



CONSORTIUM FOR POLICY RESEARCH IN EDUCATION

Evaluation of the i3 Scale-up of Reading Recovery

Year One Report, 2011-12

Henry May
Abigail Gray
Jessica N Gillespie
Philip Sirinides
Cecile Sam
Heather Goldsworthy
Michael Armijo
Namrata Tognatta

August 2013
RR-76

A COLLABORATIVE PUBLICATION BETWEEN



Center for Research in
Education & Social Policy

The research reported here was supported by the Office of Innovation and Improvement (OII), U.S. Department of Education, through Grant #U396A100027 to The Ohio State University, and in part by the Institute of Education Sciences (IES), U.S. Department of Education, through Grant #R305B090015 to the University of Pennsylvania. The opinions expressed are those of the authors and do not represent the views of OII, IES, or the U.S. Department of Education.

About Consortium for Policy Research in Education (CPRE)

Since 1985, the Consortium for Policy Research in Education (CPRE) has brought together renowned experts from major research universities to improve elementary and secondary education by bridging the gap between educational policy and student learning. CPRE researchers employ a range of rigorous and innovative research methods to investigate pressing problems in education today. Having earned an international reputation for quality research and evaluation, CPRE researchers have extensive experience conducting experimental studies, large-scale quasi-experimental research, qualitative studies, and multi-state policy research.

CPRE's member institutions are the University of Pennsylvania; Teachers College Columbia University; Harvard University; Stanford University; University of Michigan; University of Wisconsin-Madison; and Northwestern University.

About the Center for Research in Education & Social Policy (CRESP)

The Center for Research in Education and Social Policy (CRESP) within the College of Education and Human Development at the University of Delaware conducts rigorous research to help policymakers and practitioners in education, health care and human services determine which policies and programs are most promising for improving outcomes in children, youth, adults and families.

Although research in prevention sciences and health care have long used rigorous designs to assess the effectiveness of programs, it was not until the Education Sciences Reform Act of 2002 that we witnessed a dramatic increase in the quantity and quality of research to evaluate the effects of education programs and policies. The education community began to focus on research that could measure the impact of these programs through randomized experiments and other research designs that support causal conclusions and can determine whether, how well, for whom, and why new programs and interventions work.

CRESP specializes in experimental and quasi-experimental research that uses quantitative and mixed methods to evaluate how and how well programs and interventions work to improve educational, family, and health outcomes in schools and communities.

Table of Contents

1 Executive Summary

5 Introduction

- 5 Overview of the Reading Recovery i3 Scale-Up Evaluation
- 7 Reading Recovery: Theory, History, and Research
- 15 Reading Recovery Structure, Logic Models, and Theory of Action

25 Evaluation Study Design and Methods

- 25 Primary Research Questions
- 26 Evaluation Timeline
- 27 Quantitative Impact Studies of i3 Scale-up
- 32 Exploring and Explaining Variation in Impact Estimates
- 35 Overview of the Mixed Methods Evaluation of Implementation Under the i3 Scale-Up
- 35 Qualitative Evaluation of i3 Scale-Up
- 41 Survey Instruments and Derived Measures of Implementation Fidelity
- 44 RR teacher Activity Logs
- 44 Reading Recovery Intervention Document Review
- 45 Case Studies

51 Early Impacts on Student Achievement

- 51 School Participation and Data Availability
- 53 Baseline Balance Tests
- 54 Overall Impacts on ITBS Reading Scores
- 58 School-Level Variation in Overall Impacts Estimates
- 59 Impacts on ITBS Reading Scores in Rural Schools and for English Language Learners



63 Fidelity to Reading Recovery Standards and Guidelines

67 Early Implementation Findings: Recruitment

- 68 i3 Scale-up Recruitment Outcomes
- 73 i3 Recruitment Challenges
- 79 Support for Recruitment
- 81 Support to Teacher Leaders
- 82 Future Plans for Recruitment
- 85 Conclusion: Recruitment

87 Early Implementation Findings: Training

- 88 RR teacher Training: Year 1
- 91 Challenges to Training

101 Early Implementation Findings: School-Level Implementation

- 102 School-Level Implementation: Findings from Case Studies
- 104 School-level Implementation: Roles & Relationships

137 Conclusion: Changing students. Changing teachers. Changing schools?

142 Appendix A: Statistical Model for Impacts of Reading Scores

143 References

Executive Summary

Reading Recovery (RR) is a short-term early intervention designed to help the lowest-achieving readers in first grade reach average levels of classroom performance in literacy. Students identified to receive Reading Recovery meet individually with a specially trained Reading Recovery (RR) teacher every school day for 30-minute lessons over a period of 12 to 20 weeks. The purpose of these lessons is to support rapid acceleration of each child's literacy learning. In 2010, The Ohio State University received a Scaling Up What Works grant from the U.S. Department of Education's Investing in Innovation (i3) Fund to expand the use of Reading Recovery across the country. The award was intended to fund the scale-up of Reading Recovery by training 3,675 new RR Teachers in U.S. schools, thereby expanding capacity to allow service to an additional 88,200 students.

The Consortium for Policy Research in Education (CPRE) was contracted to conduct an independent evaluation of the i3 scale up of Reading Recovery over the course of five years. The evaluation includes parallel rigorous experimental and quasi-experimental designs for estimating program impacts, coupled with a large-scale mixed-methods study of program implementation under the i3 scale-up. This report presents findings through the second year of the evaluation. The primary goals of this evaluation were: a) to assess the success of the scale-up in meeting the i3 grant's expansion goals; b) to document the implementation of scale-up and fidelity to program standards; and, c) to provide experimental evidence of the impacts of Reading Recovery on student learning under this scale-up effort.

This document is the first in a series of three annual reports produced based on our external evaluation of the Reading Recovery i3 Scale-Up. This report presents early results from the experimental impact and implementation studies conducted over the 2010-11 and 2011-12 school years.

In order to estimate the impacts of the program, a sample of first graders who met the Reading Recovery target criteria for selection were randomly assigned to either receive Reading Recovery, or to continue receiving classroom instruction. The reading achievement of students in this sample was assessed using an externally standardized assessment of reading achievement - the Iowa Tests of Basic Skills (ITBS) - in order to determine the impact of Reading Recovery lessons on low-achieving students who were randomly assigned to receive the intervention. The data for the implementation study include extensive interviews and surveys with stakeholders at multiple levels: individuals at the i3 office, University Training Center directors, district-level site coordinators, teacher leaders, RR teachers, principals, and 1st-grade teachers. Case studies were also conducted with nine i3 scale-up schools to observe how Reading Recovery operates in differing contexts.

Key findings from the first two years of this five-year evaluation include the following:

Impacts on Student Reading Performance

- » Treatment students who participated in Reading Recovery outperformed students in the control group on each subscale of the ITBS Reading test.
 - » The mean of Reading Recovery students' posttest ITBS Total Reading scores was at the 36th percentile nationally, while students in the control group had posttest scores at the 18th percentile—a difference of +18 percentile points.
 - » The mean of Reading Recovery students' posttest ITBS Reading Words scores was at the 43rd percentile nationally, while students in the control group had posttest scores at the 27th percentile—a difference of +16 percentile points.
 - » The mean of Reading Recovery students' posttest ITBS Reading Comprehension scores was at the 39th percentile nationally, while students in the control group had posttest scores at the 19th percentile—a difference of +20 percentile points.
- » The estimated standardized effect of Reading Recovery on students' ITBS Total Reading Scores was .68 standard deviations relative to the population of struggling readers eligible for Reading Recovery under the i3 scale-up, and .47 standard deviations relative to the nationwide population of all first graders. These standardized effect sizes are large relative to typical effect sizes found in educational evaluations.
- » Effect estimates were similarly large for both the ITBS Reading Words and Reading Comprehension subscales.
- » The impact estimates of Reading Recovery vary substantially across schools, with most schools having moderate to large positive impact estimates (greater than .40 standard deviations).

Recruitment Successes and Challenges

- » Recruitment of schools during the first two years was 44 percent above the intended goal, recruitment of RR teachers reached 76 percent of the intended goal, and instruction of students reached 88 percent of the intended goal.

Key barriers to recruitment included the late start of the i3 project in Year 1 (after the school year had already begun), and concerns about the current economic climate. Scale-up grant funds are intended for use to train RR teacher leaders and teachers, however districts and schools are still responsible for paying RR teacher salaries. Thus, many are hesitant to hire additional staff or implement a new intervention during an economic downturn.

Training and Support of New RR teachers

- » The behind-the-glass training sessions were described as a very valuable part of RR teachers' preparation. However, logistic constraints sometimes made it difficult to conduct as many sessions as required by Reading Recovery.

RR teacher leaders provided critical feedback and intensive support to new RR teachers as they implemented the intervention during their first year.

School-Level Implementation Fidelity

- » School-level implementation of Reading Recovery was, in most respects, faithful to the Reading Recovery Standards and Guidelines.
- » There was strong fidelity to standards in the execution of Reading Recovery lessons, but less fidelity to the requirement to formally document each lesson.
- » Data from surveys, activity logs, interviews, and observations reveal that the job of a RR teacher is extremely demanding, with very busy schedules and many long days.

While most schools used the Reading Recovery Observation Survey (OS) to identify low-performing students, the students selected to receive Reading Recovery instruction were not always the students who scored the lowest on the OS.

CHAPTER 1

Introduction

As part of the 2010 economic stimulus, Reading Recovery was awarded a \$45 million “Investing in Innovation” (i3) grant from the U.S. Department of Education, along with an additional \$10.1 million from private sources, to fund the scale-up of Reading Recovery across the nation. Reading Recovery was one of only four programs awarded an i3 scale-up grant in 2010, due in part to the availability of sufficiently rigorous evidence of prior impacts to justify such a large investment. The grant was intended to expand Reading Recovery in more than 2,000 schools and provide targeted literacy assistance to over 88,000 students.

This evaluation report is the first in a three-part series that will present results from the external evaluation of the Reading Recovery i3 Scale-Up. This first report describes the evaluation research design for the five-year project, and presents the first set of results on the implementation and impacts of the Reading Recovery program.

Overview of the Reading Recovery i3 Scale-Up Evaluation

Research Objectives

- » This evaluation of the Reading Recovery (RR) i3 scale-up is focused on a set of research objectives that emphasize RR-related recruitment, training, implementation, and impacts on educators and students participating in the i3 scale-up initiative. This study addresses only schools, sites, and individuals involved with the i3 scale-up, not all those who participate in Reading Recovery generally. In this first evaluation report we will address the following research objectives:

Recruitment

- » Document the processes of district, school, RR teacher leader, and RR teacher recruitment for the i3 scale-up.
- » Document the processes used by The Ohio State University (OSU) and partner universities to monitor the recruitment of these groups to participate in the scale-up.

Training

- » Document the training and preparation of RR teachers to carry out the responsibilities associated with their roles.
- » Investigate variation in training and support provided across training centers and school sites involved with the i3 scale-up.

Implementation

- » Explore the roles of RR teacher leaders, RR teachers, and other stakeholders in the implementation of RR under the i3 scale-up.
- » Examine the school-level factors that support or hinder RR implementation.
- » Investigate the extent to which the implementation activities of RR teachers, RR teacher leaders, and site coordinators during the i3 scale-up adhere to the Standards and Guidelines of Reading Recovery in the United States (5th Edition) (RRCNA, 2009).
- » Document differences across schools and districts in terms of student selection and instructional practices.

Impacts

- » Estimate RR's impacts on the early literacy competence of 1st-grade students selected for the program during the 2011-12 school year, as measured by the Iowa Tests of Basic Skills (ITBS).
- » Estimate variation in impact estimates and describe the distribution of impacts across schools.
- » Estimate the impacts of RR on two subgroups of particular interest under the i3 scale-up: English language learners (ELLs) and students enrolled in rural schools.

