Article

Hammill Institute on Disabilities

Learning Disability Quarterly
36(4) 215–230

DOI: 10.1177/0731948712472186
ldq.sagepub.com
SAGE

Evaluating the Effects of a Systemic Intervention on First-Grade Teachers' Explicit Reading Instruction

Nancy J. Nelson-Walker, PhD[1], Hank Fien, PhD[1], Derek B. Kosty, BSc[2], Keith Smolkowski, PhD[2], Jean Louise M. Smith, PhD[1], and Scott K. Baker, PhD[1]

Abstract

This article examines the efficacy of a multitiered systemic reading intervention for increasing the intensity and quality of explicit literacy instruction that teachers provide in first-grade classrooms. Schools (j = 16) were randomly assigned to the treatment or comparison condition. In both conditions, teachers (i = 42) provided 90 min of Tier I reading instruction to first-grade students (n = 883). In the treatment condition, Tier I classroom teachers were trained to enhance their core reading instruction by making instruction more explicit and intensive through standardized protocols and ongoing coaching support. At-risk treatment students (n = 240) also received an additional 30 min of Tier II supplemental reading intervention that was highly aligned with Tier I instruction. The focus of this study is on the Tier I portion of the multitiered intervention and the impact of the Tier I Enhancing Core Reading Instruction intervention on teacher practices. Results indicate positive effects of the Tier I intervention on the quality of explicit instruction and the frequency and accuracy of group practice opportunities provided to students.

Keywords

reading, observation research, teacher, instructional strategies, at risk

Many professions require the use of evidence to inform practice (Brownson, Baker, Leet, Gillespie, & True, 2011; Chwalisz & Dollinger, 2009). For example, in the public health field, professionals have long been engaged in experimental research to identify practices that improve the health and well-being of the general public (Sterman, 2006). In the field of medicine, scientists adjudicate the evidence from drug trials to determine the potential efficacy and effectiveness of drugs before they are permitted for sale in the marketplace (Elstein, 2004; Guyatt, Kirshner, & Jaeschke, 2009). More recently, education has significantly increased its emphasis on the application of scientific inquiry to identify programs and practices that promote student learning. At the forefront of this effort, the No Child Left Behind Act (NCLB) of 2001 called for the use of "effective, scientifically based instructional strategies" to ensure all students have access to high-quality instruction (20 U.S.C. §1001). Subsequently, the Education Sciences Reform Act (2002) established the Institute for Education Sciences (IES). IES has used a number of mechanisms to conduct and disseminate the findings of rigorous research, including the establishment of the What Works Clearinghouse (WWC). A specific goal of the WWC is to assist educators with the difficult task of identifying "research with credible and reliable evidence to use in making informed decisions" (WWC, 2011).

The shift toward the use of evidence to inform school practice has not yet been clearly associated with substantive increases in student achievement (PISA, 2009). For instance, the nation's youth have experienced statistically significant increases in reading achievement since the enactment of NCLB, but these increases have been relatively small (National Center for Education Statistics [NCES], 2011). Moreover, much work remains to be done. The 2011 results of the National Assessment of Educational Progress (NAEP) in reading show that only 8% of U.S. students scored at the advanced level. Perhaps more unsettling, 33% of U.S. fourth graders scored below basic, including 68% of students with disabilities.

Likewise, the shift in policy toward the promotion of evidence-based practices in schools has not been met with

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

Corresponding Author:
Nancy J. Nelson-Walker, University of Oregon, 5292 University of Oregon, Eugene, OR 97403, USA.
Email: nnelson3@uoregon.edu

widespread implementation of evidence-based instruction in U.S. classrooms (Cook & Cook, 2011; Odom, 2009). In fact, there are many examples of evidence-based practices not making their way into U.S. classrooms on a large scale (Carnine, 1997; Walker, 2004; Walker & Gresham, 2003). For instance, there is ample evidence indicating that reading instruction should be a priority in the early grades and young students should actively engage with print during reading activities (Chard & Kame'enui, 2000; Snow, Burns, & Griffin, 1998). However, research indicates that students receive far less than the recommended amount of instruction, instructional blocks are often used inefficiently, and the instruction students do receive rarely requires them to actually read print (Kent, Wanzek, & Al Otaiba, 2012). In light of this evidence, the national effort to promote evidence-based practice in U.S. classrooms and to translate the evidence base into feasible instruction models that can take root and be sustained in authentic school settings is critical.

Perhaps no feature of reading instruction has received as much support for increasing at-risk readers' learning as explicit instruction. In the recent IES Practice Guide, *Assisting Students Struggling with Reading: Response to Intervention and Multi-Tier Intervention in the Primary Grades* (Gersten, Compton, et al., 2009), the only recommendation with a strong level of evidence is to "provide intensive, systematic instruction on up to three foundational reading skills in small group" (p. 6). In other words, experts in primary grade literacy instruction, supported by research evidence, recommend that intense, explicit teaching be a regular component of the reading instruction delivered to students with or at risk of reading difficulties. Given the merits of using evidence-based practices such as explicit instruction to support students learning to read, the need to intensify instruction for students at risk of learning difficulties, and the challenges associated with implementing these practices in schools (Kratochwill & Shernoff, 2004), this study examined the efficacy of the Enhancing Core Reading Instruction (ECRI) intervention for changing teacher instructional behaviors.

Conceptual and Empirical Support for Explicit Instruction

In beginning reading, few educational practices have garnered as much evidence as the use of systematic, explicit instruction. The National Reading Panel (NRP), convened to "assess the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read," identified evidence-based practices for beginning reading instruction in their final report, underscoring the importance of *systematic* and *explicit* instruction in each of the five key areas of beginning reading (NRP, 2000, p. 1). The panel also endorsed the recommendations of the Committee on the Prevention of Reading Difficulties in Young Children (Snow et al., 1998), which indicated that (a) primary grade reading instruction is critical for preventing reading failure and (b) explicit instructional strategies support literacy achievement for all students, but especially those at risk of reading difficulties (Snow et al., 1998).

Other reports extend the findings of the NRP and Committee on the Prevention of Reading Difficulties in Young Children. In concluding there was strong evidence for the use of systematic, explicit instruction to prevent reading difficulties, the IES Practice Guide (Gersten, Compton, et al., 2009) identified 11 studies meeting WWC evidence standards. Seven of these studies found a positive effect on at least one key early reading skill when using methods of explicit instruction in small group settings.

In one study, Gunn, Biglan, Smolkowski, and Ary (2000) conducted a 56-week, randomized control trial examining a multicomponent early reading intervention for students in kindergarten through third grade. Students in the treatment condition received explicit instruction in phonemic awareness, decoding, fluency, and comprehension for 25 min each day in small groups (1–3 students). Gunn et al. (2000) found significant effects for students in the treatment condition in decoding, vocabulary, and reading comprehension. Effects for fluency approached significance.

In another study, Vaughn et al. (2006) studied a 28-week intervention for at-risk, first-grade students targeting phonemic awareness, letter knowledge, word recognition, connected text fluency, and comprehension strategies. Students in the treatment condition received 50 min of instruction each day in small groups composed of three to five students. The intervention demonstrated significant effects favoring the treatment condition for decoding and reading comprehension in English. The Gunn et al.'s (2000) and Vaughn et al.'s (2006) studies provide strong conceptual and empirical support for the use of systematic and explicit instruction in small group settings to improve the outcomes of students at risk of difficulties in beginning reading.

Other studies have explored the efficacy of *incorporating explicit instruction for at-risk students in the context of Tier I instruction.* In our own work, we have taken the research evidence on explicit, systematic instruction and applied it to large-group settings with promising results in math and reading. In one study of a kindergarten mathematics curriculum incorporating explicit instructional strategies (Early Learning in Mathematics [ELM]), randomly assigned treatment classrooms implemented ELM with the full range of students in general education. Preliminary evidence supports the efficacy of ELM (Clarke, Smolkowski, Baker, Fien, & Chard, 2011): Students in the treatment condition outperformed their control group peers on early math curriculum-based measures and a standardized, norm-referenced assessment of whole number proficiency (Hedges' g = .13–.14). Moreover, in a condition by risk

status analysis, at-risk students significantly outperformed their at-risk counterparts on the same assessments (Hedges' g = .22–.24). In the context of a multitiered instructional approach, these effect sizes (ESs) are important with respect to identifying practices that aid the prevention of math difficulty for at-risk students.

