranged from .26 to .42 (Doabler et al., 2015).

QEI. The QEI measures fidelity of ECRI implementation and the quality of explicit reading instruction in grade 1 classrooms with 11 items scored on a three-part scale (0 = no, 0.5 = partially, 1 = yes). The scale includes items related to access and use of intervention materials (2 items), adherence to principles of high-quality explicit instruction for critical reading skill areas (6 items), completion of intervention lesson activities (2 items), and overall quality of intervention implementation (1 item). Studies of the QEI demonstrate that it can be used reliably across observation occasions and raters to document fidelity of implementation of the ECRI intervention. Nelson-Walker et al. (2013) observed an intraclass correlation coefficient (ICC) for the QEI of .71 between observation occasions and .86 across raters, which provides estimates of the stability of instruction and interobserver reliability, respectively.

Observer training and reliability. At the beginning of each observation round, observers received three 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. In the fall, only the RCMIS and QEI instruments were used in training and observations. In the winter and spring, the RCMIS, QEI, and COSTI instruments were used in training and observations. Average QEI and RCMIS field reliability was 97% (SD = 2.8%), while average COSTI field reliability was 93% (SD = 2.4%).

To examine maintenance of reliability during each observation round, at least 20% of all observations were paired. For these paired observations, we estimated the variation in scores within and between observation occurrences using an unconditional multilevel model and computed the ICC as the proportion of between-reliability pairing variance (Landis & Koch, 1977). Based on 92 occasions in which two observers collected QEI or RCMIS data simultaneously, and 69 occasions in which two observers collected COSTI data simultaneously, moderate to high interobserver reliability was obtained. Specifically, the ICC was .78 for the QEI and RCMIS, and ICCs for documenting rates of teacher models, group practice, individual practice, covert practice, errors, and feedback were .81, .96, .80, .86, .66, and .80, respectively.

Statistical Analysis

We assessed intervention effects on each of the primary outcomes with a mixed-model (multilevel) time × condition analysis (Murray, 1998) to account for the intraclass correlation associated with students nested within schools, the level of random assignment. The analysis tests net differences between conditions on change in outcomes from the fall (T1) to spring (T2) of grade 1, with gains for individual students clustered within schools. The test of net differences provides an unbiased and straightforward interpretation of the results (Cribbie & Jamieson, 2000; Fitzmaurice, Laird, & Ware, 2004). The statistical model includes time, con-

dition, and the time × condition interaction, with time coded 0 at T1 and 1 at T2, and condition coded 0 for control and 1 for ECRI. With 44 schools, tests of time × condition used 42 degrees of freedom. Analyses included the students in each school who scored at or above the 10th percentile and below the 30th percentile on the SAT-10 in the fall of grade 1 (i.e., students assigned to Tier 2).

The secondary aim required an analysis to explore differential response due to various student-level and classroom-level variables. We expanded the model to test interaction effects. The statistical model included a predictor and its interaction with condition, time, and the time × condition term, resulting in a three-way interaction, all corresponding two-way interactions, and individual (conditional) effects. The three-way interaction of the predictor, time, and condition provides an estimate of whether condition effects vary by the predictor. The analysis included dichotomous and continuous predictors, and we used continuous variables whenever possible.

Model estimation. We fit models to our data with SAS PROC MIXED version 9.2 (SAS Institute, 2009) using restricted maximum likelihood, and included all available data, whether or not students' scores were present at both time points. Maximum-likelihood estimation with all available data produces potentially unbiased 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 likely did not depend on unobserved determinants of the outcomes of interest (Little & Rubin, 2002). Most missing data involved students who were absent on the day of assessment or transferred to a new school.

The models assume independent and normally distributed observations. We addressed the first assumption (van Belle, 2008) by explicitly modeling the multilevel nature of the data. Regression methods have 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; Hannan & Murray, 1996; Murray et al., 2006). Murray et al. (2006) showed that violations of normality at either or both the individual and group levels do not bias results as long as the study is balanced at the group level.

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, 2011), represents an individual-level effect size comparable to Cohen's *d* (Cohen, 1988).

Results

Table 1 provides descriptive statistics (*M, SD, N*) for all primary measures used for tests of the impact of the ECRI intervention. Below we test for differences in attrition by condition and differential scores for students missing data by condition. We next test the efficacy of ECRI and then explore the potential for differential response to condition by student and group or teacher characteristics.