The research objectives listed here do not include all of the objectives set forth for this five-year evaluation. The goal of the first year was to accurately and rigorously assess the main inputs, outputs, and outcomes of the scale-up, which will enable deeper analysis and more targeted data collection in future years. For example, future research will include cross-case analyses of case study schools over multiple years; data pooling across cohorts to estimate RR impacts on additional subgroups of students (e.g., economically disadvantaged students); and investigation of if and how variation in program impacts is associated with school context factors and implementation fidelity.

Reading Recovery: Theory, History, and Research

Instructional initiatives that are able to identify students at greatest risk and help them achieve grade-level proficiency are essential to addressing the epidemic of low literacy levels in the U.S. (U.S. DOE NCES, 2010; NAEP Data Explorer; Annie E. Casey Foundation, 2010). Several studies have shown that most students who leave first grade reading below grade-level never catch up (Juel, 1988; Lyon et al., 2001; Shaywitz et al., 1999). Others suggest that literacy interventions are most effective if they are instructionally rigorous and focused on the early years of schooling (Strickland, 2002; National Research Council, 1998; Johnston, Allington & Afflerback, 1985; Morrow, 1993; Pikulski, 1994). However, schools often lack the structures or expertise to implement an intervention with the intensity required to reverse low literacy. Reading Recovery was designed in direct response to such challenges. Developed in the 1970s by educator and psychologist Marie Clay, Reading Recovery is one of oldest and most widely implemented literacy interventions. From its roots in New Zealand, Reading Recovery has spread to Australia and countries in North America and Europe.

This report begins with a description of the key features of the RR program, including its three-tiered structure, followed by a discussion of RR as it specifically relates to the i3 scale-up grant. The remainder of this chapter reviews prior evidence of RR impacts on student achievement, and presents the theory of action and logic models guiding this evaluation. Chapter 2 details the research methods used to understand the implementation process and measure impacts under the scale-up. Chapter 3 describes the findings of the student impact analysis, and Chapter 7 discusses emerging themes from our first year of field-based case study research. Chapters 4-7 present findings on scale-up program oversight, recruitment, training, and implementation. Finally, Chapter 8 outlines conclusions regarding program implementation, impacts on student performance, general recommendations, and plans for the evaluation in subsequent years.

Overview of Reading Recovery

Reading Recovery is an intensive intervention targeting the lowest-achieving 15-20 percent of 1st-grade readers. It takes as its underlying principle the idea that individualized, short-term, highly responsive instruction delivered by an expert can disrupt the trajectory of low literacy achievement, produce accelerated gains, and enable students to catch up to their peers and sustain achievement at grade level into the future. Reading Recovery attends to phonemic awareness, phonics, vocabulary, fluency, and comprehension—the critical elements of literacy and reading instruction identified by the National Reading Panel (2000). Ultimately, Reading Recovery is intended to help students develop a set of self-regulated strategies for problem-solving words, self-monitoring, and self-correcting that they can apply to the interpretation

of text. These strategies focus on enabling students to use meaning, structure, letter-sound relationships, and visual in their reading and writing processes (Clay, 1991; 2005c). Reading Recovery's model is based on theory that asserts that once equipped with these strategies for independent processing, struggling readers can achieve at average levels and maintain proficiency in the regular classroom without special intervention.

Literature on Reading Recovery reports that approximately 75 percent of students—both native English speakers and ELLs—typically reach grade-level proficiency within their 12- to 20-week instructional cycle and go on to maintain their progress throughout elementary school (Rodgers, Wang, & Gómez-Bellengé, 2004; Rodgers, Gómez-Bellengé, Wang, & Schultz, 2005). An additional aim of the RR model is to improve the ongoing literacy instruction received by those students who do not reach grade-level proficiency. One way to achieve this goal is to streamline the Special Education referral process for those students who fail to make adequate progress during their cycle of RR instruction. The diagnostic, achievement, and progress information collected by the in the course of their instruction is critical to this process.

According to the program model of Reading Recovery, highly trained teachers provide daily instruction to students during 30-minute, one-on-one teaching sessions. Teachers tailor their lessons to a student's individual strengths and needs based on their own observations. According to Clay (2004a), RR teachers "must be able to design a superbly sequenced series of lessons determined by the particular children's competencies, and make highly skilled decisions [at each] moment during the lesson" (p. 23). A primary assumption of the Reading Recovery model is that high-quality instruction is key to accelerated progress in literacy learning. Ensuring teacher quality is therefore considered essential to the success of the intervention, and the RR model's three-tiered structure emphasizes instructional quality at each tier.

At the first tier of the RR model, faculty at partner universities train RR teacher leaders through an in-residence, post-graduate program of study (see Chapter 6 for a full description). Through this training, as well as through mentorship experiences with faculty and experienced teacher leaders, new RR teacher leaders are expected to become expert literacy coaches with a deep understanding of how children learn and become literate, and why some children have great difficulty learning to read and write. Teacher leaders also become expert teachers of both children and adult learners; they are taught how to design and deliver literacy lessons to individual students with a focus on those students who are making particularly slow progress, and also how to deliver training and support to RR teachers.

Developing these core understandings about learning and instruction prepares RR teacher leaders for their role in the Reading Recovery model's second tier of instruction. At this tier, RR teacher leaders provide a year-long academic program of training and professional development to the school-based RR teachers who are simultaneously working directly with students. RR teachers are chosen based on their past history of instructional success, and evidence of their

.....

adaptability and problem-solving skills. Under RR teacher leader supervision, RR teachers complete graduate coursework that emphasizes the skills that are critical for individualized, highly responsive instruction, including theory of literacy acquisition, assessment, systematic observation, instruction, analysis, and reflection. As they continue past their first year, RR teachers receive ongoing professional development and supervision from their RR teacher leader to ensure continual learning and growth. At the third tier of instruction in the RR model, trained RR teachers design and deliver individualized lessons to students identified to receive RR services.

Although it is often described as first and foremost a one-on-one tutoring intervention, Reading Recovery has always been conceptualized by Clay as a system-level intervention. She describes Reading Recovery as “much more than merely a particular theory or analysis of what is needed to help children who have difficulty reading” (Clay 1987, p. 36). Clay also describes four dimensions of change that are required to “dramatically reduce the number of children with reading and writing difficulties in an education system” (Clay 1987, p.36). These changes include: behavioral change on the part of teachers; child behavior change achieved by teaching; organizational change in schools achieved by teachers and administrators; and, social/political change in funding by controlling authorities.

Underlying this conceptualization is the need for a “change agent” (Fullan, 1993) who will support school systems as they engage in and attempt to sustain the RR reform. In the program model, RR teacher leaders are generally expected to act as these change agents. With support from the Reading Recovery Council of North America (RRCNA) and faculty at their site’s partner university, RR teacher leaders are expected to enable a process that encourages schools and districts to:

- » Adopt the view that investing in young children who are just beginning to experience reading failure conserves resources by ensuring that reading problems are addressed early, rather than remediated later.
- » Invest in ongoing teacher professional development.
- » Expect students who are struggling to learn to read and write to make accelerated progress, and to reach the average of their class, regardless of the reason for the reading difficulty.

The Standards and Guidelines of Reading Recovery in the United States (RRCNA, 2009) codify Clay’s assumptions and understandings of the systemic change necessary to support implementation of the RR model, and make explicit the practices that constitute adherent implementation. Thus, the Standards and Guidelines formalize the processes by which individual actors and systems achieve the dimensions of change advocated by Clay, and facilitate the implementation of Reading Recovery with fidelity across widely dispersed U.S. sites.

The Reading Recovery Program 1984 to Present

Reading Recovery was first introduced in the U.S. in 1984 through a collaboration between Reading Recovery's New Zealand leadership, faculty at OSU, and a group of public school teachers in Columbus, Ohio. The success of this collaboration led to funding from the Ohio Department of Education, and Reading Recovery quickly spread to additional schools in Ohio (Lyons, 1998). The program received a boost in 1987, when the U.S. Department of Education's National Diffusion Network recognized Reading Recovery as an "exemplary research-based program" (Lyons, 1998, p. 77), and offered federal funds to support expansion of the intervention. By the early 1990s, Reading Recovery was operating in 48 states.

Following its rapid growth in the U.S. during the late 1980s and 1990s, Reading Recovery reached its peak in 2000-01, serving 152,000 students. Then, in the early 2000s, the program encountered a significant obstacle. Under the Reading First grant program—a cornerstone of the No Child Left Behind Act of 2001—nearly \$1 billion in annual funds for K-3 literacy instruction was made available to states. Unfortunately, the implementation of the Reading First program by the U.S. Department of Education involved numerous processes that resulted in conflicts of interest and bias against specific reading programs (OIG, 2006). More specifically, the investigation report by the Office of the Inspector General concluded that the Department:

- » Developed an application package that obscured the requirements of the statute;
- » Took action with respect to the expert review panel process that was contrary to the balanced panel composition envisioned by Congress;
- » Intervened to release an assessment review document without the permission of the entity that contracted for its development;
- » Intervened to influence a State's selection of reading programs; and
- » Intervened to influence reading programs being used by local educational agencies (LEAs) after the application process was completed.

Findings in the Inspector General's report that pertained specifically to Reading Recovery suggested that Department staff actively discouraged the inclusion of Reading Recovery in states' Reading First applications. This was evident in communications from the Reading First Director regarding the selection of expert panelists who held opinions on reading instruction that were explicitly contrary to Reading Recovery.

...a Department employee reported to the Reading First Director that the Department had received a question from a member of the media about the panel composition. The response by the Reading First Director suggests that he

.....

may indeed have intended to “stack” the expert review panel. The employee stated: “The question is...are we going to ‘stack the panel’ so programs like Reading Recovery don’t get a fair shake[?]” The Reading First Director responded, “Stack the panel?...I have never *heard* of such a thing... <harumph, harumph>” (OIG, 2006, p. 18).

A few days before the Department publicly announced the panelists it had chosen to serve, one of the Department-nominated panelists contacted the Reading First Director and shared his strong bias against Reading Recovery and his strategy for responding to any State that planned to include Reading Recovery in its application. The Reading First Director responded: “I really like the way you’re viewing/approaching this, and not just because it matches my own approach :-), I swear!” This individual later served as the panel chair for the subpanel that reviewed Wisconsin’s State application and in response to the State’s plans to use Reading Recovery, he included an 11-page negative review of Reading Recovery in his official comments on the application (OIG, 2006, p. 18).

The investigation also uncovered evidence of direct communication with states that suggested bias towards Reading Recovery.

State officials [from Kentucky] informed us that they spoke with the Reading First Director in a conference call and informed him that they believed Reading Recovery and Rigby were sufficiently SBRR. State officials said that the Reading First Director told them he had concerns about Reading Recovery and urged them not to use Reading First funds on the program. The State officials said that they asked for this request in writing, but the Reading First Director told them that he would not do so and invited them to defend the two programs instead. The State officials subsequently provided support for the programs in writing, but they informed us that they did not receive a response from the Department (OIG, 2006, p. 26).

Given this evidence of explicit bias against Reading Recovery by those directly involved in the solicitation, review, and approval of states’ applications for the Reading First program, implementation of Reading Recovery began to regress during the Reading First era. Between 2001 and 2008, the number of students participating in Reading Recovery each year decreased by nearly 40 percent.