In addition, we have studied the effect of a whole-class, Tier I read aloud intervention that used an explicit instructional approach in before-, during-, and after-reading segments of read aloud lessons to develop student listening comprehension and vocabulary knowledge (Baker, Fien, Santoro, Chard, & Park, in press). In one study, 12 first-grade classrooms were randomly assigned to the read aloud intervention condition or a comparison condition. Intervention students outperformed comparison students on a standardized narrative retell measure (d = +0.41). Risk moderation analyses found positive trends for the intervention condition on a receptive listening comprehension measure favoring at-risk students (d = +0.41), and a narrative retell measure (d = +0.29). On the vocabulary measure, the overall effect favoring students in the intervention condition was also significant (d = +1.00). Akin to our work in early mathematics, the read aloud study provides preliminary evidence that explicit instructional design technology can be applied to large group, Tier I instructional settings to benefit all students, particularly those at risk of failure in critical academic content.

Conceptual and Empirical Support for Instructional Intensity

Instructional intensity has received increasing attention as various policies and initiatives have pushed for better understanding of instructional practices that are effective for preventing learning difficulties and ameliorating persistent learning problems (e.g., Batsche et al., 2006). A number of studies indicate there are features of instruction that can be manipulated to intensify instruction in the early grades. For example, O'Connor, Harty, and Fulmer (2005) examined the efficacy of a longitudinal, tiered intervention designed to provide increasing levels of support to students across multiple years of intervention. Kindergarten students scoring in the lower third of their class on a measure of beginning reading achievement were eligible for Tier II intervention, delivered in small groups for 10 to 15 min, three times per week. Supports were intensified for students that continued to demonstrate need for Tier II intervention beyond kindergarten. Students who received intervention significantly outperformed control group peers on measures of beginning reading achievement at the end of third grade (O'Connor et al., 2005).

Other studies demonstrate that students who receive more intensive intervention display significant improvements in achievement when compared with students who receive typical instruction. Simmons et al. (2008) studied the longitudinal effects of a small group reading intervention in kindergarten. Students meeting criteria for the intervention condition received at least 30 min of supplemental reading instruction, 5 days per week in groups of three to five students. Results indicate that students who received additional, small group (i.e., intensified) instruction in kindergarten were more likely to be on track for average reading achievement at the end of the year, and students who scored above the identified at-risk category on reading outcome measures at the end of kindergarten continued to perform above the at-risk range through third grade (Simmons et al., 2008). Vellutino et al. (1996) also implemented a longitudinal study of intensive reading instruction and manipulated group size and instructional time to intensify instruction for students with reading difficulties. The authors examined the efficacy of a tutoring intervention for first-grade students identified as "poor readers" on the basis of their scores on intelligence and reading achievement measures. Students in the intervention group received daily, individualized, one-on-one tutoring in reading for at least 15 weeks. Students in the control group received business-as-usual reading intervention, which varied considerably from school to school. More than two thirds of students classified as poor readers scored in the average range on standardized measures of reading achievement after only one semester of intensive tutoring intervention (Vellutino et al., 1996).

The literature is ripe with instances of intense instruction supporting the achievement of students who struggle to learn to read in the early grades. However, intensity is often poorly defined in studies and frequently limited to descriptions of duration, which diminishes generalizability of findings and prevents the systematic study of intensity as a malleable characteristic that can aid intervention effectiveness (Warren, Fey, & Yoder, 2007). More precise measurement estimates of intensity in intervention studies can contribute to the field by clearly articulating the features of an intervention that lead to attained outcomes.

Features and Goals of Explicit and Intensive Instruction

Explicit instruction is a structured, systematic instructional approach that can be used to effectively teach students from a range of backgrounds and skill levels (Archer & Hughes, 2011; Coyne, Kame'enui, & Carnine, 2006). Instructional design experts describe a number of elements inherent to the effective design and delivery of explicit instruction. Coyne et al. (2006) recommended that six instructional design principles (i.e., emphasizing critical content, making learning strategies conspicuous, introducing and fading instructional scaffolds, activating prior knowledge, helping

students strategically integrate knowledge, and providing systematic and judicious review of concepts) and five instructional delivery features (i.e., appropriate pacing, frequent student responding, adequate think time, response monitoring, and specific, immediate student feedback) be used to structure interactions between teachers and students and make content more accessible to students.

Researchers have also emphasized the critical role of guided and supported practice in the learning of new skills and argue that providing frequent opportunities for students to engage in practice is a key feature of effective, explicit instruction (Archer & Hughes, 2011; Rupley, Blair, & Nichols, 2009; Watkins & Slocum, 2004). However, simply providing large amounts of practice is unlikely to result in significant advances in student achievement (Ericsson, Roring, & Nandagopal, 2007; Silverman, 1985). Instead, practice must be thoughtfully structured to facilitate high rates of accuracy and gradual exposure to new and challenging content, for students to improve their skills (Archer & Hughes, 2011). Ericsson et al. (2007) propose that practice must be *deliberate* for a student to transcend from novice to expert:

> Deliberate practice does not involve a mere execution or repetition of already attained skills, but repeated attempts to reach beyond one's current level . . . by engaging in practice activities designed to change and refine particular (skills), requiring problem-solving and successive refinement with feedback. (pp. 18–19)

Explicit instruction allows the teacher to carefully design deliberate and meaningful student practice, and to adjust the level of challenge and scaffolding to support student success.

A major goal of explicit, systematic reading instruction is to provide intensive instruction for students that struggle to acquire early literacy skills. Intervention researchers focus on malleable variables such as group size, intervention duration, and intervention frequency, and manipulate these variables to increase intervention intensity (Fuchs, Compton, Fuchs, Bryant, & Davis, 2008; Harn, Linan-Thompson, & Roberts, 2008; Simmons et al., 2008). Warren et al. (2007) provide a comprehensive framework for defining and measuring intervention intensity that can be used to guide intervention development and to study the effects of intensive interventions. Briefly, Warren et al. (2007) posit that intervention intensity can be conceptualized and quantified as the product of Dose × Dose frequency × Total intervention duration. The critical element of this algorithm is what constitutes a "dose." We conceptualize "dose" as a structured interaction between a teacher and a student around a task or series of tasks involving critical concepts (i.e., "big ideas"), in which a number of essential features of instructional design and delivery are integrated (e.g., teacher demonstrations, checks for understanding, corrective and confirmatory feedback). Using this definition of "dose," instruction can be intensified by increasing the rate of student–teacher interactions that occur within an instructional lesson or activity. Within this framework, factors such as group size, frequency, and intervention duration could be viewed as proxy variables associated with systematic increases in the number of interactions between students and teachers. For example, reducing group size would only increase intervention intensity if it was actually accompanied by an increase in the number of academic interactions between teachers and students.

In the context of explicit instruction, one way to increase intensity is to structure lessons and activities so that teachers elicit frequent student responses, including opportunities to practice lesson content (Archer & Hughes, 2011; Coyne et al., 2006). Eliciting frequent responses from students as a way to increase student engagement and retention is well documented (e.g., Brophy & Good, 1986; Sutherland & Wehby, 2001). The use of choral student responding is one technique often used in classrooms to increase response rates and the engagement of all students, particularly in large-group settings (Archer & Hughes, 2011; Coyne et al., 2006; Howell & Nolet, 2000). Because choral responding allows the teacher to provide *all* students with frequent, structured practice opportunities, check student understanding, and deliver specific and immediate feedback, it can increase the overall efficiency of student practice (Archer & Hughes, 2011). Thus, when provided in the context of carefully designed, systematic and explicit instruction, choral responding can increase student learning (Fien et al., in press; Watkins & Slocum, 2004).