Table 1. Descriptive Statistics for Primary Outcome Measures

	Fall		Winter		Spring	
	ECRI	Control	ECRI	Control	ECRI	Control
NWF-WRC:						
M	8.8	8.6	16.8	13.4	21.6	19.2
(*SD*)	(5.0)	(5.3)	(10.0)	(9.1)	(13.9)	(13.8)
N	379	402	371	392	355	388
ORF:						
M	10.8	11.3	29.9	25.8	55.2	53.6
(*SD*)	(6.9)	(8.0)	(17.0)	(16.1)	(22.9)	(23.4)
N	379	402	371	393	355	388
WRMT Word ID:						
M	16.2	16.3	–	–	44.7	43.2
(*SD*)	(6.9)	(7.4)	–	–	(9.5)	(9.6)
N	365	402	–	–	348	381
WRMT Word Attack:						
M	7.5	7.7	–	–	21.0	19.3
(*SD*)	(4.3)	(4.5)	–	–	(7.6)	(7.6)
N	365	402	–	–	348	381
SAT-10 Total Reading:						
M	468.3	469.1	–	–	546.9	544.4
(*SD*)	(11.1)	(11.5)	–	–	(29.3)	(29.2)
N	394	417	–	–	362	385
SAT-10 Word Reading:						
M	433.4	433.4	–	–	539.3	530.7
(*SD*)	(19.9)	(19.9)	–	–	(45.4)	(46.0)
N	394	417	–	–	362	387
SAT-10 Sentence Reading:						
M	455.7	459.3	–	–	554.0	553.8
(*SD*)	(25.1)	(27.6)	–	–	(40.0)	(36.8)
N	394	417	–	–	363	386

Attrition

Student attrition was defined as students with data at T_1 but missing data at T_2, and we examined attrition with respect to the Tier 2 sample of 811 students, 417 in comparison schools and 394 in ECRI schools. For DIBELS data, we experienced 8.4% attrition at T_2, with 29 students missing T_2 data in comparison schools and 39 students missing T_2 data in ECRI schools ($\chi^2(1) = 2.29$, $p = .1306$). SAT-10 scores were missing for 7.5% of students at T_2, with 30 students missing T_2 data in comparison schools and 31 students missing T_2 data in ECRI schools ($\chi^2(1) = 0.13$, $p = .7161$). Students were missing WRMT scores at T_2 for 10.1% of the sample, including 36 students in comparison schools and 46 students in ECRI schools ($\chi^2(1) = 2.06$, $p = .1509$). Given that none of these differences were statistically significant, we conclude that attrition rates did not differ between conditions.

Although differential rates of attrition are undesirable, differential scores on literacy tests by condition present a greater threat to validity (Barry, 2005). We conducted an analysis to test whether student scores were differentially affected by attrition across conditions. We examined the effects of condition, attrition status, and the interaction between the two on pretest scores within a mixed-model analysis of variance (Murray, 1998), which nests students' T_1 scores within schools and

condition. We tested scores for NWF-WRC, ORF, SAT-10 Total Reading Score, SAT-10 Word Reading, SAT-10 Sentence Reading, WRMT Word Identification, and WRMT Word Attack. We found no evidence of differential attrition for any of our dependent variables: $p > .2400$ for all tests.

ECRI Efficacy

We tested the hypothesis that students in ECRI classrooms would perform better than those in comparison classrooms with 11 measures. We first examined differences between conditions in gains on NWF-WRC and ORF from fall to winter and report the results in Table 2. Students in schools that implemented ECRI outperformed students in comparison schools on all three measures, with effect sizes (g) ranging from .21 to .30. We also report the ICC for gains as described by Murray (1998, see p. 301).

Table 2. Results from Mixed-Model Time × Condition Analysis for Tests of Condition Effects on Fall-to-Winter Gains in Student Achievement

Effect or Statistic	NWF-WRC	ORF
Fixed effects:		
Intercept	8.64****	11.39****
	(.77)	(1.09)
Time	4.88****	14.24****
	(.77)	(1.24)
Condition	−.10	−.77
	(1.09)	(1.56)
Time × condition	2.87*	4.22*
	(1.10)	(1.76)
Variances:		
Residual	31.18****	65.04****
	(1.64)	(3.46)
Student	17.87****	83.78****
	(1.94)	(6.36)
School intercept	5.33*	4.71
	(2.28)	(4.54)
School gains	4.63**	12.74***
	(1.41)	(3.83)
ICC (ρ):		
School gains	.129	.164
Hedges's g:		
Time × condition	.301	.255
p-values:		
Time × condition	.0128	.0214

Note.—Table entries show parameter estimates with standard errors in parentheses except for intraclass correlations (ICCs), Hedges's g values, and p-values. Tests of fixed effects (first four rows) used 42 df to account for the school as the unit of analysis. NWF = Nonsense Word Fluency; WRC = Words Read Correctly; ORF = Oral Reading Fluency; ICC = intraclass correlation coefficient. ICC calculated as per Murray (1998, p. 301).