In 2008, Reading Recovery’s U.S. trajectory was again altered by a federal initiative when the U.S. Department of Education What Works Clearinghouse released its report on Reading Recovery. Based on a meta-analysis of research on RR’s effectiveness, the What Works Clearinghouse reported “positive effects on alphabetic and general reading achievement, and potentially

positive effects on fluency and comprehension” (What Works Clearinghouse, 2008). With this, Reading Recovery became the only program among 171 reviewed (as of fall 2011) to receive “positive” or “potentially positive” ratings across all four targeted domains, making it the highest-rated early literacy program overall. Reading Recovery personnel recount that the What Works Clearinghouse report has been an important tool in their efforts to reinvigorate Reading Recovery’s U.S. expansion on the heels of Reading First.

In 2010, OSU received a Scaling Up What Works grant from the U.S. Department of Education’s Investing in Innovation (i3) Fund. This grant signaled the start of another important growth phase for Reading Recovery. The \$45 million award, supplemented by an additional \$9 million in private contributions, was approved to fund the scale-up of Reading Recovery’s model in thousands of U.S. schools. Key goals for the scale-up included the training of 15 new RR teacher leaders and 3,675 new RR teachers, with priority placed on training teachers in low-performing schools, rural schools, and schools with significant numbers of ELLs. All told, the scale-up proposed to bring the intervention to an additional 1,470 high-need schools, and serve 88,200 new students. Another related goal of the scale-up was to learn if and how the effectiveness of Reading Recovery is impacted when the program is implemented on a large scale in natural settings.

Prior Research on Reading Recovery

Reading Recovery has been the focus of considerable study over the course of its nearly 30-year history. Research has focused on various issues, ranging from the merits of the intervention’s instructional approach (Iverson & Tunmer, 1993; Moats, 2007; Pinnell et al., 1994; Chapman, Tunmer & Prochnow, 1999; Tunmer & Chapman, 2001) to its cost-effectiveness (Dyer & Binkney, 1995; Hiebert, 1994; Rasinski, 1995; Gómez-Bellengé, 2002). Most relevant to the current study is the subset of the Reading Recovery research that examines the intervention’s efficacy in terms of its goal of producing accelerated and long-lasting gains in students’ literacy achievement. This literature includes research on both the extent of Reading Recovery’s impact on student achievement and the persistence of students’ gains over time.

A significant number of the efficacy studies focused on short-term outcomes have arrived at the same basic conclusion: that by the end of the 12- to 20-week intervention cycle, most students who participate in Reading Recovery achieve at a level equivalent to the average of their classmates in reading and writing (Ashdown & Simic, 2000; Allington, 2005; Center et al., 1995; D’Agostino & Murphy, 2004; Pinnell et al., 1994; Schwartz, 2005; Pinnell, 1989; Quay et al., 2001; Rodgers et al., 2004; Rodgers et al., 2005). Reading Recovery’s internal research corroborates these findings: The International Data and Evaluation Center (IDEC) based at OSU collects, analyzes, and reports on data on nearly two million children served by Reading Recovery in the U.S. These data reveal that 75 percent of the students who have completed Reading Recovery since its introduction in 1984 have met grade-level standards in reading and writing by the end of the intervention (<https://www.idecweb.us/>).

From within this group of efficacy studies, findings regarding the size of Reading Recovery's effects vary widely. A number of studies reported large to very large effect sizes (Schwartz, 2005; Rodgers, Wang, & Gómez-Bellengé, 2004). Others found smaller treatment effects or none at all (Tunmer & Chapman, 2002). Shanahan & Barr (1995) found gains in Reading Recovery students, but indicated that these gains were smaller than anticipated and in fact only slightly larger than the gains made by other low-achieving students in the same schools. Elbaum et al. (2001) found that trained volunteers were able to elicit similar progress in struggling readers to that achieved by RR teachers. In their 2004 meta-analysis, D'Agostino and Murphy identified a total of 1,379 different effect sizes in the 36 Reading Recovery studies they analyzed.

Additional research has examined the persistence of Reading Recovery students' literacy growth over time. These findings are also mixed. Some studies suggest long-lasting impacts from the intervention (Wasik & Slavin, 1993; Brown, Denton, Kelly, & Neal, 1999; Rowe, 1995; Schmitt & Gregory, 2005; Pinnell, 1989; Briggs & Young, 2003; Ruhe & Moore, 2005). For instance, a five-year study of 760 California students who participated in Reading Recovery between 1993 and 1998 found that "more than three-fourths of the children who successfully completed Reading Recovery achieved standardized test scores in the average or above average range" in grades 2-5 (Brown et al., 1999, p. 10). Similarly, an Indiana study found that "one, two, and three years beyond the intervention, Reading Recovery children were performing roughly as well as or better than their cohort sample peers on the task of oral text reading" (Schmitt & Gregory, 2005, p. 1).

Other research suggests that Reading Recovery students do not consistently maintain their gains (Tunmer & Chapman, 2002; Shanahan & Barr, 1995). Hiebert (1994) examined the literacy achievement of 4th-grade students who had completed Reading Recovery, compared with a comparison group. While she found that the Reading Recovery group's overall performance suggested little long-term growth, she noted that the two groups' levels of proficiency varied "as a function of the [reading] task" (p. 21), with RR students performing better on oral reading, but showing no difference from the comparison groups on tasks requiring comprehension or recognition of nonsense and unfamiliar words.

Despite the volume of research on Reading Recovery's short- and long-term impacts, questions linger as to the quality of the literature overall. Several reviews note that factors inherent to the intervention and the students it serves have consistently confounded researchers' efforts to implement rigorous research designs (D'Agostino & Murphy, 2004; Shanahan & Barr, 1995; Pinnell, 1997). D'Agostino and Murphy (2004) posit that Reading Recovery "has proven a very difficult program to evaluate, given its student selection and attrition policies, the barriers to locating an equivalent comparison group, the reliance on outcome measures designed for the program, and the problems inherent with accurately measuring students' achievement levels in first grade" (p. 24).

In part as a result of these inherent challenges, the literature on Reading Recovery's impact has been criticized for a lack of methodological rigor, with scholars noting a range of issues. In their systematic review of all U.S.-based studies that employed a pretest-posttest design to measure the effectiveness of Reading Recovery, Shanahan and Barr (1995) observed that most studies contained some selection bias that inflated impact estimates. As one example, Reading Recovery students who received fewer than 60 lessons were excluded from one study's impact analysis. Similarly, Torgeson (2000) noted that data on children who do not complete the Reading Recovery intervention—or do not complete it successfully—are often omitted from studies. Rasinski (1995) questioned the extent to which studies adequately explored alternative explanations for observed effects.

Others have raised concerns about the measures used in the evaluation of Reading Recovery. Wasik & Slavin (1993) observed that posttests used to assess students' gains are often closely aligned with the intervention itself, thus predisposing students with Reading Recovery experience to perform well. Tunmer and Chapman (2001) concluded that the most important outcome measure of the Reading Recovery program, text reading level, was highly unreliable and yielded inflated estimates. Further, a meta-analysis on one-on-one tutoring programs conducted by Elbaum (2000) supported the finding that text reading level produced larger effect size estimates for Reading Recovery than any other measure.

The 2008 report on Reading Recovery by the What Works Clearinghouse underscored the shortcomings of much of the extant literature (U.S. Department of Education, Institute of Education Sciences, 2008). Of the 106 studies reviewed for the What Works Clearinghouse's original analysis, just five were determined to meet IES' evidence standards. Four of these five were randomized control trials (Baenen et al., 1997; Pinnell et al., 1988; Pinnell et al., 1994; Schwartz, 2005). In each case, a treatment group of 1st-grade students who received Reading Recovery services was compared with a control group of first graders who did not participate in the intervention. Treatment and control were randomly assigned. The fifth study was quasi-experimental; the treatment and control groups were not assigned randomly. The studies differed from one another in terms of assessment strategy and sample size and dispersion, as well as the scope of the research.

Nearly half of the Reading Recovery studies reviewed by IES were excluded from the What Works Clearinghouse analysis because of their failure to establish baseline balance between the treatment and control groups. This issue was also raised in D'Agostino and Murphy's meta-analysis:

... We found that all pretest effect sizes [in a subset of the studies analyzed] were negative and significantly different from zero, indicating that RR students (discontinued, not-discontinued, and combined) indeed scored significantly lower on all pretest measures than low-achieving and regular students. Thus,

.....

one cannot rule out that at least some of the observed RR effects have been the result of a selection by regression artifact. Given that a fundamental goal of RR is to serve students the most in need, many studies on RR failed to construct equivalent groups (2004, p. 34)

The second most common reason Reading Recovery studies were excluded from the What Works Clearinghouse analysis was the lack of a comparison group, an issue that occurred in roughly 20 percent of the 106 studies reviewed.

On the basis of the empirical impact analyses included in the five studies that met IES' evidence standards, the 2008 report on Reading Recovery found positive effects in two of the key literacy indicators assessed by IES, alphabets and general reading achievement. "Potentially positive effects" were identified in the other two areas, fluency and comprehension. The report's authors write: "Based on these five studies, the What Works Clearinghouse considers the extent of evidence for Reading Recovery to be medium to large for alphabets, small for fluency and comprehension, and medium to large for general reading achievement" (What Works Clearinghouse, 1998).

While on the whole very positive, the What Works Clearinghouse report's findings—coupled with the shortage of studies that met IES' evidence standards—strongly suggest the need for a large-scale and highly rigorous examination of the impacts of Reading Recovery and the implementation factors that underlie variations in impact.

Reading Recovery Structure, Logic Models, and Theory of Action

The implementation of Reading Recovery requires coordination among stakeholders at several different levels (e.g. universities, districts, schools) to execute different components of the program (e.g., training and data collection). The following chapter includes: a) a description of the structure of Reading Recovery in the United States both pre- and post-i3 scale-up; b) a logic model for the implementation of the Reading Recovery program; c) a logic model for the nationwide scale-up of the Reading Recovery program; and, d) CPRE's theory of action showing relationships between University Training Centers (UTCs), Regional Training Centers (UTCs), districts, and RR schools and teachers. The two logic models are based on a template provided by Abt Associates through their technical assistance to all i3 grantees. The logic models are intended to include enough detail so that our evaluation procedures could be replicated. Thus, though they are complex, the logic models are included here in the interest of comprehensive clarity.

Reading Recovery’s Historical Structure

Throughout its history in the United States, Reading Recovery has been sustained and developed by a cooperative network of 19 partner universities. This structure remains intact, and is represented in Figure 2, the Implementation Logic Model (further described below). Each partner university supports the operation of a University Training Center (UTC). UTCs are staffed by at least one faculty member who is partially or wholly dedicated to the training of new teacher leaders and the operation of the UTC. In very few cases, two trainers share the role of UTC director. Some UTCs are also supported by a small administrative staff and/or graduate assistants.

Each UTC is responsible for overseeing all Reading Recovery-related activity in its affiliate teacher training sites. UTC directors carry out these responsibilities with the assistance of teacher leaders and district site coordinators, who are often employed directly by a school district and carry out their Reading Recovery responsibilities in addition to their district-specific responsibilities. While most UTC regions are geographically defined, they frequently encompass more than one state and, in a few cases, overlap.

As the first UTC for Reading Recovery in the country, OSU holds the trademark for Reading Recovery in the United States. Other UTCs are licensed to use the trademark on a royalty-free basis. Figure 1 illustrates the collaborative but autonomous nature of the pre-i3 relationships between UTCs; this illustration includes the UTC at OSU.