In the ECRI intervention, we trained teachers to use instructional materials that were designed to increase the frequency of student choral practice in whole and small group formats. We also emphasized the role of engaging students in multiple, varied opportunities to practice foundational reading skills and to apply comprehension strategies, as a way to support high-quality delivery of explicit reading instruction. Thus, using the Warren et al. (2007) framework, we conceptualized and measured instructional intensity as the number of student responses that teachers elicited in a 90-min Tier I reading block.

Changing Teacher Behavior: Implementation Challenges

Although research indicates delivering intense, explicit instruction is an evidence-based practice that can be used to achieve positive learning outcomes for students at risk of reading difficulties, there are a number of challenges that frequently prevent evidence-based practices from taking hold in schools and classrooms. For example, research findings are sometimes ignored by practitioners—because they

may be deemed untrustworthy, of little utility, or inaccessible on a day-to-day basis (Carnine, 1997). Beyond the "research-to-practice gap" (Carnine, 1997), there are other challenges that make implementation of evidence-based practices difficult. Even when practitioners are knowledgeable about the content and instructional practices that have been shown to improve student achievement, it is uncommon for teacher knowledge on its own to translate to student gains (Carlisle, Kelcey, Rowan, & Phelps, 2011). In addition, implementation of new practices is difficult to sustain, because initial execution often tapers off over time, and indirect attributes of the environment (e.g., administrative support, staff turnover) have the potential to counteract the efforts of the evidence-based practice (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005).

A great deal of professional development (PD) is intended to improve the implementation of evidence-based practices in classrooms with the ultimate goal being increased student learning. However, a number of PD studies report no impact on changing teacher behavior or on improving student achievement. Several recent research publications highlight the paucity of rigorous studies demonstrating positive effects for PD (e.g., Garet et al., 2011; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). For instance, less than half of the PD studies reviewed by Yoon et al. (2007) that met evidence standards demonstrated positive, significant effects for reading or language arts outcomes. Commonly, even when studies report ESs for student achievement, they do not do so for teacher behavior change. Marsh et al. (2008) studied the effects of a statewide coaching initiative to improve teachers' literacy instruction and found very small effects (0.04–0.06) on student achievement in half of their cohorts but did not report effects for changes in teacher instructional behavior.

A handful of studies demonstrate that efforts to alter teacher behavior result in positive changes in teacher practices, with mixed effects on student learning targets. Gersten, Dimino, Jayanthi, Kim, and Santoro (2009) conducted a study of the Teacher Study Group (TSG) PD model for increasing teacher knowledge, changing teacher practice, and improving student vocabulary knowledge and reading comprehension skill in first-grade classrooms. TSG participants scored higher than control group teachers on posttest measures of instructional vocabulary knowledge (ES = 0.73), and classroom observations indicated TSG teachers provided vocabulary and comprehension instruction that was of higher quality than instruction provided by control group teachers (Gersten, Dimino, et al., 2009). However, changes in TSG teacher practice did not translate to significant differences in vocabulary or comprehension achievement for students in these classrooms. In contrast, Al Otaiba et al. (2011) studied a PD intervention designed to support individualized instruction for kindergarten students receiving explicit, systematic instruction. Teachers in the treatment condition provided significantly more individualized instruction than teachers in the control condition and students in classrooms receiving the intervention significantly outperformed peers in control classrooms on a latent measure of basic reading skills (ES = 0.52).

Congruent with the small set of studies that have demonstrated positive effects on teacher behavior, a body of implementation research recommends the use of PD with built-in opportunities for practitioners to practice skills and ongoing coaching to provide support for new practice (Fixsen et al., 2005). In addition to incorporating active learning opportunities and using reform-based methods that rely on ongoing training, well-designed PD is frequently provided to a collective, cohesive group (e.g., all of the first-grade teachers in a school building); is coherent and aligned with teachers' and students' learning goals; and focuses on the content teachers need to know to provide instruction (Desimone, Porter, Garet, Yoon, & Birman, 2002). In addition, in-service PD—in particular, training delivered directly to teachers and cumulating in more than 14 hrs of total training—has the potential to significantly increase student achievement (Yoon et al., 2007). The ECRI Intervention, which emphasizes explicit, systematic instruction, relies on a PD model that incorporates these elements to aid the effectiveness of the intervention.

Purpose of the Study and Research Hypotheses

The purpose of this study was to test the efficacy of the ECRI intervention for supporting the implementation of evidence-based practices that teachers deliver in Tier I settings by increasing the *quality* and *intensity* of explicit instruction provided to students. Although the distal outcomes of the ECRI systemic intervention included student measures of reading achievement, our proximal outcomes were focused on changing teachers' instructional behaviors in the context of Tier I instruction. In other words, we intended to increase student reading achievement by increasing the degree to which teachers delivered intensive and high-quality explicit instruction in Tier I. The rationale for providing more explicit instruction in Tier I is that the majority of at-risk readers' instructional time occurs in Tier I. For example, many schools provide a 90-min Tier I reading block and an additional 30 min of Tier II supplemental intervention, where 75% (90/120 min) of at-risk readers' instructional time is spent in Tier I.

The ECRI intervention was designed to increase (a) the quality of explicit instruction and (b) the intensity of Tier I instruction via the provision of more frequent student response opportunities (e.g., choral responses) to support the accessibility of Tier I instruction for students at risk of reading difficulties. The ECRI intervention attempted to

improve the quality of explicit instruction in several ways. Throughout the school year, supplemental instruction materials, PD, and ongoing coaching were used to support implementation of explicit instruction in Tier I. Specifically, the ECRI intervention incorporated explicit instructional design technology into Tier I *lesson maps* and *instruction templates.* The purpose of the lesson maps and instruction templates was to provide tools for teachers, which would increase the likelihood that they would successfully deliver high-quality explicit instruction to at-risk students in the context of Tier I instruction. The ECRI intervention attempted to increase the intensity of instruction (i.e., the overall frequency of student responses) by working with teachers to incorporate choral responding into their teaching methodology. We trained and coached teachers to provide students with more choral response opportunities, respond to student errors with corrective feedback, and strategically distribute differentiated individual response opportunities to students based on students' skill levels.

We hypothesized the ECRI intervention would increase the quality of explicit instruction teachers provided as measured by the *Quality of Explicit Instruction* (QEI) scale, including clearer teacher models, higher quality guided individual and choral practice opportunities, and the appropriate use of teacher feedback. Our rationale for posing this hypothesis was that the lesson maps and instruction templates would provide an advance organizer (Ausubel, 1968) for teachers to attend to qualitative features of explicit instruction. In addition, we reasoned that the ongoing PD and coaching would provide teachers an opportunity to practice implementation quality explicit instruction, receive ongoing feedback from their coach, and set goals for continued improvement.

In addition, we hypothesized that ECRI would increase *intensity* of Tier I instruction as measured by the Classroom Observations of Student–Teacher Interactions (COSTI) instrument (Smolkowski & Gunn, 2012). The COSTI tool is used to record the frequency of teacher models, group practice opportunities, individual practice opportunities, student errors, and corrective feedback. Because the ECRI intervention emphasized explicit instruction and supported practice, we expected more choral practice opportunities in the treatment condition and operationalized this type of responding as one major component of intervention intensity (Archer & Hughes, 2011). Although we had no specific hypotheses regarding the frequency of teacher models, we did anticipate improving the quality of teacher models as a critical component of explicit instruction. Likewise, although we anticipated an increase in the quality of individual practice opportunities (i.e., differentiating individual practice opportunities based on student skill), we did not anticipate an increase in the frequency of individual practice opportunities. We further hypothesized there would be (a) an overall decrease in the number of errors students committed during instruction (i.e., student practice would be more accurate) and (b) a relative increase in teacher corrective feedback when student errors occurred (i.e., more frequent attempts to clarify student misconceptions).