* $p < .05$.

** $p < .01$.

*** $p < .001$.

**** $p < .0001$.

Next we examined differences between conditions in gains from fall to spring with NWF-WRC, ORF, WRMT Word ID and Word Attack, and SAT-10 Total Reading, Word Reading, and Sentence Reading. Table 3 presents the results for these eight models. Students in ECRI schools outpaced their peers in comparison schools on Word Attack and Word Reading, with marginal gains favoring ECRI schools on Word ID. All differences favored the ECRI condition, and effect sizes (g) ranged from .12 for ORF to .32 for Word Attack. The effect sizes for Word Reading and Word ID were both .24, but the analysis did not produce statistically significant differences for Word ID due to the larger ICC.

To aid interpretation, we describe SAT-10 Word Reading results in terms of percentiles. The difference between conditions represents four percentile points; control students were at the 18th percentile at posttest while ECRI students scored

Table 3. Results from Mixed-Model Time × Condition Analysis for Tests of Condition Effects on Fall-to-Spring Gains in Student Achievement

Effect or Statistic	DIBELS NWF-WRC	DIBELS ORF	WRMT Word ID	WRMT Word Attack	SAT-10 Total Reading	SAT-10 Word Reading	SAT-10 Sentence Reading
Fixed effects:							
Intercept	8.58****	11.40****	16.17****	7.81****	469.26****	433.55****	459.53****
	(.95)	(1.41)	(.74)	(.52)	(1.97)	(2.63)	(2.53)
Time	9.87****	41.38****	26.62****	11.17****	73.48****	95.87****	92.83****
	(1.14)	(1.86)	(.85)	(.64)	(2.66)	(3.64)	(3.83)
Condition	−.01	−.82	−.39	−.45	−1.69	−1.02	−4.62
	(1.35)	(2.02)	(1.06)	(.74)	(2.81)	(3.76)	(3.62)
Time × condition	2.64	2.79	2.24+	2.42*	6.12	10.76*	6.98
	(1.63)	(2.65)	(1.22)	(.91)	(3.80)	(5.22)	(5.47)
Variances:							
Residual	70.71****	178.52****	33.42****	23.05****	325.43****	1000.60****	865.97****
	(3.76)	(9.59)	(1.83)	(1.25)	(16.97)	(52.18)	(45.63)
Student	21.54****	86.13****	29.63****	11.39****	94.28****	125.80**	124.24***
	(3.55)	(10.48)	(2.65)	(1.38)	(15.67)	(41.57)	(37.35)
School intercept	4.15	1.66	2.27	.82	2.96	1.39	−23.97
	(3.08)	(7.09)	(2.03)	(.96)	(13.46)	(23.87)	(22.46)
School gains	9.93**	26.59**	5.86**	2.98**	57.90**	85.98**	108.41**
	(3.01)	(8.56)	(1.79)	(.99)	(17.52)	(32.44)	(35.35)
ICC (ρ):							
School gains	.123	.130	.149	.114	.151	.079	.111
Hedges's g:							
Time × condition	.191	.120	.235	.318	.209	.235	.182
p-values:							
Time × condition	.1140	.2998	.0743	.0112	.1154	.0455	.2093

Note.—Table entries show parameter estimates with standard errors in parentheses except for intraclass correlations (ICCs), Hedges's g values, and p-values. Tests of fixed effects (first four rows) used 42 df to account for the school as the unit of analysis. NWF = Nonsense Word Fluency; WRC = Words Read Correctly; ORF = Oral Reading Fluency; SAT-10 = Stanford Achievement Test (10th ed.); WRMT = Woodcock Reading Mastery Test; ID = identification; ICC = intraclass correlation coefficient. ICC calculated as per Murray (1998, p. 301).

+ $p < .10$.