Figure 1: The National Reading Recovery Network, Pre-i3

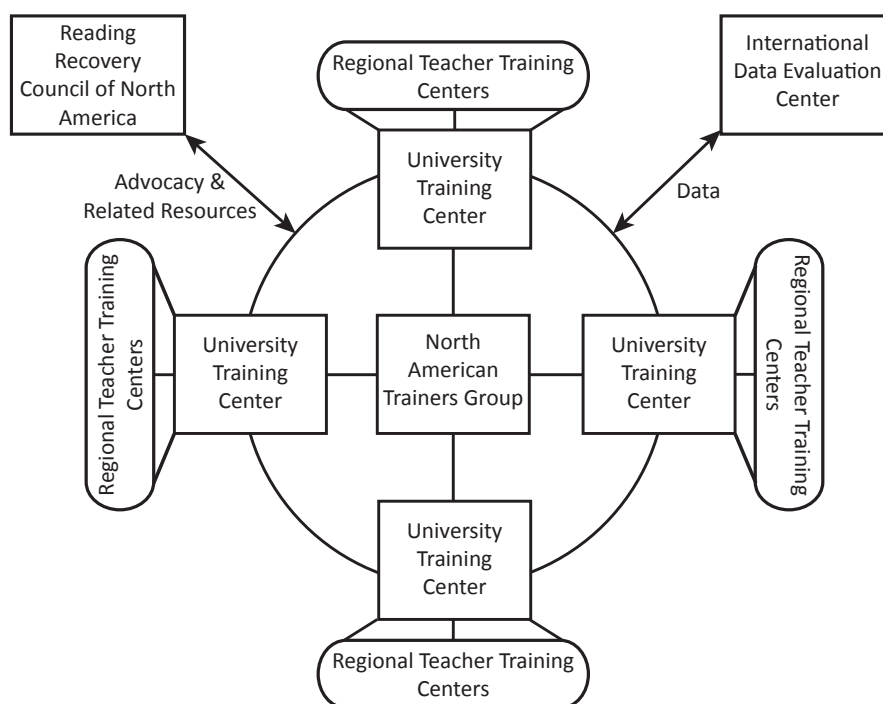


Figure 1 also includes other components of the national RR network. The Reading Recovery Council of North America is the membership agency responsible for carrying out national advocacy efforts, and coordinating resource-sharing and communication across RR regions. The North American Trainers Group (NATG) serves as the primary vehicle for collaboration among trainers at different UTCs. The IDEC, which is housed at OSU, is the partner agency responsible for collecting and processing data related to Reading Recovery services and outcomes.

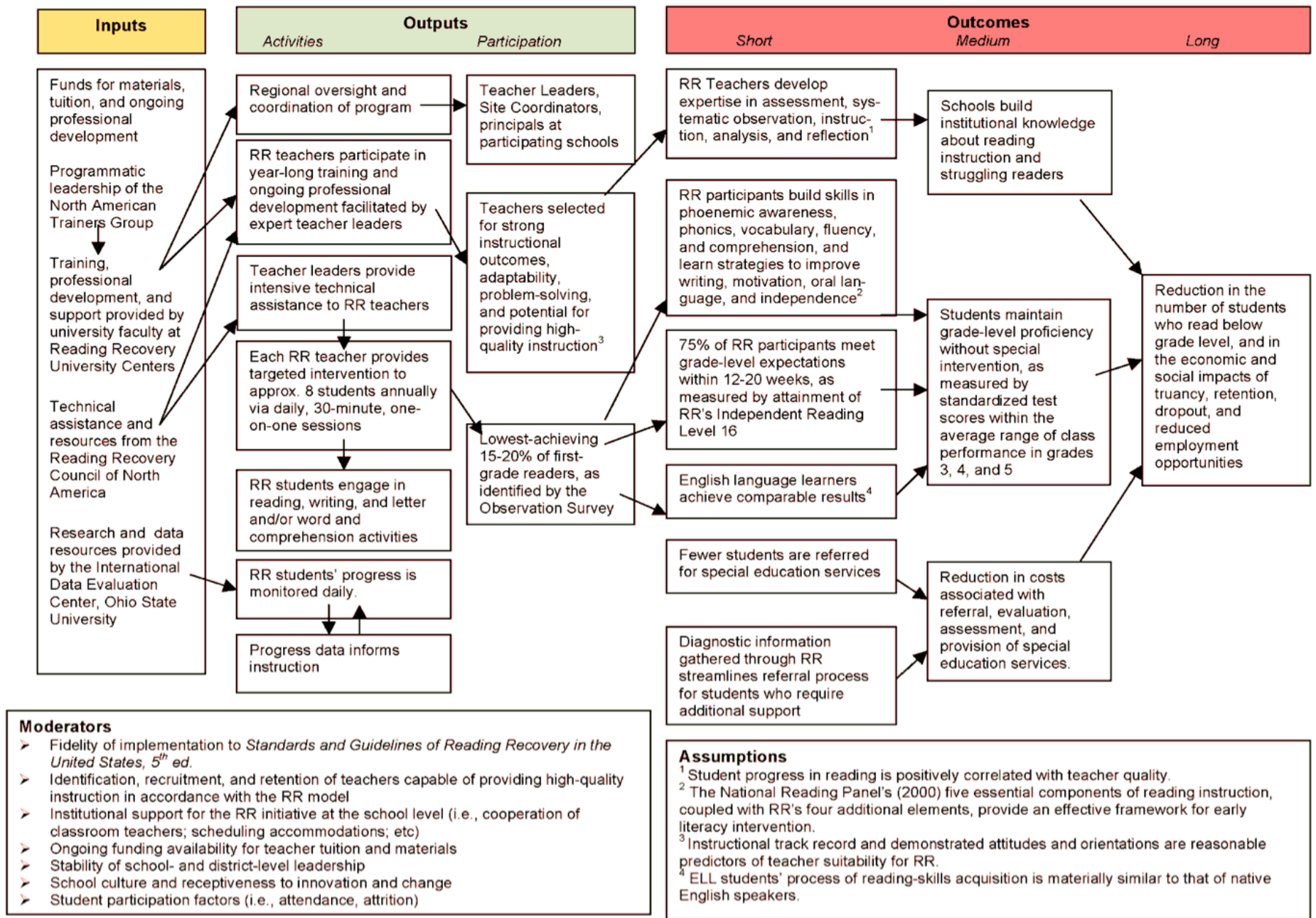
Logic Model of Reading Recovery Implementation

Figure 2 is a detailed diagram of the key inputs, outputs, and outcomes that characterize the Reading Recovery model. Inputs include: the i3 funds that pay for the training of RR teacher leaders and teachers; the district funds that cover RR teacher salaries; programmatic guidance provided by the NATG; training and support for RR teacher leaders from the 19 UTCs across the country (purchased with i3 funds); support from the RRCNA; and data collection, analysis, and research support provided by IDEC.

Outputs primarily relate to the selection of RR teachers and instruction of students. Ideally, according to the Reading Recovery model, teacher leaders, district-based site coordinators, and principals at participating schools work together to identify future RR teachers. RR teachers are then trained by regionally based teacher leaders on the specific skills that underlie the program's instructional approach. Once fully trained, RR teachers provide approximately eight students—typically the lowest-achieving 15-20 percent of 1st-grade readers in their school – with RR services each year. These eight students are typically served during half of the RR teacher's work day. During the other half of the day the RR teacher performs additional duties that vary by individual; common duties for RR teachers include classroom instruction, small-group work, or instructional coaching. As they continue in their positions, RR teachers receive ongoing professional development and technical support from teacher leaders both one-on-one and in group meetings with other RR teachers. This leads to increased teacher expertise and eventually, changes in schools' knowledge about instruction for struggling readers.

Anticipated outcomes for students range from improved reading achievement in the short term, to decreases in student retention and dropout rates over the long-term. Students who do not sufficiently respond to the intervention to be discontinued are referred for further evaluation. This process of referral can be streamlined for RR students by the screening and diagnostic information gathered through the RR assessment process and the administration of lessons. As its ultimate goal, Reading Recovery aims to reduce not only the number of students who read below grade level, but the future social and economic costs of associated truancy, retention, dropout, and underemployment as well.

Figure 2. Logic Model for the Implementation of Reading Recovery



Reading Recovery's Shifting Structure: A New Role for Ohio State

Though much of the infrastructure required to support a national scale-up of RR was in place prior to allocation of the i3 grant, receipt of the grant has required OSU to assume new fiscal, administrative, and managerial responsibilities. In response, it has been necessary for the RR system to evolve its historical structure to support the scale-up.

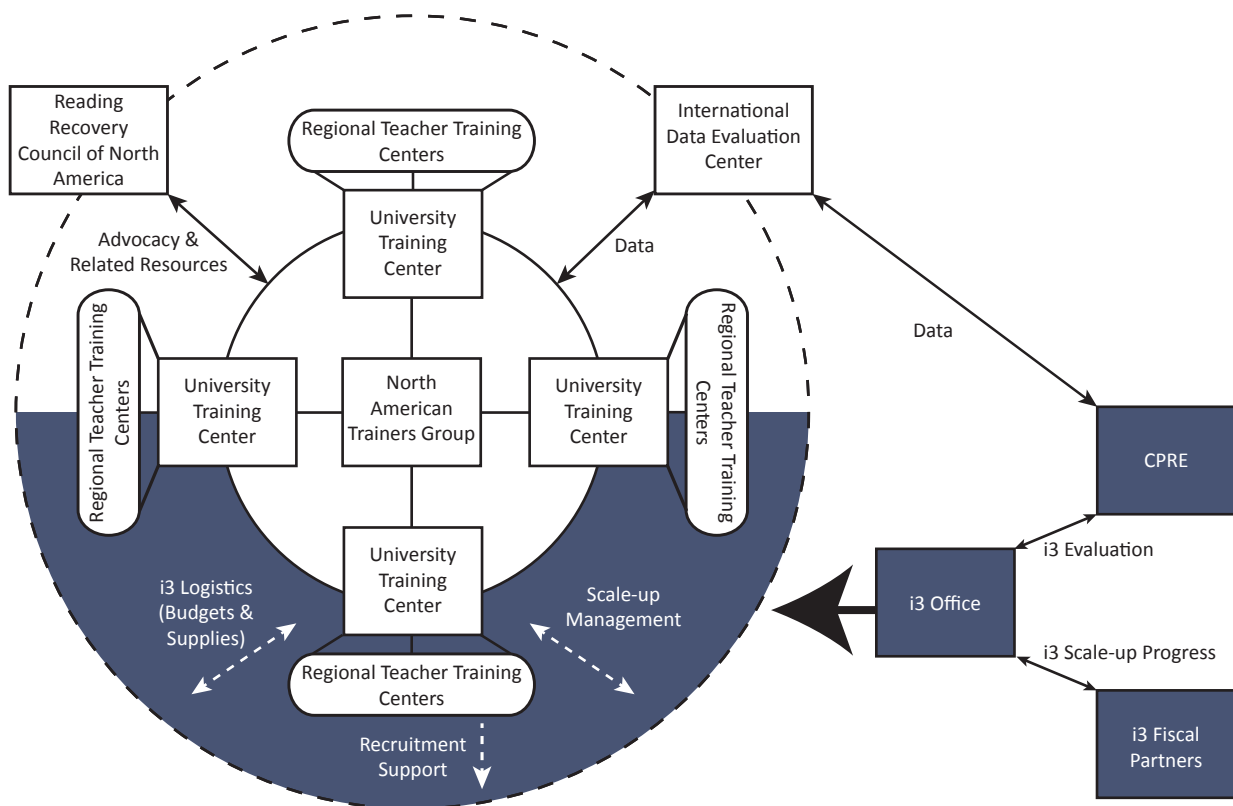
In response to new fiscal and administrative requirements, OSU has created what has become known throughout the Reading Recovery system as "the i3 office." This office, which is based at OSU, is staffed by a fiscal and project manager, as well the director and co-director of the grant. This team works in close collaboration with the administrator of the Reading Recovery IDEC to oversee the grant and support the i3 evaluation.