Method

Project Design

ECRI is a 4-year project designed to study the efficacy of a multitiered intervention system using a cluster-randomized controlled trial, nesting students and first-grade teachers within schools. Within district, schools were randomly assigned to either the ECRI treatment or a wait-list comparison condition. By blocking on districts, we controlled for core curricula and other important factors. We assessed all first-grade students in treatment and comparison classrooms on preliteracy and literacy measures in the fall of first grade to assign students to Tiers I, II, and III and again in the spring to measure gains on important skills. Treatment students in Tier I received enhanced core reading instruction, whereas students in Tier II received the same enhanced core instruction plus additional small group supports. Treatment and comparison schools provided Tier III instruction and intervention consistent with typical practices in each building. We conducted systematic observations in all treatment and comparison Tier I classrooms three times during the year to document implementation fidelity and to measure instructional variables hypothesized to mediate student achievement. This article presents findings from observations conducted in Tier I classrooms participating in the 1st year of the project. Although the systemic ECRI intervention included Tier I and Tier II components, given we are reporting teacher outcomes from Tier I, we emphasize the features of the Tier I component of the intervention.

ECRI Tier I Intervention. ECRI treatment classrooms committed to providing at least 90 min of daily Tier I whole group, core reading instruction and an additional 30 min of highly aligned, Tier II small group instruction. Lesson maps and instruction templates were used to enhance the content, design, and delivery of instruction in ECRI treatment classrooms. Many core reading programs that are used to guide Tier I instruction do not provide sufficient explicit, scaffolded instruction or practice opportunities for learners at risk of reading difficulty (Baker, Fien, & Baker, 2010; Gersten, 1999). Our experience is that explicit teacher models or demonstrations are frequently absent from instruction, or are poor in quality (e.g., vague or inconsistent). The practice opportunities necessary for deep understanding are limited in number, do not provide enough guided support for struggling readers, and are often restricted in range so that students end up practicing what they have learned under a narrow set of examples. In addition, many students do not receive opportunities for supported practice because practice

opportunities may be distributed unsystematically, time does not allow each student to receive necessary individual practice, and feedback designed to improve student understanding is lacking.

To address these limitations, ECRI lesson maps and instruction templates emphasize the following features of explicit instruction: (a) learning strategies are made conspicuous through visual models, verbal directions, full and clear explanations, and outlined steps; (b) instruction is integrated to build connections between new and previously taught content; (c) instruction activates background knowledge so that students can more easily integrate new knowledge; (d) instructional scaffolds are provided and removed systematically as students demonstrate greater control of learning over time; (e) sufficient practice to achieve automaticity is provided; and (f) previously learned material is reviewed systematically (Carnine & Kame'enui, 1992; Chard & Jungjohann, 2006; Coyne et al., 2006). In addition, lesson maps constructed for each core program prioritize for teachers the essential components of reading instruction; these maps were designed to ensure that instructional time is spent teaching essential content. Instruction templates were designed to increase the quality and intensity of instructional interactions between teachers and students by incorporating clear models of key reading skills and concepts, ample guided and structured practice opportunities for students, and consistent teacher feedback (Baker et al., 2010).

For example, the core curriculum in the comparison condition might identify the target words to teach students in the day's lesson, but it typically will not offer guidance about the language teachers should use when instructing students, the amount of practice to include, or the way to provide corrective or confirmatory feedback to students when they respond. In the ECRI condition, instruction templates matched to each target skill in the core curriculum provided a structure teachers used for distributing practice to the whole class and to select individual students. The templates provided the language teachers should use when modeling content and seeking responses from students, and the method for providing immediate feedback to students. The careful integration of lesson maps and templates was intended to provide a highly specified teaching routine to increase student engagement.

Teachers in the ECRI treatment condition received five 8-hr days (3 days in the fall and 2 days in early winter) of PD. Expert ECRI coaches provided the PD, emphasizing the science of beginning reading, principles of effective instruction, and procedures for ECRI through the use of lesson maps and instruction templates. During this PD, expert coaches modeled instructional routines and teachers practiced using routines in small groups. The content emphasized during PD was explicit instruction in each of the five critical areas of reading identified by the NRP (2000). Additional PD occurred throughout the year for treatment teachers, primarily through monthly meetings with expert coaches, who observed instruction and provided supplementary training on the effective delivery of the ECRI intervention. A standardized coaching protocol was used across sites to ensure a consistent PD focus, including targets of the coaching observations and feedback provided to ECRI treatment teachers.

Comparison schools. Comparison schools (a) used a comprehensive core program identified and adopted by standard district procedures for 90 min each day during Tier I instruction and (b) provided students identified for Tier II with 30 min of additional daily small group instruction. However, comparison teachers did not receive enhanced materials or ECRI PD. Districts agreed to schedule and provide typical PD activities to schools in the comparison condition.

Comparison schools differed from treatment schools in important ways related to key components of the intervention. Some comparison schools supplemented core instruction with extra support materials included in the core program materials; other comparison schools used supplemental materials (published or teacher developed) in addition to the core program. By design, ECRI treatment classrooms used a greater proportion of whole group instruction during core reading time, whereas comparison classrooms used a variety of formats in more equally distributed amounts, including whole groups, small groups, and reading centers.

Year 1 Participants

We originally recruited elementary schools from three school districts in the Pacific Northwest to participate in the first wave of the study. These school districts were targeted for recruitment because they had adopted a first grade, core reading program; served students from diverse backgrounds (e.g., urban and rural, English learner [EL] and English only, free and reduced-price lunch [FRL] vs. non-FRL); and were implementing a multitiered instructional delivery system. Twenty-two schools expressed interest in participating, but because of changes in school leadership between recruitment and the beginning of the study, four schools decided not to participate prior to randomization. Thus, 18 schools were randomly assigned to the ECRI treatment or a wait-list comparison condition. Two schools, one treatment and one comparison, left the project at the beginning of the 1st year, leaving 16 schools in the study (a rate of 11% overall attrition with no differential attrition by condition).

In these 16 schools, 42 first-grade teachers (23 treatment, 19 comparison) taught 883 student participants. Teachers were female (except one) and averaged 14 years of teaching experience (SD = 8.71). Of the student participants, 240 students were identified as needing Tier II support (27%) on the basis of their performance on the Stanford

Achievement Test–Tenth Edition (SAT-10) in the fall of first grade. Student ethnicity as reported by schools to the National Center on Education Statistics was 65.27% White, with another 21.38% of students identified as Hispanic, and 2.52% identified as African American. Approximately 20% of students were identified as ELs, 46% received FRL, and 14% were identified as needing special education services.

Project coaches ($n = 2$) were nondistrict employees with expertise in instructional design and delivery, intervention development, and coaching. They received initial training and ongoing supervision from a principal investigator with extensive knowledge of instructional design principles and experience designing interventions and coaching schools. The coaching team, including the principal investigator, met monthly for the duration of the school year to calibrate support provided to schools and districts and improve intervention implementation in accordance with the standardized coaching protocol.

Observation Measures

QEI. The QEI measure was designed to document (a) fidelity of implementation of the intervention and (b) the QEI in critical early literacy skills. The QEI contains five items that address the fidelity of implementation of the intervention, including the degree to which the intervention materials (i.e., lesson maps and instruction templates) were used in the classroom. We collected this information in treatment (to measure fidelity) and comparison (to measure treatment diffusion) classrooms. Six additional QEI items address the quality of explicit instruction, documenting the degree to which instruction consistently contained high-quality teacher demonstrations, student practice, and corrective or confirmatory feedback for each of the foundational concepts of reading (i.e., phonological awareness, alphabetic principle, accurate and fluent reading of connected text, vocabulary, and comprehension). An explicit instruction summary score is provided, representing the average of the six explicit instruction quality items.