** $p < .05$.

** $p < .01$.

*** $p < .001$.

**** $p < .0001$.

at the 22nd percentile. In the comparison condition, 42% of students reached or exceeded the 30th normative percentile on Word Reading, while 55% of ECRI students reached or exceeded the 30th percentile. In terms of meeting or exceeding the 40th percentile on the SAT-10, 32% of comparison students met this standard and 43% of ECRI students met this standard.

Differential Response to ECRI

The tests of differential response included a predictor and its interactions with time, condition, and time × condition. The predictor × time × condition term indicates differential response to treatment. For example, students appeared to respond to the ECRI intervention differentially by the quality of explicit instruction (i.e., $t = 3.50$, $df = 40$, $p = .0012$). Students in classrooms with higher-quality explicit instruction experienced greater benefit from the intervention than those exposed to poorer-quality explicit instruction. Figure 1 depicts the differential response. The vertical axis shows the *difference* between conditions on gains in SAT-10 Sentence Reading. The horizontal axis shows the quality of explicit instruction. The heavy line shows the difference between conditions at each level of explicit instruction

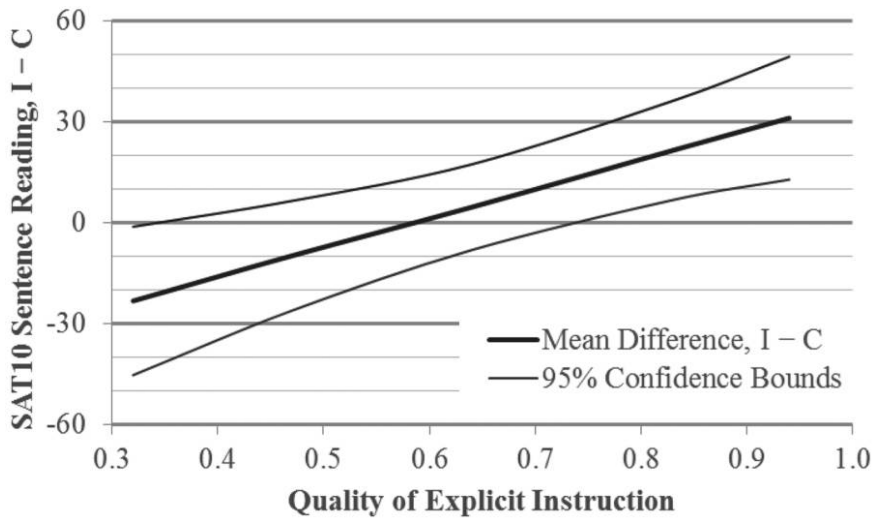

Figure 1. Differences in student gains between conditions plotted by teachers' quality of explicit instruction, measured on a 0 to 1 scale. The heavy line depicts the estimate of the difference between conditions (I–C) on gains in SAT-10 Sentence Reading across the range of the measure of explicit instruction quality, from the 5th to the 95th sample percentile (.32 to .94). The two thin lines show the 95% confidence interval around the mean estimate. Students in classrooms where with poorer quality of explicit instruction (left end of the chart) did not differ between conditions because the 95% confidence bounds include zero. The mean crosses zero, implying a positive difference between conditions, at .58 and becomes statistically significantly different from zero at .74 or more, where the confidence bounds exclude zero. This indicates that the distribution of student gains in sentence reading across the year were not similar across condition in classrooms with high-quality explicit instruction (.74 or higher).

quality. At an explicit instruction quality rating of .32 (5th percentile), ECRI students performed 23 points worse than control students, and with an explicit instruction quality rating of .94 (95th percentile), ECRI students outperformed their peers in control schools by 31. The two thin lines show the 95% confidence bounds, and the difference between conditions becomes statistically significant at .74, where the lower 95% confidence line crosses zero on the difference between conditions.

For each outcome measure, we tested differential response for eight pretest measures of student performance: NWF-WRC, ORF, PPVT, WRMT Word ID, WRMT Word Attack, SAT-10 Total Score, CTOPP Rapid Naming Composite, and the CTOPP Phonological Memory Composite. Next we tested eight other student characteristics: reading enjoyment, reading value, perceived reading competence, implicit theory of academic ability, implicit theory of behavior, LEP status, special education status, and gender. Finally, we tested 11 group or teacher characteristics, including the quality of explicit instruction (QEI), specific ECRI fidelity (QEI), classroom management and instructional support (RCMIS), rate of group practice opportunities, number of years teaching, teacher knowledge of reading concepts, minutes of daily Tier 2 reading instruction, minutes of daily total reading instruction, the number of at-risk readers per class, class size, and school size. Because these exploratory analyses involved 27 predictors of 10 outcome measures, or 270 total tests, we have a 95% chance of between 7 and 21 Type I errors (false positives). We will therefore discuss only patterns of results, which we emphasize must be interpreted cautiously and in an exploratory manner given the number of tests involved.