The work of the i3 office is focused solely on i3 related activities. The i3 office serves as the primary source of logistical and problem-solving support for the 19 partner universities involved with the i3 scale-up. While the i3 office is a coordinating agent and does not act in a supervisory role, the introduction of a centralized agency responsible for organizing implementation in the United States is nonetheless a shift for Reading Recovery. Figure 3 represents this new configuration.

In 2011-12, the i3 office, working in collaboration with IDEC staff, undertook the following key administrative tasks associated with the scale-up:

- » interfacing with OSU personnel responsible for overseeing sponsored projects;
- » overseeing and tracking the expenditure of grant funds;
- » reporting on grant progress;
- » securing the cost-share agreements and overseeing associated activities;
- » coordinating with the CPRE evaluation team and managing tasks associated with the evaluation, including the ITBS testing process;
- » ordering and distributing books for use in lessons with RR students;
- » collecting data to monitor the progress and impact of the scale-up;
- » updating/monitoring a shared data system to track i3 teachers and teacher leaders;
- » problem-solving support in response to requests from 19 university partners.

Figure 3: The National Reading Recovery Network, Post-i3



As depicted in Figure 3, the i3 office's supportive role for UTCs was primarily responsive in nature. In 2011-12, the office did not engage in preemptive troubleshooting. While UTC directors reported that the i3 office was the first place they turned for assistance with issues related to the grant, few systems or supports had been created to proactively assist UTC directors in navigating scale-up issues. CPRE will continue to explore the evolution of the national RR network in future years.

Logic Model of Reading Recovery Scale-Up Goals and Activities

Figure 4 provides an overview of the i3-funded scale-up of Reading Recovery, as described in the i3 grant proposal submitted by OSU and the cooperative agreement between OSU and OII. Key inputs, outputs, and outcomes that characterize the scale-up are represented. Along with grant funds for project costs, other primary inputs include expert knowledge to support essential scale-up activities. OSU, the funded organization, has engaged expertise and support from several sources. OSU contracted the Consortium for Policy Research in Education (CPRE) to evaluate the impact, implementation and scale-up of RR; expanded relationships with the North American

Trainers Group (NATG), a consortium that provides scale-up support; and worked closely with the 19 UTCs to ensure the successful implementation of the RR model in schools nationwide.

The expertise of the Project Director and Co-Directors enables two primary scale-up activities or outputs: UTC recruitment efforts, and development of training facilities. The Project Director and Co-Directors provide oversight and coordination of UTC efforts to recruit RR teacher leaders, districts, and schools, and ensure that UTCs receive resources and support for their efforts to recruit districts and schools in their region to participate in the i3 project. They also support UTCs as they establish new RTCs. In addition, the Project Directors and Co-Directors are also responsible for developing and maintaining partnerships with private sector organizations.

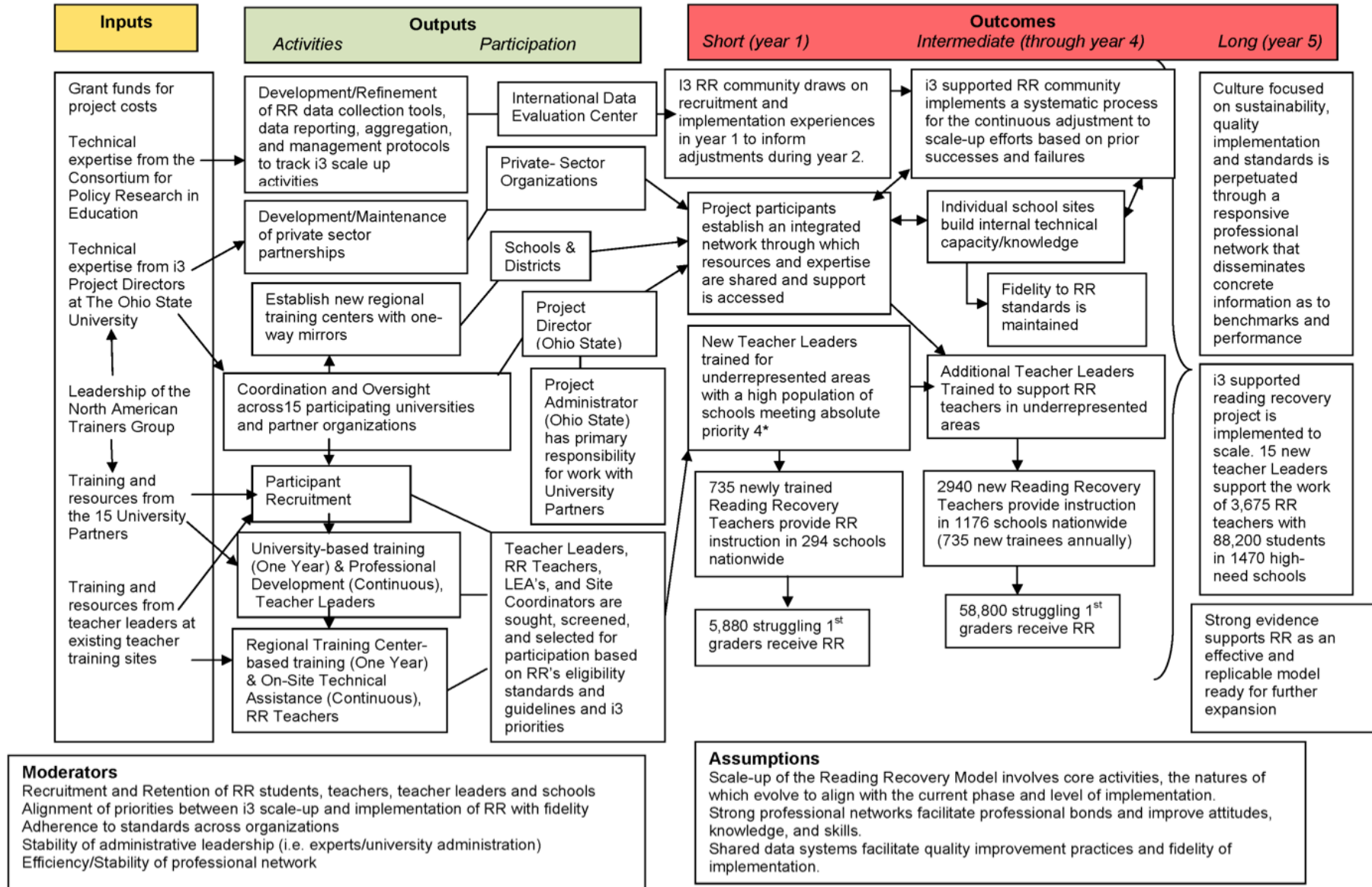
Other direct inputs for scale-up activities are training and resources provided by each of the 19 UTCs. These inputs enable several outputs, specifically the regional recruitment of teacher leaders and site coordinators, and teacher leader training. In addition to recruiting and training, UTCs are also responsible for ensuring that teacher leaders and site coordinators meet Reading Recovery's standards for qualified personnel. Training and resources from experienced RR teacher leaders are also direct inputs for scale-up activities; teacher leaders are responsible for recruiting new RR teachers and delivering to them a year-long training program, in addition to providing continued formal and informal professional development to both new and experienced RR teachers.

Anticipated outcomes of these inputs and output activities include the expansion of: a) a high-quality, shared data system that allows RR stakeholders to monitor and assess student performance, and b) an integrated professional network through which information is distributed and obtained. In the short term, teacher leaders provide support to RR teachers in i3 priority areas (i.e., low-performing schools, rural schools, and schools with significant numbers of ELLs). In the intermediate future, new RR teachers provide support to struggling first graders. And in the long term, the RR scale-up is anticipated to train 15 new teacher leaders and 3,675 new RR teachers, and impact over 88,000 below-average first graders in 1,470 high needs schools nationwide.

Reading Recovery Theory of Action

During the first year of our external evaluation of the i3 scale-up, CPRE used a mixed-methods approach to investigate the quality of the regional implementation of Reading Recovery. The framework for this approach was based on two components: a) the principles espoused by Clay, and further developed by other Reading Recovery scholars and practitioners; and, b) the concrete RR implementation practices standardized through the Standards and Guidelines. By including both of these components, CPRE sought to examine both the spirit and the letter of adherent RR implementation. The RR Implementation Logic Model (Figure 2) is complimented by the School-Level Theory of Action (Figure 5), which represents the systemic nature of school-level implementation of Reading Recovery.

Figure 4. Logic Model for the Scale-Up of Reading Recovery



At the school level, RR implementation includes specific interactions and exchange of resources between stakeholders both within a school and across its boundaries. The theory of action represented here assumes that the relationships that develop through these interactions and exchanges will accomplish two things. First, they will protect the instructional core of the intervention, enabling implementation with fidelity, and facilitating the expected student-level outcomes. And second, they will build a base of support for Reading Recovery in a building, which is key to the expected school-level outcomes and promotes sustainability of the intervention.

This report addresses several closely related but distinct questions regarding the regional implementation of Reading Recovery:

1. What is the nature of the implementation of Reading Recovery in the schools involved in the i3 scale-up?
2. What are the regional factors that support or hinder the school-level implementation of Reading Recovery?
3. To what extent are school-level enactments of Reading Recovery reflecting the aspirations of Reading Recovery embodied in the theory of action represented in Figure 5?

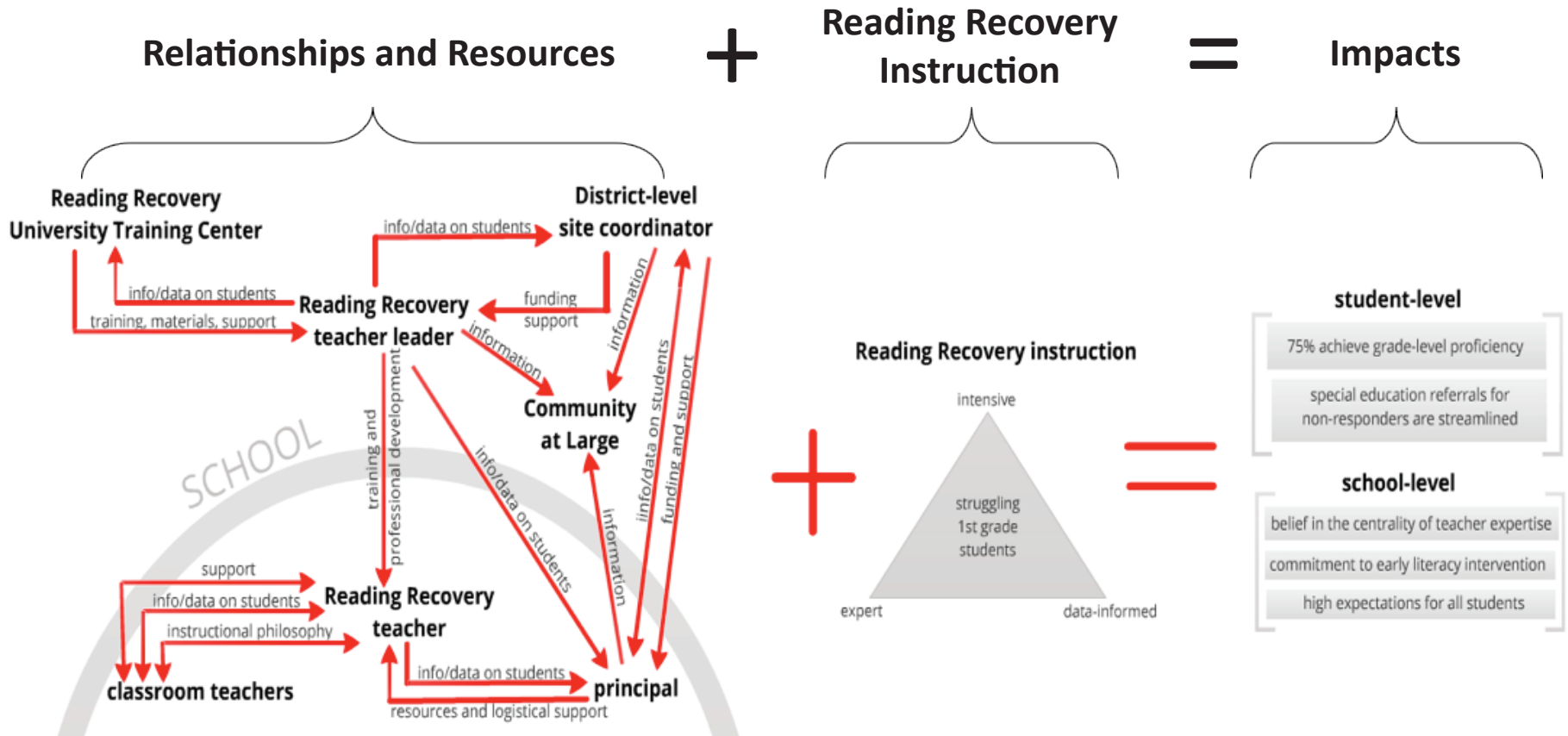
Interviews and surveys were conducted with more than 1000 Reading Recovery stakeholders, and provided rich descriptions of the educational contexts in which RR is being implemented in this national scale-up. These data sources informed CPRE's analysis of the relationships between context and fidelity, and between idealized and actual enactment of the RR model.