On 41 occasions, two observers collected data simultaneously to assess interobserver reliability. As an index of interobserver reliability, we estimated the variation in QEI 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. According to guidelines provided by Landis and Koch (1977), observers were highly reliable in rating the specific and general fidelity of implementation (ICC = .82 and .86, respectively). Analogous to test–retest reliability estimates, we estimated the stability of QEI scores using an unconditional multilevel model that nested three observations per classroom within each of the 42 classrooms. A stability ICC was computed as the proportion of between-classroom variance. Ratings of the specific and general fidelity of implementation were reasonably stable over time (ICC = .46 and .71, respectively) suggesting that three observations per classroom provided a reliable estimate of fidelity of implementation (Shoukri, Asyali, & Donner, 2004).

COSTI. A modified version of the COSTI (Smolkowski & Gunn, 2012) was used to document instructional intensity in the study, which we conceptualized broadly as the rate of student–teacher interactions, and more specifically as the frequency of choral practice opportunities. The COSTI is used to record four types of student–teacher interactions, including teacher demonstrations, teacher corrective feedback, student independent practice opportunities, and student errors (see Smolkowski & Gunn, 2012, for definitions of these behaviors), along with several contextual variables (e.g., group size, primary content focus of the instructional activity, start and stop times for the activity). Because the frequency of student–teacher interactions are documented within blocks of time, rates of individual student and teacher behaviors can be calculated for each observation. Research on the use of the COSTI indicates it is a valid and reliable measure of teacher initiated behavior in early reading (Smolkowski & Gunn, 2012).

The COSTI was modified for use in the ECRI study in several ways. First, student independent practice was parsed into three separate categories: 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). In addition, observers using the modified COSTI were trained to code confirmatory *and* corrective feedback from the teacher, when delivered immediately following a student practice opportunity to confirm, extend, clarify, or correct the student response. In this study, COSTI data were reported as rates of student and teacher behaviors, calculated by dividing the total frequency of each behavior by the number of instructional minutes observed.

Interobserver reliability and stability estimates were obtained by following the same procedures as described for the QEI. 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). Rate of group practice was reasonably stable over time (ICC = .78) suggesting that two observations per classroom likely provided an authentic estimate of the rate of group practice (Shoukri et al., 2004). Teacher demonstrations, individual practice, covert practice, student mistakes, and feedback were less stable over time (ICCs = .28, .34, .17, .36, and .16, respectively)

suggesting that additional observations are necessary to obtain accurate estimates of these behaviors.

Observation Procedures

Trained data collectors observed core reading instruction in all ECRI treatment and comparison classrooms three times during the school year. Per observation, reading instruction was observed for an average of 99.86 min (*SD* = 19.22, range = 80.50–146.50) in ECRI treatment classrooms and 91.17 min (*SD* = 9.27, range = 74.50–114.50) in comparison classrooms. To support the feasibility of collecting QEI and COSTI data in a single observation, observers did not use the COSTI until the winter observation period. Consequently, COSTI data were collected in all classrooms twice during the school year—during the winter and spring observations—whereas the QEI was collected in all classrooms during the fall, winter, and spring.

Prior to each observation round, observers received training targeting four components of study observations: (a) background information about the observation system and/or a review of the previous training session's content, (b) introduction of new information, including procedures for completing the observation form(s), (c) practice using the observation form(s) with videos of reading instruction, and (d) opportunities for independent practice and feedback using videos and correctly coded protocols. At the end of each 3-day training session, data collectors watched a video of instruction and completed a training reliability checkout to examine initial interobserver agreement (calculated as the number of coding agreements divided by the total number of behaviors coded by the observation coordinator or "gold standard" observer). Observers were also required to complete a field reliability checkout and meet an interobserver agreement standard of .80 before observing independently during each observation round. To examine maintenance reliability throughout each observation round, 23% of all observations were paired. Average field reliability across the year for all field reliability checkouts and paired observations was .94.

Statistical Analysis

We assessed the effects of intervention condition on observer ratings of the quality of explicit instruction and instructional intensity within a series of conditional multilevel models in which classrooms (Level 1) were nested within schools (Level 2). Observer ratings were predicted by a dichotomous school-level variable indicating ECRI treatment and comparison conditions (1 and 0, respectively). The statistical models can be represented by the equation:

$$\text{Obs}_{ij} = \beta_{00} + \beta_{01}\text{ECRI}_j + \mu_{0j} + r_{ij}$$

Obs_{ij} is the value of the mean observer rating across all observation occasions for classroom i in school j. The two fixed effects are the average observer rating for comparison schools (β_{00}) and the effect of ECRI treatment condition on observer ratings (β_{01}). The two random effects are the classroom-level residual (r_{ij}) and the school-level residual (μ_{0j}). Tests of fixed effects used 14 degrees of freedom. Multilevel modeling was conducted using hierarchical linear modeling (HLM; Raudenbush, Bryk, Cheong, & Congdon, 2004), and parameters were estimated using restricted maximum likelihood (RML).

To ease the interpretation of results, we computed two ESs for each fixed effect of ECRI condition: Hedges' g (Hedges, 1981) and Pseudo R^2 (Singer & Willett, 2003). Hedges' g, recommended by the WWC (Seftor et al., 2011), represents an ES comparable to Cohen's d (Cohen, 1988), except that Cohen's d uses the sample standard deviation, whereas Hedges' g uses the population standard deviation (Rosenthal & Rosnow, 2008). The Pseudo R^2 represents the decrease in classroom-level variance between unconditional and conditional models, or the proportion of variance explained in the outcome measure by a predictor or set of predictors. We reported Hedges' g and Pseudo R^2 because each index provides unique and meaningful information. Hedges' g describes how the conditions differ with respect to the outcome measures in standard deviation units. Pseudo R^2 describes the strength of association between intervention condition and the outcomes.

Results

Table 1 presents classroom-level means and standard deviations of observer ratings by intervention condition, including features of reading curriculum observed, quality of explicit instruction, instructional intensity, and quality of practice. This study aimed to test the hypotheses that teachers in the ECRI treatment condition would outperform teachers in the comparison condition on the quality of explicit instruction and instructional intensity. As described in the Method section, the hypotheses were tested within a series of conditional multilevel models in which observer ratings were predicted by a dichotomous variable indicating experimental condition. Tables 2 and 3 summarize complete model results for the effects of the ECRI intervention on the outcome measures, including measures of ESs (Hedges' g and Pseudo R^2) and ICCs.

Quality of Explicit Instruction

The main effect of condition statistically significantly favored ECRI treatment schools with respect to the overall quality of explicit instruction (t = 3.71, p = .003, Hedges' g = 1.31, Pseudo R^2 = .74). Exploratory post hoc analyses were conducted to determine the specific content areas in which the

Table 1. Classroom-Level Descriptive Statistics for Observation Measures by Intervention Condition.

Measure	Control (n = 24)	Intervention (n = 18)
	M (*SD*)	*M* (*SD*)
Features of reading curriculum[a]		
Lesson maps used	0.11 (.16)	0.90 (.16)
Instruction templates used	0.08 (.13)	0.91 (.16)
Quality of explicit instruction[a]		
Overall[b]	0.57 (.16)	0.78 (.16)
Phonemic awareness	0.17 (.28)	0.85 (.19)
Alphabetic principle	0.62 (.26)	0.94 (.13)
Word reading	0.69 (.30)	0.90 (.16)
Reading connected text	0.87 (.17)	0.85 (.21)
Vocabulary	0.50 (.33)	0.60 (.31)
Comprehension strategy	0.59 (.25)	0.57 (.33)
Instructional intensity[c]		
Teacher model rate	0.81 (0.41)	0.82 (0.35)
Group response rate	1.25 (0.88)	2.92 (1.16)
Individual response rate	0.99 (0.48)	0.69 (0.34)
Covert response rate	0.15 (0.08)	0.17 (0.11)
Proportion of practice followed by error	0.06 (0.05)	0.04 (0.02)
Proportion of errors followed by feedback	0.76 (0.20)	0.79 (0.13)

[a]Items were rated as 0 (*no*), 0.5 (*partially*), or 1 (*yes*). [b]Represents the mean across the six content areas (e.g., phonemic awareness, alphabetic principle, etc.). [c]Instructional intensity was measured using the Classroom Observation of Student–Teacher Interactions. Items and summary scores were averaged across three observation occasions.