For two predictors we found statistically significant interactions with condition for two or more dependent variables. These predictors were pretest ORF and quality of explicit instruction. Pretest ORF produced statistically significant interactions with condition for NWF-WRC and ORF from fall to winter. Students with higher ORF scores in the fall made greater gains on ORF and NWF-WRC by winter, but the differential response was not maintained for fall-to-spring gains on these or other measures. We found differential response for the quality of explicit instruction on SAT-10 Sentence Reading (Fig. 1) and Total Reading. For both dependent variables, students in ECRI schools made greater gains than those in comparison schools when the quality of explicit instruction was high. We also found differential response by (*a*) SAT-10 Total Score on ORF fall to winter, (*b*) implicit theory of academic ability on SAT-10 Word Reading, (*c*) number of at-risk readers per class on ORF from fall to spring, (*d*) classroom management and instructional support on SAT-10 Sentence Reading, (*e*) rate of group practice on SAT-10 Sentence Reading, (*f*) number of years teaching on SAT-10 Word Reading, and (*g*) minutes of total daily Tier 2 reading instruction on NWF-WRC from fall to spring.

Given the large number of tests, few predictors suggested the possibility of a differential response. Just 11 of the 270 tests (4.1%) were statistically significant with $\alpha = .05$, which is just below the expected Type I error rate. With a Bonferroni correction, we would set $\alpha = .00019$ (.05/270). No individual test met this level of significance. Thus our analyses were unable to offer clear evidence of differential response to intervention condition.

Discussion

The primary aim of this study was to evaluate the impact of a highly specified and aligned Tier 1 and Tier 2 multitiered intervention, Enhanced Core Reading Instruction (ECRI), on reading outcomes in grade 1 for students at risk for reading difficulty. The multitiered intervention supported implementation through Tier 1 enhanced core reading instruction, Tier 2 small-group intervention that was highly aligned with Tier 1 instruction, and comprehensive professional development. We expected at-risk readers in the ECRI intervention to make stronger gains on reading outcome measures, when compared to at-risk readers in participating districts' standard-practice comparison condition, which was also multitiered. We also expected to observe stronger gains for at-risk readers in the intervention group on more proximal reading outcome measures (e.g., DIBELS) relative to more distal, comprehensive measures of reading achievement (e.g., SAT-10).

Summary of Findings and Implications for Research and Practice

ECRI impact. Mixed-model time x condition analyses reveal a statistically significant difference favoring treatment students for NWF-WRC and ORF ($g = .30$ and .26, respectively) from fall to winter. Although the fall-to-spring difference was not significant, the trend favored the treatment condition. On more distal measures of reading performance, at-risk readers in the ECRI treatment condition demonstrated larger gains on WRMT Word Attack ($g = .32$) and SAT-10 Word Reading ($g = .24$), relative to at-risk readers in comparison schools. No statistically significant effects for WRMT Word ID, SAT-10 Sentence Reading, or SAT-10 Total Reading were observed. One interpretation of this pattern of findings is that because the ECRI Tier 2 intervention focused heavily on beginning word-reading skills early in the school year, fall-to-winter gains on very proximal measures (e.g., NWF) were more pronounced for treatment students during this timeframe, with differences between treatment and comparison on these measures diminishing as the focus of the intervention expanded to include greater emphasis on reading comprehension and vocabulary later in the school year.

At the end of the year, treatment students demonstrated improvements in early literacy skills (e.g., SAT-10 Word Reading), but significant differences were not realized for measures of reading achievement that focused on vocabulary and comprehension. The findings of overall positive trends for the ECRI intervention for beginning reading skills are similar to those observed in other evaluations of interventions that align Tier 2 supports with Tier 1 instruction for at-risk readers (see Chambers et al. 2011; Scanlon et al., 2008; Wonder-McDowell et al., 2011).

Differential response to the ECRI intervention. In this study, we also explored differential response to the intervention using student- and classroom-level predictors identified in the literature as having an impact on student learning (e.g., Fletcher et al., 2011; Nelson et al., 2003). Our exploratory analysis included a total of 270 tests, with evidence of differential response to intervention for 4.1% of the tests. This mirrors the estimated Type I error rate (5%), and no tests would have been statistically significant with a conservative Bonferroni correction. Although

our differential response results should be interpreted with caution, we think two patterns of results deserve more discussion and possibly further study.