The Bigger Picture: This Report in the Context of the Longitudinal Evaluation of Reading Recovery

This first report presents early results from the randomized impact and implementation studies conducted over the 2010-11 and 2011-12 school years. As they are part of an ongoing evaluation, the findings presented here are preliminary, yet provide the foundation for future exploration of how Reading Recovery influences and is influenced by educational systems.

The subsequent chapters of this report are organized as follows. First, we describe the research study design and methods, and a timeline for the evaluation through 2015. Next, we present the estimates of the impacts of Reading Recovery based on the first cohort of students and schools participating in multi-site randomized experiment during the 2011-12 school year. We then present four chapters detailing results from our mixed-methods study of RR implementation under this i3 scale-up. These four chapters focus on a) scale-up management and oversight, b) recruitment of schools and teachers, c) training of RR teachers, and d) school-level implementation of the Reading Recovery intervention.

Figure 5. Reading Recovery Theory of Action for Systemic Implementation



Evaluation Study Design and Methods

This evaluation utilizes two parallel research designs to estimate short-term and long-term impacts of Reading Recovery on student achievement. Short-term impacts are estimated using a randomized controlled trial (RCT) field experiment; long-term effects (through third grade) are estimated via a quasi-experimental regression discontinuity design (RDD). Across the full population of i3 schools, a mixed-methods approach is used to collect implementation fidelity data from a large sample of sites, and to provide sufficient detail to understand the RR intervention and its implementation in context. The data collection tools include interviews, focus groups, document review, activity logs, and surveys. To maximize efficient use of resources in this evaluation, the RCT and RDD studies are being conducted in random samples of schools from the larger population of Reading Recovery i3 schools, while the survey components of the implementation study are being conducted in all Reading Recovery i3 schools. Furthermore, a sample of schools is recruited for more in-depth case studies, including interviews with RR and school stakeholders, and observation of classrooms and RR lessons. The overarching goals of this evaluation are to document the implementation of Reading Recovery under the i3 scale-up, and produce rigorous estimates of the causal impacts of the intervention on student literacy achievement.

Primary Research Questions

The primary research questions for CPRE's overall, multi-year evaluation are:

1. For 1st-grade students who begin the school year struggling to read (i.e., they have the lowest text reading levels in their class), what are the impacts of Reading Recovery on reading achievement at the end of the 12-20 week intervention, as measured by the ITBS?
2. What are the longer-term impacts of Reading Recovery on student reading performance in the 3rd and 4th grades, as measured by state reading tests?

3. Are the recruitment goals (for teacher leaders, RR teachers, and schools) of the scale-up being met, and what factors have supported or hindered efforts to meet these goals?
4. Is training of RR teachers being implemented as intended, and what is the perceived quality of the training?
5. Is the Reading Recovery intervention being implemented as intended, and what factors support or hinder fidelity of implementation?

This first report presents early results that provide preliminary answers to four of these research questions. Some will require additional years of data collection and analysis to address, such as Question 2, which focuses on long-term impacts. The fidelity of implementation referred to in Question 5 is, for this first report, understood as adherence in RR lessons to the Standards and Guidelines related to instruction. In future years, this question of implementation will expand to focus on the quality of instruction beyond fidelity. It should also be noted that the data collection instruments used in this evaluation are revised and adapted each year based on the issues and themes that emerge during analysis. The overall research design and project timeline, along with the instruments, protocols, and analysis methods used in the 2011-12 year of the evaluation, are described below.

Evaluation Timeline

Year 1 of this project began in October 2010, after the start of the 2010-11 school year. As it was not possible to control or document the selection of students to receive RR services at this point in the cycle of implementation, Year 1 was dedicated to long-term planning, and impact study and instrument design.

In Year 2 of the study, the RCT and RDD designs were implemented, each in a separate group of randomly selected and mutually exclusive schools (see Table 1). Going forward, these groups of schools will rotate through the research designs they participate in each year, so that schools are not forced to implement random assignment of students more than once every three years. Estimates of short-term impacts of RR on reading achievement were made based on data produced through the administration of the ITBS to treatment and control students in the RCT schools. The ITBS was administered midway through the school year, after the first round of students (i.e., the treatment students) completed their intervention cycle (see the section describing the RCT study below for more details). Also in Year 2, interviews, focus groups, observations of instruction, activity logs, and surveys were conducted with Reading Recovery and school staffs to gather data on scale-up activities and implementation fidelity. These same procedures will be repeated in Years 3-5 of the evaluation, using revised instruments when necessary to address issues and themes raised through data analysis.

In Years 3-5 of the study, data requests will be prepared and submitted for access to 3rd- and

4th-grade state test scores for students participating in Reading Recovery at schools selected for the RDD study of long-term impacts.

As previously discussed, this evaluation will ultimately result in a series of three reports. The second report, to be released in 2014, will delve further into issues of potential supports or hindrances to successful implementation of Reading Recovery—specifically the contextual factors that explain variation in impacts across schools. The third report, to be released in 2015, will present final estimates of impacts on student outcomes, both short- and long-term. It will also include a cumulative perspective on the scale-up effort, and recommendations for maximizing the fidelity of implementation and the size of impacts for Reading Recovery across the nationwide population of schools.

Quantitative Impact Studies of i3 Scale-up

Selection of Schools into Impact Studies

As a way to secure agreement from scale-up schools to participate in a rigorous impact evaluation, the study team agreed to vary the study designs used in each school over the course of the evaluation. Each of the 628 i3 schools contributed to three independent evaluation samples over the four-year evaluation period—the RCT, the RDD, and an internal-only evaluation. The schools were randomly assigned to three blocks, with block assignment determining the nature of the evaluation they would receive each year (See Table 1).

The most burdensome evaluation design was the RCT, which required that eligible students be randomly assigned to receive either RR or general classroom literacy instruction. Random assignment of students to treatment and control conditions is a significant departure from typical Reading Recovery procedures, and feedback from program and school leaders suggested that individual schools would be unwilling to implement the RCT design every year of the evaluation. In order to minimize the burden of this design on individual schools, each was rotated through the different evaluations each year, meaning a school only implemented the RCT once or twice over the course of the study. This sampling procedure ensured that representative samples of schools with varying lengths of RR implementation are included in each of the three evaluation designs.

Multi-Site Randomized Controlled Trial for Estimating Short-Term Impacts

Prior to the start of the 2011-12 school year, 209 schools participating in the i3 scale-up were randomly selected for inclusion in an RCT. At each school, a subsample of low-performing students was identified using the Reading Recovery Observation Survey of Early Literacy Achievement (the OS). These students were then rank-ordered according to their Text Reading Levels according to the OS. The eight students with the lowest scores were matched according to

pretest scores and ELL status. One student in each pair was randomly assigned to the treatment (RR instruction) group, and the other was assigned to the control (classroom literacy instruction) group. Upon the conclusion of the intervention (i.e. 12-20 weeks), each treatment group student and control group counterpart were assessed at approximately the same time using the ITBS. After the ITBS was administered to both students from a matched pair, the control group student became eligible to receive the one-on-one Reading Recovery intervention.

Table 1. Random Selection and Rotation of Schools Across Three Research Designs

Study Year	Group	2011-2012	2012-2013	2013-2014	2014-2015
1	n/a	No Impact Evaluation			
2	1	Randomized Experiment	RR Internal Evaluation	Regression Discontinuity	Randomized Experiment
2	2	Regression Discontinuity	Randomized Experiment	RR Internal Evaluation	Regression Discontinuity
2	3	RR Internal Evaluation	Regression Discontinuity	Randomized Experiment	RR Internal Evaluation
3	1		Randomized Experiment	RR Internal Evaluation	Regression Discontinuity
3	2		Regression Discontinuity	Randomized Experiment	RR Internal Evaluation
3	3		RR Internal Evaluation	Regression Discontinuity	Randomized Experiment
4	1			Randomized Experiment	RR Internal Evaluation
4	2			Regression Discontinuity	Randomized Experiment
4	3			RR Internal Evaluation	Regression Discontinuity
5	1				Randomized Experiment
5	2				Regression Discontinuity
5	3				RR Internal Evaluation

The “blocking” of students in matched pairs was intended to address the variability of the length of the intervention cycle for students. Blocking students into pairs ensured that the outcome for each treatment student was compared to the outcome for a control student who experienced the counterfactual for the same length of time as the treatment. In addition, blocking students is a mechanism that increases the likelihood of the baseline equivalence of treatment and control in regards to pretest scores (i.e., text reading levels) and ELL status.

The Outcome Measure: Iowa Tests of Basic Skills

The ITBS is a well-regarded, group-administered, norm- and criterion-referenced, standardized assessment designed to “assess the extent to which a child is cognitively ready to begin work in the academic aspects of the curriculum” (Hoover et al., 1994, as cited in Tang & Gomez, 2007), and to “measure growth in fundamental areas of school achievement” (Hoover et al., 2003, p.1).

Originally published in 1955, the ITBS is currently available in two forms, A and B, which are broken into multiple parts and subtests that measure achievement for students in kindergarten through the eighth grade. Part 1 of the ITBS is a teacher-administered subtest, while Parts 2-6 are student-administered following a teacher-modeled example. The analysis performed in this evaluation to determine overall program impacts used three scores from the ITBS Reading subtest. These were the Reading Words and Reading Comprehension subtests, used for exploratory analysis, and the Reading Total score, used for confirmatory analysis. The parts that comprise each score are as follows:

- » Reading Words: Words (Part 1), Pictures (Part 2), and Word Attack (Part 3)
- » Reading Comprehension: Sentences (Part 4), Picture Story (Part 5), and Story (Part 6)
- » Reading Total: Parts 1-6

These test components were chosen for several reasons. First, the battery of tests utilized during this study (ITBS form A, level 6), is appropriate for students who are six-years-old and whose level of academic development ranges between K.8 to 1.9. Second, ITBS raw scores can be converted to several other types of scores, including developmental scores (grade equivalents), developmental standard scores, and status scores. Finally, the national standardization of the ITBS was conducted with a normative sample designed to represent the national population of school children, grades kindergarten to eight (Hoover, Dunbar, & Frisbie, 2011).