Table 2. Effects of Intervention Condition on Quality of Explicit Instruction Implementation (Overall and by Content Area).

Effect or statistic	Overall	Phoneme awareness	Alphabetic principle	Word reading	Reading connected text	Vocabulary	Comprehension strategy
Fixed effects							
Intercept (β_{00})	.57**** (.04)	.20** (.08)	.62**** (.05)	.69**** (.05)	.87**** (.04)	.49**** (.09)	.59**** (.07)
ECRI (β_{01})	.21*** (.06)	.67**** (.11)	.32*** (.08)	.21** (.08)	−.02 (.06)	.11 (.13)	.01 (.11)
Variance components							
Residual[a] (σ^2)	.02 (.15)	.03 (.18)	.04 (.19)	.06 (.25)	.04 (.19)	.07 (.27)	.07 (.26)
Classroom (τ^2)	.00 (.06)	.03**** (.19)	.01* (.10)	.00 (.00)	.00 (.00)	.04*** (.19)	.02** (.14)
Summary statistics							
Hedges' g	1.31	2.73	1.49	0.84	0.11	0.34	0.03
Pseudo R^2	.74	.77	.69	.80	.00	−.06	−.24
ICC (ρ)	.13	.51	.21	.00	.00	.34	.22

Note. ECRI = Enhancing Core Reading Instruction; ICC = intraclass correlation coefficient. Tests of fixed effects used 14 degrees of freedom. Entries for fixed and random effects include parameter estimates with standard errors and standard deviations, respectively, in parentheses. Pseudo R^2 represents the decrease in classroom-level variance between unconditional and conditional models.

[a]Significance test not conducted.

*p < .10. **p < .05. ***p < .01. ****p < .001.

quality of explicit instruction was increased by the ECRI intervention. ECRI treatment schools provided higher quality explicit instruction in the areas of phonemic awareness (t = 6.01, p < .001, Hedges' g = 2.73, Pseudo R^2 = .77), alphabetic principle (t = 3.98, p = .002, Hedges' g = 1.49, Pseudo R^2 = .69), and word reading (t = 2.69, p = .018, Hedges' g = 0.84, Pseudo R^2 = .80). The analyses did not produce statistically significant main effects of condition for the quality of explicit instruction in reading connected text (p = .685), vocabulary (p = .403), or comprehension strategies (p = .952).

Instructional Intensity

ECRI treatment schools provided higher rates of group practice opportunities than the comparison condition (t = 4.01, p = .001, Hedges' g = 1.63, Pseudo R^2 = .63). We

Table 3. Effects of Intervention Condition on Measures of Instructional Intensity.

		Practice per min			Proportion of	
Effect or statistic	Teacher demos per min	Group	Individual	Covert	Practice with errors	Errors with feedback
Fixed effects						
Intercept (β_{00})	0.81**** (.08)	1.26**** (.28)	0.97**** (.11)	.15**** (.02)	.06**** (.01)	.76**** (.05)
ECRI (β_{01})	0.01 (.13)	1.64*** (.41)	−0.28 (.16)	.02 (.04)	−.03** (.01)	.03 (.07)
Variance components						
Residual[a] (σ^2)	0.14 (.38)	0.65 (.80)	0.13 (.37)	.01 (.09)	.00 (.04)	.02 (.15)
Classroom (τ^2)	0.01 (.08)	0.40*** (.63)	0.05** (.22)	.00* (.04)	.00 (.00)	.01** (.09)
Summary statistics						
Hedges' g	0.03	1.63	0.66	0.21	0.75	0.17
Pseudo R^2	−1.25	0.63	0.26	−1.69	0.92	−0.09
ICC (ρ)	.04	.38	.26	.19	.01	.28

Note. ECRI = Enhancing Core Reading Instruction; ICC = intraclass correlation coefficient. Tests of fixed effects used 14 degrees of freedom. Entries for fixed and random effects include parameter estimates with standard errors and standard deviations, respectively, in parentheses. Pseudo R^2 represents the decrease in classroom-level variance between unconditional and conditional models.

[a]Significance test not conducted.

*$p < .10$. **$p < .05$. ***$p < .01$. ****$p < .001$.

found no effects of condition on the rate of teacher demonstrations ($p = .956$), individual practice opportunities ($p = .105$), or covert practice opportunities ($p = .552$). The main effect of condition statistically significantly favored ECRI treatment schools with a lower rate of student errors ($t = -2.27$, $p = .039$, Hedges' $g = 0.75$, Pseudo $R^2 = .92$). The analyses did not produce statistically significant main effects of condition for the proportion of errors followed by teacher corrective feedback ($p = .679$).

Discussion

The purpose of this study was to examine the extent to which the ECRI systemic intervention improved the quality and intensity of teachers' explicit teaching behaviors. We hypothesized that the ECRI intervention would increase the quality and intensity of explicit instruction that teachers provide in Tier I settings. Results provide preliminary support for use of the ECRI intervention to increase the quality and intensity of explicit reading instruction in first-grade classrooms.

Regarding the quality of explicit instruction, we found large ($g = 1.31$), significant effects for the overall quality of explicit instruction favoring the treatment condition. Post hoc analyses revealed significant effects for the quality of phonemic awareness, decoding, and word reading instruction. In addition, we found a positive trend for the quality of explicit vocabulary instruction favoring the treatment condition. Although this effect was not significant, the moderate ES warrants attention (i.e., Hedges' $g = 0.34$). The WWC describes nonsignificant findings with ESs larger than $g = 0.25$ as substantive and rate the effectiveness as "potentially positive" (WWC, 2011). Arguably, it is more difficult to change teacher behavior when addressing instructionally complex topics, such as vocabulary and comprehension. Still, the potentially positive effect on improving the quality of explicit vocabulary instruction is promising.

The nonsignificant findings for improving the quality of comprehension and fluency instruction are likely the result of two factors. First, reading comprehension is arguably the most difficult construct to teach in beginning reading, and accordingly, affect through PD and coaching efforts. A recent meta-analysis conducted by Yoon et al. (2007) supports this proposition: Despite identifying 11 PD studies with positive effects on student outcomes, none of these studies reported positive effects on reading comprehension or general reading achievement. Rather, effects were primarily reported for foundational reading skills. If PD intends to improve instructional practice with the goal of improving student achievement, it is important to acknowledge that it may be more difficult to change teacher behavior as it relates to higher order skill instruction (Klingner, Vaughn, Hughes, & Arguelles, 1999). The second factor related to the nonsignificant finding for improving the quality of fluency instruction is that the primary setting for differentiated fluency instruction was in Tier II small groups. In other words, the majority of fluency instruction that at-risk students received was outside of Tier I instruction (the focus of this study). Thus, the nonsignificant finding for the quality of fluency instruction is not surprising.

Improving the overall quality of explicit instruction for at-risk students is a critical evidence-based practice supported by numerous meta-analyses and research syntheses (Gersten, Compton, et al., 2009; NRP, 2000; National Reading Council [NRC], 1998). Although studies frequently examine effects of PD on teacher knowledge (e.g., Garet et al., 2011), teacher perceptions (e.g., Garet, Porter,

Desimone, Birman, & Yoon, 2001; Heck, Banilower, Weiss, & Rosenberg, 2008; Marsh et al., 2008), or student achievement (e.g., Yoon et al., 2007), few studies have examined the impact of PD on changes to explicit teaching behaviors, and those that do not often report ESs for teacher behavior differences (e.g., Al Otaiba et al., 2011; Franke, Carpenter, Levi, & Fennema, 2001). Pianta, Mashburn, Downer, Hamre, and Justice (2008) reported "small effects" for the impact of an individualized PD intervention on teachers' language modeling behaviors. Gersten, Dimino, et al. (2009) found moderate to large effects (ES = 0.58–0.86) for the impact of targeted vocabulary and comprehension PD on observed teaching practice in first grade. The effect of the ECRI intervention for improving high-quality explicit reading instruction is commensurate with or slightly larger than the recent slate of studies of PD effects on teacher behaviors. The findings in this study further contribute to the literature on how to implement and improve evidence-based, high-quality explicit reading instruction in schools, and inform the field about the magnitude of ESs that might be expected in the context of early literacy PD and coaching efforts.