We observed differential effects using ORF pretest as a predictor of response to intervention. Among students at risk for reading difficulty, those who began the year with lower scores on ORF responded less well to the ECRI intervention in the first half of the year compared to students with higher pretest scores on ORF. The differential effects for students with lower versus higher ORF scores could be a function of the content focus of the Tier 1 and Tier 2 materials. It is possible that students with lower ORF scores at the beginning of first grade, those near the 10th percentile, had insufficient beginning reading skills to benefit from the content in the grade 1 core program and were in need of more intensive instructional support (e.g., increased instructional time, additional practice opportunities with careful corrective feedback, systematically sequenced content addressing students' knowledge and skill gaps). This pattern also resembles the "Matthew effect," in which student achievement fans across time. Although there has been mixed support for the Matthew effect in the research literature (see Shaywitz et al., 1995; Sideridis, 2011; Stanovich, 1986), this study provides support for this pattern within the context of an intervention for at-risk readers in grade 1 and implies that lower-scoring students require additional support. Because students at the lower end of the distribution made less progress than students at the higher end, initial skill levels may have usefully selected students who would benefit from more intensive intervention (Tier 3) instead of waiting for a demonstrated lack of response to Tier 2 instruction to make that determination.

We also observed a pattern of differential effects for quality of explicit instruction on SAT-10 Sentence Reading and Total Reading. Students in ECRI treatment schools made significantly greater gains than those in comparison schools when the quality of explicit instruction was high (see Fig. 1). Despite using a comprehensive professional development approach, in treatment classrooms average quality of explicit instruction (0.77; $SD = 0.15$)—a distinguishing feature of the ECRI intervention—may not have reached optimal levels. Future ECRI intervention studies could examine how initial program training, coaching, and other methods of teacher support might be used to increase the quality of explicit instruction, and how increases in the quality of explicit instruction might differentially impact student achievement.

Limitations

Because ECRI is a multitiered intervention, we are not able parse the relative contributions of intervention components on student achievement. For instance, we are not able to examine the effects of Tier 2 intervention independent from the effects of Tier 1 instruction. These trade-offs in intervention research are inevitable, and in this case raise future research opportunities to examine the relative value of the components that comprise the ECRI multitiered intervention, including Tier 1 instruction, Tier 2 support, alignment between tiers of support, and professional development and coaching protocols. A second, related limitation of this study is the lack of information on the nature of the alignment of Tier 1 instruction

and Tier 2 intervention in the comparison condition, which limits our ability to analyze the specific benefit associated with tier alignment across conditions.

A third limitation is the number of predictors included in our analyses of differential response across treatment and comparison conditions. We chose to explore a wide array of potential predictors of differential response. Our analyses produced few statistically significant results, which serves to help eliminate potential predictors more than confirm their potential role in the effects of ECRI. However, patterns of differential effects and support from the literature give us reason to believe that some predictors of differential response identified in this study may have merit. For example, students in the treatment condition whose teachers demonstrated greater teacher knowledge and delivered higher-quality explicit instruction made greater gains on the SAT-10 across the school year.

Conclusion

In the present study, we tested the efficacy for at-risk readers of the ECRI multitiered intervention, which carefully aligned Tier 1 and Tier 2 content and intensified the instruction within the Tier 2 intervention. We attained support for our research hypothesis that participation in the ECRI intervention resulted in significant increases in reading achievement for at-risk readers in grade 1. Students at risk for reading problems who were provided with the ECRI intervention achieved greater gains on reading measures of decoding and word reading compared to similarly at-risk peers in comparison schools. These findings are particularly compelling given the nature of the comparison condition, where comparison schools implemented the district-adopted core reading program and a multitiered service-delivery model.

The ECRI multitiered intervention (enhanced Tier 1 with a highly aligned Tier 2 intervention) represents one model of a cohesive system of intervention support. Although the design of this study does not allow us to isolate the features of the intervention that were most effective, we believe the evidence from this study suggests that the unique combination of features used in the ECRI intervention should be considered during the design of multitiered systems of support for at-risk readers, including (*a*) high specificity provided to teachers and interventionists to implement the model, (*b*) highly explicit and intensive instruction during Tier 1 instruction and Tier 2 intervention, and (*c*) alignment between Tier 1 instruction and Tier 2 intervention content and materials. Additionally, the study provides evidence that multitiered interventions can be used to provide explicit instruction in Tier 1 and Tier 2 to improve reading outcomes for students at risk for reading difficulties in grade 1.