Information regarding the technical characteristics of the ITBS was obtained through the Guide to Research and Development (GRD Manual), and the ITBS technical manual. The GRD Manual contains multiple reliability coefficients (internal consistency, equivalent forms, test-retest), most of which range between the middle .80s to low .90s. Designed to “measure growth in the fundamental areas of school achievement” (Hoover et al., 2003, p.1)—including vocabulary and reading comprehension—the ITBS manual provides sound evidence to support the instruments’

content validity and high discriminant ability (item p-values and discrimination indices) (Hoover et al, 2003). Additionally, the ITBS has often been used as an outcome measure for both experimental and quasi-experimental impact studies (Kim & White, 2008; Reis et al., 2008; Jenner & Jenner, 2007). In all, the ITBS is regarded as a well-developed assessment with sound technical qualities established through rigorous processes.

The Pretest Measure: The Observation Survey of Early Literacy Achievement

The Observation Survey of Early Literacy Achievement (OS), developed by Marie Clay, is the primary screening, diagnostic and monitoring instrument for Reading Recovery. It is a one-to-one, teacher-administered, standardized assessment that includes six sub-scales: Letter Identification, Concepts about Print, Ohio Word Test, Writing Vocabulary, Hearing and Recording Sounds in Words, and Text Reading Level. The Text Reading Level (TRL) is used to block students during the random assignment process, and later as a pretest covariate in the statistical models of impacts.

RR teachers administer the Text Reading OS task in order to determine a student's text reading level. Text reading level is defined by RR as the level at which a student can read a set of texts, specifically the Scott Foresman Specialty Practice Books. This OS task requires the administration of a running record, which teachers use to record the speed and accuracy with which a student reads a selected text with known difficulty. Using this method, RR teachers can determine a student's text reading level with at least 90 percent accuracy (NATG, 2005). Multiple methods have been employed to estimate the reliability of the OS. Reported test-retest and internal consistency reliability estimates range from moderate to high on the individual OS Tasks (Clay, 2002, as cited in Denton, Ciancio, & Fetcher, 2006); measures of the inter-assessor reliability of the Text Reading and Writing Vocabulary tasks yielded coefficients of .92 and .87 (Denton et al., 2006). In addition, evidence of the validity of the information gathered by administering the OS has been proved by several studies. These studies assess the construct and criterion validity of the OS tasks using the sub-tests of various norm-referenced tests, including the ITBS. Across these studies, researchers have found that scores can be validly interpreted for the following purposes: a) identification of at-risk students (Gomez, Rogers, Wang, & Schultz, 2005), b) measurement of early reading constructs (Tang & Bellenge, 2007; Gomez, Gibson, Tang, Doyle, & Kelly, 2007), and c) prediction of the attainment of performance benchmarks (Denton et al., 2006).

Recording Student Demographic and Achievement Data

The Reading Recovery International Data Evaluation Center (IDEC) provides data collection, management, and analysis support to Reading Recovery programs throughout the U.S. IDEC's existing infrastructure supported large-scale data collection at low cost under this evaluation. IDEC has developed a two-step data entry and review process. First, RR teachers are required to enter data regarding their personal characteristics, along with student and school information;

teacher leaders are then required to review this data as an additional quality assurance measure. IDEC data is used by UTC Directors, teacher leaders, RR teachers, and school administrators for their annual reporting. As part of the scale-up evaluation, CPRE coordinated with IDEC to allow RR teachers to enter pretest OS data and posttest ITBS and OS data for students in the RCT each year. To address possible test administrator bias, the post-intervention assessments with each student are conducted by someone other than the teacher who provided RR lessons to that student (i.e., typically another RR Teacher or a Teacher Leader). CPRE received data files for i3 schools that included student, teacher, and building characteristics. These data were used by CPRE to: a) generate sampling frames for the RCT and qualitative data collection components of the evaluation; b) report aggregate statistics regarding the i3 scale-up and Reading Recovery's implementation; and, c) compare outcomes for treatment and control students participating in the RCT.

Statistical Analyses of Impacts on Student Reading Achievement

Impacts on student reading performance were estimated by comparing mid-year reading achievement of students randomly assigned to participate in Reading Recovery at the beginning of 1st-grade to students randomly assigned to the control condition. Using a three-level hierarchical linear model (HLM) (Raudenbush & Bryk, 2002) with students nested within matched pairs, and matched pairs nested within schools. Differences in the posttest performance of the treatment and control students were estimated after controlling for pretest performance. This HLM included the OSText Reading Level scores as a covariate, random effects for blocks (i.e., matched pairs), a random effect for overall school performance (i.e., random school intercepts), and a random effect for the impact of Reading Recovery (i.e., random treatment effects across schools). Models were estimated using PROC MIXED in SAS 9.3 via Restricted Maximum Likelihood (REML), with model-based standard errors and degrees of freedom based on within- and between-cluster sample sizes.

Impact estimates from the HLM models represent mean differences in ITBS scale scores between treatment and control groups after adjusting for initial text reading level on the OS. These raw impact estimates are standardized using the standard deviation of the outcome for the control group to produce Glass' D. The choice to use Glass' D is based on the expectation that the impact of Reading Recovery will vary across students and schools, resulting in an increase in not only mean posttest achievement, but also an increase in the variance of posttest achievement scores. By using the control group standard deviation, we are better able to benchmark the impact estimate against the counterfactual (i.e., the distribution of potential outcomes in the absence of the intervention).

In addition to Glass' D, which represents a standardized effect relative to the distribution of outcomes for only study participants (i.e., the lowest eight students in each school who were

selected for the RCT), we present a population-based standardized effect size. This effect size is calculated by dividing the raw impact estimate by the standard deviation of ITBS scores for the national norming sample. This allows the impact of Reading Recovery to be benchmarked against the full population of 1st-grade students, not just the struggling readers in the study sample. These impact estimates should be smaller than the Glass' D estimates because the variance in outcomes for the full population of first graders is larger than the variance for struggling readers.

Exploring and Explaining Variation in Impact Estimates

The large number of schools involved in this multi-site study allows for HLM analyses of school-level factors associated with variability in program effects. The use of random treatment effects for schools in a multilevel modeling framework allows estimation of variability in treatment effects across schools, and supports estimation of cross-level interactions to explain variability in treatment effects. Analyses in this report include estimates of the standard deviation of program effects across the sample of schools. These estimates are based on the random effects covariance estimates from the HLM analyses. In addition, we present a histogram of empirical Bayes estimates for site-specific impacts; however, raw empirical Bayes estimates have an underestimated variance due to over-shrinkage (Bloom, Raudenbush, & Weiss, 2012; Raudenbush & Bryk, 2002, p. 88). Therefore, we calculate adjusted site-specific estimates by rescaling the variance of the empirical Bayes impact estimates to equal the estimate of the true impact variance from the HLM covariance parameter estimate.

Beginning in Year 2, data will be pooled across multiple cohorts of schools participating in the RCT component of this evaluation. Data from surveys and activity logs completed by teacher leaders, RR Teachers, and site coordinators will be linked and used as predictors of school-level variability in impact estimates, while student demographic variables will be used to investigate variability in effects across subgroups of students. School-level variability in impacts is expected to be substantial given that the intervention in most schools is delivered by a single teacher, and that there is substantial variability in school contexts, student demographics, and school and district program supports.

The multiple sources of data just described are also leveraged to explore the strategies for, and progress of, the scale-up effort. Interviews with program directors and trainers across the 19 UTCs provide detailed information about recruitment and training; teacher surveys, RR teacher activity logs, and document reviews provide information about quality and fidelity of implementation across hundreds of schools. Indicators of variation in program contexts, supports, and components derived from these data are also used to explore potential predictors of variation in program impacts in subsequent statistical analyses of student reading achievement scores.

An important component of future analyses in this evaluation of the i3-funded scale-up will involve combining data from the multi-site RCT with individual and contextual factors that

emerge from the survey sources just described. This will enable us to identify predictors of variation in program impacts when implemented at scale. Potential moderators of program effects include teacher characteristics (e.g., years of experience, education, certifications, prior training), student characteristics (e.g., race, gender, initial reading performance, ELL status), and school characteristics (e.g., school size, leadership support, coherence with other programs/interventions, fidelity of implementation). Numerous indicators of implementation fidelity are derived from the Reading Recovery Standards and Guidelines. By linking these teacher, student, school context, and implementation indicators to impact results, we can begin to identify those factors that are predictive of larger effects. These results can inform future implementations of Reading Recovery, including future waves under the i3 scale-up, in order to maximize effects on student outcomes.

Multi-Site Regression Discontinuity Design for Estimating Long-Term Impacts

In the final year of this evaluation (2014-15), long-term program impacts of RR will be estimated through an RDD. Given this timeline, data from the RDD will be included in subsequent reports.

The RDD classifies the eight lowest-performing 1st-grade students (as identified by the OS) as the treatment group. All of these students will receive the RR intervention. The next five to ten students above the cut-score (i.e. the point between the lowest eighth and ninth ranked student) are assigned to the control group. These students will not receive the RR intervention at any time during first grade or subsequent grades.

By the third year of this evaluation, 2013-14, most students enrolled in the study during the 2011-12 and 2012-13 school years will be in grades 3-4. At that time, reading achievement scores from state assessments will be available for students who were enrolled in the study during the first two years of the evaluation. The availability of state test data during the final year of the study assumes that scores will be available by the end of the summer following testing. Analyses of data for students enrolled in subsequent years of the study (i.e. 2013-2014 and 2014-2015) is possible, but will not be completed within the timeframe of this i3 grant. Plans to seek supplemental funding from other sources are in development.

State reading achievement scores will be used as a longer-term outcome measure, and the analysis of these data will focus on impacts within each grade level (i.e., with separate estimates for third and fourth grades). Cross-state impact estimates will be produced using methods described by May et al. (2009). More specifically, test scores in each state will be standardized by rescaling individual student scores using the standard deviation of outcome scores for each statewide population of students. These standard deviations will be obtained from published technical manuals for each state test. Data will then be pooled across states to estimate overall impacts.

Secondary Study of Changes in Referral and Grade Retention Rates

Data collection is also underway for a non-experimental study of changes in grade retention and special education referral rates for first graders. As a secondary non-experimental study, we will compare the retention and referral rates for Reading Recovery schools before and after their first year of implementation. Data from this study of retention and referral will be pooled across multiple cohorts of i3 schools, and included in the final evaluation report.

The key to valid inference in this multi-cohort design (Shadish, Cook, and Campbell, 2002, p. 148) is the relative comparability, without the need for statistical adjustment, of pre-intervention and post-intervention cohorts of first graders. Because the outcomes of interest tend to happen near the end of first grade (i.e., special education referrals) or after the end of the school year (i.e., retention in first grade), it is not feasible to do comparisons of treatment and control students because these groups are maintained only through the first half of the school year. The treatment and comparison groups from the RDD study are not comparable without statistical adjustment, and the rarity of retention and referral makes it difficult to estimate the relationship between the pretest measures and these binary outcomes with sufficient precision. Therefore, the best available counterfactual for first graders during the first year of Reading Recovery implementation is the cohort of first graders in the same school during the previous year. To further ensure the best cohort comparability, we chose to forgo a longitudinal study, and instead compare only the cohort from the first year of RR implementation to the cohort from the year just prior to implementation in each participating school.