Our findings also indicate that instruction in treatment classrooms was delivered with significantly greater intensity (g = 1.61)—in terms of a relative increase in group practice opportunities—than instruction in comparison classrooms. Our PD and coaching emphasized the use of choral, group responses as a means for classroom teachers to increase the rate of student responding, particularly in large-group settings, and efficiently elicit more frequent responses from students. This is an important finding, given the clear and compelling evidence base for providing at-risk students with intensive reading supports and multiple and varied practice opportunities in the early grades (Gersten, Compton, et al., 2009; NRP, 2000; NRC, 1998). The rate of student choral practice opportunities provided in treatment classrooms was more than twice the rate of group practice provided in comparison classrooms in the same amount of time dedicated to Tier I instruction. Thus, we significantly increased the number of student responses in the treatment group, a goal set forth by proponents of explicit and intensive instruction (Archer & Hughes, 2011; Baker et al., 2010). We also found that practice in the treatment condition was significantly more accurate than practice in control classrooms. In other words, treatment teachers provided twice as much reading practice that was more aligned with students' skill level (as evidenced by significantly more accurate practice in the treatment group relative to the control group). Although there were no significant findings for the increased frequency of teacher models, our PD stressed improving the quality of teacher models, not necessarily the quantity of teacher models, and the intervention did have a significant impact on the overall quality of teacher models.

Implications for Research and Practice

Research indicates explicit and intensive instruction is critical when teaching students to read, in particular for students with or at risk of reading disabilities (Gersten, Compton, et al., 2009; NRP, 2000). Increasing the likelihood of teachers implementing research-based strategies in authentic school settings is a major goal of education leaders (Gersten, Chard, & Baker, 2000; Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). To address this issue in the context of ECRI, we developed highly specified lesson plans and teaching routines to support standard implementation of instruction and intervention materials. These routines provided clear expectations to teachers for what content to cover during instruction and intervention lessons and highly specified guidance for initiating explicit and engaging student–teacher interactions. Clear routines have been used in other fields to increase treatment integrity and adherence. For example, routines or "fidelity" checklists are regularly used in the field of medicine to reduce errors and complications, improve patient safety, reduce patient deaths from postsurgical complications, and increase efficiency in surgery (cf. Checklist Manifesto; Gawande, 2009). The approach of using highly specified instruction and intervention routines tied to a robust PD and coaching framework can also guide coaches and school leaders to define and measure implementation fidelity and to provide subsequent implementation goals for teachers.

In the context of intervention research, there is a call to identify the "active ingredients" of effective interventions (Carroll et al., 2007). Treatment intensity, as defined by Warren et al. (2007), provides a framework for considering the active ingredients of an intervention in so much as it captures *what* participants receive and *how* the intervention works at a molecular level. We used the COSTI instrument to observe and record teaching events in the context of explicit instruction (i.e., teacher models, guided practice opportunities, and corrective feedback), to measure instructional intensity. Coupled with other instruments (e.g., QEI), we were able to reliably document the quality and intensity of explicit instruction as an active ingredient in the ECRI intervention. These tools could likely assist other researchers explore the relationship between explicit and intensive instruction and student achievement in the context of intervention research.

Beyond their utility as tools for documenting hypothesized active ingredients in this particular research study and potentially other research projects, the COSTI and QEI instruments may also serve as blueprints to develop practitioner-friendly measures of the quality and intensity of explicit instruction in schools, especially those implementing multitiered approaches. With training and support, school leaders can reliably collect these kinds of implementation data on effective teaching behaviors and link these explicit teacher behaviors to the local student reading data.

Although project researchers provided rigorous training to support reliable collection of implementation data, the great majority of data collectors hired to observe in classrooms had previous experience as certified teachers. Thus, with similar training opportunities, districts could potentially use this type of implementation data to precisely describe highly interactive classrooms (e.g., judicious amounts of teacher modeling with high rates of student practice opportunities) and differentiate them from those classrooms that may be too didactic (e.g., high rates of teacher models and very low rates of student practice opportunities) or classrooms not supportive for at-risk learners (i.e., low rates of teacher modeling and high rates of practice opportunities with concomitant student errors).

Likewise, districts and schools can identify patterns of teacher behaviors that are associated with improved academic achievement or strong rates of academic growth. School and district leaders could use such data sets to develop goals for teachers with either consistently low student achievement or that are normatively discrepant from their peers in terms of effective teaching practices. In the current era of teacher accountability and the move toward models of teacher merit-based pay (Duncan & Murnane, 2011; Hanushek, 2011), the availability of formative measures, similar to the ones used in this study that provide data that can be used to support teachers implement evidence-based instruction (vs. punish teachers for poor performance), will become increasingly important. Because the success of a multitiered model is premised upon a solid core curriculum designed to support the majority of learners and principles of data-based decision making, the COSTI and QEI instruments are well suited for use by school leaders to support teacher use of evidence-based instructional practices.

Limitations

One of the primary limitations of the current ECRI study is that we are not able to determine whether teachers continue to implement evidence-based practices after 1 year of participation. Because the teachers are participating in a multiyear implementation of the ECRI intervention, we will not be able to measure maintenance of teaching behaviors until the final year of the project. Thus, we must wait until the conclusion of the 4-year ECRI study to determine whether teachers continue to implement explicit teaching behaviors to the degree observed during their participation in the formal intervention.

A second limitation in this study is the stability of a subset of teacher behaviors we measured with the COSTI instrument. Although many teacher behaviors were relatively stable across observations, results indicate a fair amount of variability within teachers across observation occasions for teacher models, teacher feedback, and student covert responses, which limit the precision of the estimates of the rates of these behaviors. More measurement occasions may be needed to increase the stability of these particular behaviors. A final issue that may limit the generalizability of our findings is the relatively modest sample size (i.e., 42 teachers nested within 16 schools) and the fact that all of the schools were sampled from districts in one state in the Pacific Northwest. We plan to replicate the analyses conducted in this study with a larger sample of teachers and schools. In a follow-up replication study, we are including approximately 100 teachers in 28 schools across two diverse regions in the United States.

Conclusion

Implementation of the systemic ECRI intervention increased the explicitness of instruction in meaningful ways. Teachers in the treatment condition provided higher quality explicit instruction and increased opportunities for choral practice in beginning reading skills when compared with teachers in the comparison condition. The findings in this study have important implications for research in terms of documenting the effectiveness of interventions for changing teacher behaviors, and for school leaders looking to generally improve teacher practice. At a broader level, it has implications for the current push for incorporating evidence-based practices in U.S. classrooms. However, the findings should be tested with a larger sample and should be further examined with respect to sustained implementation of explicit teaching behaviors.

Declaration of Conflicting Interests

The authors declared no potential conflict of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors 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 R324A090104 to the University of Oregon. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *Elementary School Journal*, *111*, 535–560.

Archer, A. L., & Hughes, C. A. (2011). Exploring the foundations of explicit instruction. In A. L. Archer, & C. A. Hughes (Eds.), *Explicit instruction: Effective and efficient teaching* (pp. 1–22). New York, NY: Guilford.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York, NY: Holt, Rinehart & Winston.

Baker, S., 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*, 1–20.

Baker, S. K., Fien, H., Santoro, L., Chard, D., & Park, Y. (in press). An evaluation of an explicit read aloud intervention taught in whole-classroom formats in first grade. *Elementary School Journal*.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., & Tilly, W.D. (2006). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. Witrock (Ed.), *The third handbook of research on teaching* (pp. 328–375). New York, NY: Macmillan.

Brownson, R. C., Baker, E. A., Leet, T. L., Gillespie, K. N., & True, W. R. (2011). *Evidence-based public health* (2nd ed.). New York, NY: Oxford University Press.