Notes

This research was supported in part by Project ECRI, grant no. R324A090104, IES, U.S. Department of Education. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Scott K. Baker, Hank Fien, and Jean Louise M. Smith are eligible to receive a portion of royalties from the University of Oregon's future dis-

tribution and licensing of certain ECRI-based works. Potential conflicts of interest are managed through the University of Oregon's Research Compliance Services. Jean Louise M. Smith is a research assistant professor at the University of Oregon, Nancy J. Nelson is a research assistant professor at the University of Oregon and director of the CTL Reading Clinic, Keith Smolkowski is a research scientist at Oregon Research Institute, Scott K. Baker is a research professor at the Center on Teaching and Learning at Southern Methodist University, Hank Fien is an associate professor in school psychology and director of the Center on Teaching and Learning at the University of Oregon, and Derek Kosty is a senior data analyst and an associate research scientist at Oregon Research Institute. Address all correspondence to Jean Louise M. Smith, Center on Teaching and Learning, University of Oregon, Eugene, OR 97403; e-mail: jmercier@uoregon.edu.

1. NWF can also be scored for Correct Letter Sounds (NWF-CLS), which obtains the number of correct letter sounds that the students provided regardless of their production of complete pseudo-words. NWF-CLS correlated with NWF-WRC .64 in the fall, .79 in the winter, and .85 in the spring for students selected for Tier 2 intervention. The smaller correlation in the fall is likely due to the restricted range of Tier 2 student performance in the fall. In subsequent samples comprising scores from grade 1 students across all tiers from the same schools, NWF-CLS correlated with NWF-WRC between .91 and .95 ($N = 1{,}267$ to 2,273).

References

Baker, L., & Scher, D. (2002). Beginning readers' motivation for reading in relation to parental beliefs and home reading experiences. *Reading Psychology*, **23**, 239–269. doi:10.1080/713775283

Baker, S. K., Fien, H., & Baker, D. (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**(9), 1–20.

Baker, S. K., Smolkowski, K., Chaparro, E. A., Smith, J., & Fien, H. (2015). Using regression discontinuity to test the impact of a tier 2 reading intervention in first grade. *Journal of Research on Educational Effectiveness*, **8**, 218–244. doi:10.1080/19345747.2014.909548

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.1746-1561.2005.00035.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

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment*, **23**, 381–394. doi:10.1177/073428290502300406

Carnine, D. W., & Kame'enui, E. J. (Eds.). (1992). *Higher order thinking: Designing curriculum for mainstreamed students*. Austin, TX: Pro-Ed.

Carnine, D. W., Silbert, J., & Kame'enui, E. J. (1997). *Direct instruction reading* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

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 graders. *Elementary School Journal*, **111**, 625–640. doi:10.1086/659035

Chard, D. J., Harn, B., Sugai, G., Horner, R., Simmons, D. C., & Kame'enui, E. J. (2008). Core features of multi-tiered systems of reading and behavioral support. In C. R. Greenwood, T. R. Kratochwill, & M. Clements (Eds.), *Schoolwide prevention models: Lessons learned in elementary schools* (pp. 87–114). New York: Guilford.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

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

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *Elementary School Journal*, **110**(1), 1–18. doi:10.1086/598840

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

Doabler, C. T., Baker, S. K., Kosty, D., Smolkowski, K., Clarke, B., Miller, S. J., & Fien, H. (2015). Examining the association between explicit mathematics instruction and student mathematics achievement. *Elementary School Journal*, **115**(3), 303–333.

Doabler, C., & Nelson-Walker, N. (2009). *Ratings of classroom management and instructional support*. Unpublished observation instrument, Center on Teaching and Learning, College of Education, University of Oregon, Eugene.

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

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test—IV* (4th ed.). Upper Saddle River, NJ: Pearson.

Fien, H., Smith, J., Smolkowksi, K., Baker, S. K., Nelson-Walker, N. J., & Chaparro, E. (2015). An examination of the efficacy of a multitiered intervention on early reading outcomes for first grade students at risk for reading difficulties. *Journal of Learning Disabilities*, **48**, 602–621.

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

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review*, **40**, 3–22.