Carlisle, J. F., Kelcey, B., Rowan, B., & Phelps, G. (2011). Teachers' knowledge about early reading: Effects on students' gains in reading achievement. *Journal of Research on Educational Effectiveness*, *4*, 289–321.

Carnine, D. (1997). Bridging the research-to-practice gap. *Exceptional Children*, *63*, 513–521.

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

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*, 1–9.

Chard, D. J., & Jungjohann, K. (2006). *Scaffolding instruction for success in mathematics learning, intersection: Mathematics education sharing common grounds*. Houston, TX: Exxon-Mobil Foundation.

Chard, D. J., & Kame'enui, E. J. (2000). Struggling first-grade readers: The frequency and progress of their reading. *Journal of Special Education*, *34*, 28–38.

Chwalisz, K., & Dollinger, C. S. (2009). Evidence-based practice with family caregivers. In R. G. Frank, B. R. Caplan, & M. Rosenthal (Eds.), *Handbook of rehabilitation psychology* (2nd ed., pp. 301–312). Washington, DC: American Psychological Association.

Clarke, B., Smolkowski, K., Baker, S.K., Fien, H., & Chard (2011). The impact of a comprehensive tier 1 kindergarten curriculum on the achievement of students at-risk in mathematics. *Elementary School Journal*, *111*, 561-584.

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

Cook, B. G., & Cook, S. C. (2011). Unraveling evidence-based practices in special education. *Journal of Special Education*. Advance online publication.

Coyne, M. D., Kame'enui, E., & Carnine, D. (2006). *Effective teaching strategies that accommodate diverse learners* (3rd ed.). Upper Saddle River, NJ: Pearson.

Desimone, L. M., Porter, A. C., Garet, M., Yoon, K. S., & Birman, B. (2002). Does professional development change teachers' instruction? Results from a three-year study. *Educational Evaluation and Policy Analysis*, *24*, 81–112.

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity? Rising inequality, schools, and children's life chances*. New York, NY: Russell Sage.

Education Sciences Reform Act. (2002). *H.R. 3801—107th Congress*. Retrieved from http://www.govtrack.us/congress/bills/107/hr3801

Elstein, A. (2004). On the origins and development of evidence-based medicine and medical decision making. *Inflammation Research*, *53*, 184–189. doi:10.1007/s00011-004-0357-2

Ericsson, K. A., Roring, R. W., & Nandagopal, K. (2007). Giftedness and evidence for reproducibly superior performance: An account based on the expert performance framework. *High Ability Studies*, *18*, 3–56. doi:10.1080/13598130701350593

Fien, H., Smith, J. L. M., Smolkowski, K., Baker, S. K., Nelson-Walker, N. J., & Chaparro, E. A. (in press). An examination of the efficacy of a systemic intervention in early reading. *Journal of Learning Disabilities*. Manuscript under review.

Fixsen, D. L., Naoom, S. F., Blase, K., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI publication No. 231). Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute.

Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal*, *38*, 653–689.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (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.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., & Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance.

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

Gersten, R. M. (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. *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* (No. NCE 2009-4045). Washington, DC: National Center for Education Evaluation.

Gersten, R. M., Dimino, J., Jayanthi, M., Kim, J., & Santoro, L. (2009). *An investigation of the impact of the teacher study group as a means to enhance the quality of reading comprehension and vocabulary instruction for first graders in reading first schools: Technical report*. Los Alamitos, CA: Instructional Research Group.

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.

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

Guyatt, G. H., Kirshner, B., & Jaeschke, R. (2009). Measuring health status: What are the necessary measurement properties? *Journal of Clinical Epidemiology*, *45*, 1341–1345.

Hanushek, E. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*, 466–479.

Harn, B., Linan-Thompson, S., & Roberts, G. (2008). Intensifying instruction. *Journal of Learning Disabilities*, *41*, 115–125.

Heck, D. J., Banilower, E. R., Weiss, I. R., & Rosenberg, S. L. (2008). Studying the effects of professional development: The case of the NSF's local systemic change through teacher enhancement initiative. *Journal for Research in Mathematics Education*, *39*, 113–152.

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

Howell, K. W., & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making* (3rd ed.). Belmont, CA: Wadsworth.

Kent, S. C., Wanzek, J., & Al Otaiba, S. (2012). Print reading in general education kindergarten classrooms: What does it look like for students at-risk for reading difficulties? *Learning Disabilities Research & Practice*, *27*, 56–65.

Klingner, J. K., Vaughn, S., Hughes, M. T., & Arguelles, M. E. (1999). Sustaining research-based practices in reading: A 3-year follow-up. *Remedial and Special Education*, *20*, 263–274.

Kratochwill, T. R., & Shernoff, E. S. (2004). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Quarterly*, *18*, 389–408.

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

Marsh, J. A., McCombs, J. S., Lockwood, J. R., Martorell, F., Gershwin, D., Naftel, S., & . . .Crego, A. (2008). *Supporting literacy across the sunshine state*. Santa Monica, CA: RAND.

National Center for Education Statistics. (2011). *National Assessment of Educational Progress (NAEP):* 2011 Reading Assessment. Institute of Education Sciences, U.S. Department of Education.

National Reading Panel. (2000). *Report of the National Reading Panel*. Washington, DC: National Institute of Child Health and Human Development.

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

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.

Odom, S. L. (2009). The tie that binds. *Topics in Early Childhood Special Education*, *29*, 53–61.

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *23*, 431–451.

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

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). Boston, MA: McGraw-Hill.

Rupley, W. H., Blair, T. R., & Nichols, W. D. (2009). Effective reading instruction for struggling readers: The role of direct/explicit teaching. *Reading and Writing Quarterly*, *25*, 125–138. doi:10.1080/10573560802683523

Seftor, N., Constantine, J., Cody, S., Ponza, M., Knab, J., Deke, J., & Monahan, S. (2011). *What works clearinghouse: Procedures and standards handbook 2011* (NCEE 2011-XXXX). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Shoukri, M., Asyali, M., & Donner, A. (2004). Sample size requirements for the design of reliability study. *Statistical Methods in Medical Research*, *13*, 251–271.

Silverman, S. (1985). Relationship of engagement and practice trials to student achievement. *Journal of Teaching in Physical Education*, *5*, 13–21.

Simmons, D. C., Coyne, M. D., Kwok, O. M., McDonagh, S., Harn, B. A., & Kame'enui, E. J. (2008). Indexing response to intervention a longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities*, *41*, 158–173.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

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 Academy Press.

Sterman, J. D. (2006). Learning from evidence in a complex world. *American Journal of Public Health*, *96*, 505–514. doi:10.2105/AJPH.2005.066043

Sutherland, K. S., & Wehby, J. H. (2001). Exploring the relation between increased opportunities to respond to academic requests and the academic behavioral outcomes of students with emotional and behavioral disorders: A review. *Remedial and Special Education*, *22*, 113–121.

Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E., & Francis, D. J. (2006). Section on teaching, learning, and human development: Effectiveness of a Spanish intervention and an English intervention for English learners at risk for reading problems. *American Educational Research Journal*, *43*, 449–489.

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

Walker, H. M. (2004). Use of evidence-based intervention in schools: Where we've been, where we are, and where we need to go. *School Psychology Review*, *33*, 398–407.

Walker, H. M., & Gresham, F. (2003). School related behavior disorders. In W. Reynolds, & G. Miller (Eds.), *Handbook of psychology: Educational psychology* (Vol. 7, pp. 511–530). New York, NY: Wiley.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*, 70–77.

Watkins, C., & Slocum, T. (2004). The components of direct instruction. In N. E. Marchand-Martella, T. A. Slocum, & R. C. Martella (Eds.), *Introduction to direct instruction* (pp. 28–65). Boston, MA: Allyn & Bacon.

What Works Clearinghouse. (2011). *What Works Clearinghouse: Procedures and standards handbook (Version 2.1)*. Washington, DC: National Center for Education Evaluation.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues and Answers Report, REL 2007 No. 033). Washington, DC: U.S. Department of Education.