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, **3**, 30–37. doi:10.1111/j.1750-8606.2008.00072.x

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*. (No. NCE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Glover, T. A. (2010). Key RTI service delivery components: Considerations for research-informed practice. In T. A. Glover & S. Vaughn (Eds.), *The promise of response to intervention: Evaluating current science and practice* (pp. 7–22). New York: Guilford.

Good, R. H., & Kaminski, R. A. (2002). *DIBELS Oral Reading Fluency passages for first through third grades*. (Technical Report No. 10). Eugene: University of Oregon.

Greenwood, C., Kratochwill, T., & Clements, M. (2008). *Schoolwide prevention models: Lessons learned in elementary schools*. New York: Guilford.

Gunn, B., Smolkowski, K., & Vadasy, P. (2011). Evaluating the effectiveness of Read Well Kindergarten. *Journal of Research on Educational Effectiveness*, **4**, 53–86. doi:10.1080/19345747.2010.488716

Hannan, P. J., & Murray, D. M. (1996). Gauss or Bernoulli? A Monte Carlo comparison of the performance of the linear mixed-model and the logistic mixed-model analyses in simulated community trials with a dichotomous outcome variable at the individual level. *Evaluation Review*, **20**, 338–352.

Harcourt Educational Measurement. (2002). *Stanford Achievement Test (SAT-10)*. San Antonio, TX: Harcourt.

Harn, B. A., Chard, D. J., Biancarosa, G., & Kame'enui, E. J. (2011). Coordinating instructional supports to accelerate at-risk first-grade readers' performance. *Elementary School Journal*, **112**, 332–355. doi:10.1086/661997

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

Heyman, G. D., & Dweck, C. S. (1998). Children's thinking about traits: Implications for judgments of the self and others. *Child Development*, **69**, 391–403. doi:10.1111/j.1467-8624.1998.tb06197.x

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

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage.

Judd, C. M., & Kenny, D. A. (1981). *Estimating the effect of social interventions*. New York: Cambridge University Press.

Kame'enui, E. J., Carnine, D., Dixon, R. C., Simmons, D., & Coyne, M. (2002). *Effective teaching strategies that accommodate diverse learners* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice-Hall.

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

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

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

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.

Moats, L. (2006). *Teacher reading essentials*. Longmont, CO: Sopris West Educational Services.

Moats, L. (2009). Knowledge foundations for teaching reading and spelling. *Reading and Writing*, **22**, 379–399. doi:10.1007/s11145-009-9162-1

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

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 Center for Education Statistics. (2009). *The nation's report card: National assessment of educational progress at grade 4, 8, and 12* (Report No. NCES 2011-451). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Nelson, J. R., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice*, **18**, 255–267. doi:10.1111/1540-5826.00080

Nelson-Walker, N. J. (2010). *Enhanced core reading instruction classroom observation manual*. Available from the Center on Teaching and Learning at the University of Oregon, Eugene.

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.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal*, **37**, 123–151. doi:10.3102/00028312037001123

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.

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

Scanlon, D. M., Gelzheiser, L. M., Vellutino, F. R., Schatschneider, C., & Sweeney, J. M. (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.

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

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly*, **30**, 894–906.

Sideridis, G. D. (2011). Exploring the presence of Matthew effects in learning disabilities. *Journal of Learning Disabilities*, **44**, 399–401. doi:10.1177/0022219411410041

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.

Speece, D. L., & Walker, C. Y. (2007). What are the issues in response to intervention research? In D. Haager, J. K. Klingner, & S. Vaughn (Eds.), *Evidence-based reading practices for response to intervention* (pp. 287–301). Baltimore, MD: Brookes.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, **21**, 360–406. doi:10.1598/RRQ.21.4.]

van Belle, G. (2008). *Statistical rules of thumb* (2nd ed.). New York: Wiley.

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

What Works Clearinghouse. (2011). *What Works Clearinghouse: Procedures and standards handbook (Version 2.1)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Williams, K. T. (2007). *EVT-2: Expressive Vocabulary Test*. New York: Pearson Assessments.

Wonder-McDowell, C., Reutzel, D. R., & Smith, J. A. (2011). Does instructional alignment matter? *Elementary School Journal*, **112**, 259–279. doi:10.1086/661524

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

Yoder, P., & Compton, D. (2004). Identifying predictors of treatment response. *Mental Retardation and Developmental Disabilities Research Reviews*, **10**, 162–168. doi:10.1002/mrdd.20013