Although this multi-cohort design is observational, the pooling of data across hundreds of schools and dozens of states addresses many of the potential confounds threatening valid causal inference. In order for a particular threat to causal inference to stand, it would need to explain population-wide shifts in retention and referral rates from one year to the next. In other words, those confounds that are specific to individual schools, districts, or states are likely to explain an increase or decrease in referral or retention rates only in that specific school, district, or state. One would be hard-pressed to find a confound that explains away systematic differences in referral and retention rates across all schools in the sample. However, because it is not entirely impossible for such a systematic confound to exist, and because site-specific confounds may combine to produce overall positive or negative differences, results from this simple pre-post analysis are non-causal.

Overview of the Mixed Methods Evaluation of Implementation Under the i3 Scale-Up

Several qualitative instruments and surveys were developed to collect data on recruitment and implementation of Reading Recovery under the i3 scale-up. These instruments are listed and linked to our scale-up and implementation research questions in Table 2. Subsequent sections detail the format, purpose, and content of each qualitative and survey instrument.

Qualitative Evaluation of i3 Scale-Up

The qualitative components of this evaluation explored the RR scale-up as a series of processes (e.g., recruitment, training, and implementation) enacted in regional and local contexts. The core research questions guiding this implementation evaluation are:

- » Are the recruitment goals of this scale-up being met, and what factors have supported or hindered efforts to meet these goals?
- » Is training of RR teachers being implemented as intended, and what is the perceived quality of the training?
- » Is the instructional component of the Reading Recovery intervention being implemented as intended, and what factors support or hinder implementation fidelity?

Participant Interviews

Much of the qualitative data about implementation was derived from participant interviews. Members of the research team conducted extensive interviews with core i3 stakeholders whose professional responsibilities involved the recruitment of i3 participants, the training of i3 teacher leaders and RR teachers, and the oversight/support of RR implementation activities. Most broadly, the interviews explored how information sharing processes were used to inform and refine practices designed to document, facilitate, and support the attainment of scale-up benchmarks, and to ensure faithful implementation of the Reading Recovery intervention.

OSU i3 Director/Co-Director Interviews

The technical expertise of the i3 Project Director and Co-Director at OSU is a primary input of the scale-up activities related to coordination, oversight, recruitment, and sustainability. The i3 Project Directors participated in annual interviews. The interview was conducted in September and provided reflections on past experiences and a forecast for the upcoming year; questions were framed around understanding and benchmarking performance expectations.

Table 2. Cross-linking of Scale-Up and Implementation Research Questions and Instruments

Scale-Up and Implementation Evaluation Research Questions	UTC Directors	i3 Staff	Teacher Leaders		Site Coordinators		RR teachers			Principals	1st Grade Teachers	School Case Studies
	Interviews	Interviews	Focus Groups	Surveys	Interviews	Surveys	Interviews	Surveys	Activity Logs	Interviews	Surveys	
<p>Research Question 3: Are the recruitment goals (for teacher leaders, RR teachers, and schools) of the scale-up being met, and what factors have supported or hindered efforts to meet these goals?</p> <ul style="list-style-type: none"> » Document the processes for recruitment of districts, schools, RR teacher leaders, and RR teachers » Document monitoring of recruitment by OSU and UTCs 	✓	✓	✓	✓		✓		✓				
<p>Research Question 4: Is training of RR teachers being implemented as intended, and what is the perceived quality of the training?</p> <ul style="list-style-type: none"> » Document the training and preparation of RR teachers » Investigate variation in training and support across training centers and sites 			✓	✓			✓	✓	✓			✓
<p>Research Question 5: Is the Reading Recovery intervention being implemented as intended, and what factors support or hinder fidelity of implementation?</p> <ul style="list-style-type: none"> » Investigate the extent to which the implementation activities of RR teachers, RR teacher leaders, and site coordinators during the i3 scale-up adhere to the Standards and Guidelines of Reading Recovery in the United States (5th Edition). » Document differences across schools and districts in terms of student selection and instructional practices. » Explore the roles of RR teacher leaders, RR teachers, and other stakeholders in the implementation of RR under the i3 scale-up. » Examine the school-level factors that support or hinder RR implementation. 			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Project Directors had multiple recruitment responsibilities within the i3 scale-up framework, which included the direct support and oversight of recruitment efforts at UTCs. As such, questions pertaining to recruitment processes, communication, and adaptations were multifaceted. The interview protocol focused on:

- » a narrative overview of the foundational components of the i3 scale-up;
- » the Project Directors' work to educate UTCs about expectations for i3 scale-up recruitment;
- » the processes for informing and adapting recruitment support provided directly to UTCs;
- » the mechanisms that facilitate or impede the exchange of recruitment information and provision of support to UTCs, directly or indirectly; and,
- » the nature, flow, and use of communication to provide support.

OSU i3 Project Administrator Interview

A key source of support for recruitment and scale-up activities was the i3 Project Administrator, an employee of OSU. The i3 Project Administrator participated in an annual interview. This interview explored recruitment issues and coordination across sites, including:

- » the nature of the information he collects;
- » the process for collecting and distributing information; and,
- » his understanding of his role in the i3 scale-up.

OSU i3 Fiscal Administrator Interview

An annual interview was also conducted with the i3 project's Fiscal Administrator. The general purpose of this interview was to flesh out the key input in the Scale-Up Logic Model labeled "grant funds for project costs." More specifically, the goal of the interview was to monitor the expenditure of i3 funds - for the project overall and for each UTC—in relation to original budgeted amounts. Interview questions focused on:

- » OSU's process for distributing funds across UTCs;
- » the amount of i3 funds allocated and distributed to each UTC per year;
- » relative expenditure by budget category (e.g., staff, books, materials);
- » changes in budgeting and justifications; and,
- » projections for future spending.

IDEC Director Interview

An annual interview was conducted with the Director of the Reading Recovery IDEC. The key activities from the Scale-Up Logic Model captured in this interview were “Development/refinement of RR data collection tools, data reporting, aggregation, and management protocols to track i3 scale-up activities.” This interview collected information about:

- » data collection tools and mechanisms for monitoring scale-up processes;
- » tools for supporting and monitoring evaluation design fidelity (i.e., random assignment for the RCT and cutoff assignment for the RDD); and,
- » data for monitoring program implementation and impacts, and use of data by constituents at various levels (e.g., teacher leaders, UTCs, OSU) to support program scale-up and implementation.

UTC Director Interviews

The expertise, training, and technical supports provided by the 19 UTCs were primary inputs that supported the scale-up activities of recruitment, training, and oversight of implementation. A total of 14 interviews were conducted with directors of different UTCs. Through the aggregate of the responses to these questions, the UTC Director interviews provided a qualitative overview of the training and recruitment of teachers, the regional oversight of implementation, the functioning of the professional network, and the variation in these activities across UTCs. Specific interview topics addressed:

- » the role of the UTC Director and other identified stakeholders in recruiting RR teacher leaders and site coordinators;
- » the specific activities that comprise recruitment efforts;
- » the resources used to support the recruitment efforts; and,
- » the internal oversight systems used to support adherent recruitment, and to revise or expand recruitment efforts.

The UTC Director interview also focused on the processes used to gather information on participant behavior and program outcomes, to disburse support (i.e. professional development, resources, technical assistance), and to inform practice in each of these areas. The roles of other stakeholder in these processes were also explored.

RR teacher Interviews

In the 2011-12 school year, the research team conducted 37 interviews with randomly sampled RR teachers. The RR teachers were working in schools across the United States. These interviews focused on activities related to i3 scale-up as well as program implementation. Specifically, the teachers were asked to discuss the following:

- » the training and mentoring they receive from teacher leaders, with a focus on content, quality, utility, and standards;
- » the implementation of the RR intervention itself, including the lesson structure and content, the assessment process, the process of discontinuing and/or referring students, and any challenges and adaptations that have developed;
- » the support and oversight the teachers receive from teacher leaders, school and district administrators, other teachers in their schools, parents, and the network of other RR teachers;
- » the perceptions of institutional change at their schools related to the implementation of RR;
- » the nature and frequency of communication regarding RR with school-level stakeholders; and,
- » data use by RR teachers and schools.

School Principal Interviews

The building-level principal is a primary participant in the implementation of Reading Recovery. In 2011-12, interviews were conducted with eight principals selected from a stratified random sample of schools (i.e., by region, urbanicity, grade-span, and school size). Principal interviews provided information that, combined with the perceptions of other stakeholders, generated a qualitative overview of the school context.

These interviews broadly focused on the nature, purpose, frequency, and processes for building-level communications about RR. Specific interview question topics included:

- » the principal's understanding of the RR intervention (e.g. ability to articulate key goals and practices);
- » the principal's involvement in the RR intervention (e.g. monitoring the outcomes of RR students);
- » the principal's perception of the impact of RR instruction on student literacy skills, and the impact of the RR intervention on school-wide literacy instruction practices;

- » the principal's perceptions of the fit of the RR program within the school context (e.g. alignment of RR with the school literacy strategy and other instructional initiatives, and building level modifications); and,
- » the building-level cycle of sharing information, adjusting practices, and improving capacity.

Teacher Leader Focus Groups

RR teacher leaders provide the most direct support and training to the RR teachers delivering the intervention. Teacher leaders offer a perspective on the i3 scale-up that spans the boundary between school-level and regional implementation. In 2011-12, 49 teacher leaders were interviewed in five focus groups. The content areas for the focus groups included:

- » the role of the teacher leader in recruiting RR teachers;
- » the specific activities that comprise their recruitment efforts (e.g. district/school selection criterion);
- » the resources used to support recruitment efforts;
- » training of RR teachers;
- » challenges and supports pertaining to recruitment, training, and implementation; and,
- » the internal systems used to provide oversight and support for adherent recruitment, and to revise or expand recruitment efforts.

Interview Coding and Analysis

All interviews and focus groups were externally transcribed, and all identifying information was removed. A coding scheme was created that aligned with the evaluation research questions and RR's underlying theory of action (see Figure 5). Umbrella codes were used to focus on the broad dimensions of RR implementation, such as recruitment, training, and lesson instruction; sub-codes were also used to explore specific aspects of these dimensions. After inter-rater reliability was established among the members of the research team, individuals coded transcripts by using Dedoose, a qualitative data analysis software program. In the first round of coding, themes were generated from each set of participant interviews. A second round of analysis examined themes across groups to triangulate experiences, find common threads, and to gain a broader perspective of the i3 scale-up.

Survey Instruments and Derived Measures of Implementation Fidelity

In spring of 2012, CPRE conducted surveys of RR site coordinators, teacher leaders, teachers, and 1st-grade teachers who have RR students in their class. These four separate surveys were designed to collect data on implementation efforts in relation to the expectations for implementation of Reading Recovery. The findings allowed for quantification of implementation fidelity across sites and by subdomain based on the Standards and Guidelines of Reading Recovery in the United States (5th Edition) (RRCNA, 2009), which provides detailed specifications for the Reading Recovery program.

The CPRE evaluation team operationalized the Reading Recovery Standards and Guidelines as questions on the various survey instruments. The “Standards” for the implementation of Reading Recovery are mandated by the program; however, the “Guidelines” are recommendations that may improve the quality of the program, but are not required. To acknowledge this distinction, CPRE quantified fidelity to the RR model separately for standards and guidelines. The applicable standards and guidelines for program implementation were each measured by at least one survey item for each stakeholder. Fixed-response data from each survey was used to create indices of implementation fidelity as percentages of standards and guidelines met.

As previously discussed, RR teachers and teacher leaders who are in training often have a different set of relevant standards and guidelines than those who are trained. For example, RR teachers in their training year are required to attend weekly classes, whereas trained teachers only have to meet occasionally. To account for these differences, CPRE calculated the percent of standards and guidelines being met for each respondent using only the standards relevant to his/her specific position. Overall percentages were calculated as the un-weighted average of standards (or of guidelines) that were applicable to a respondent.

The surveys were administered online via weblinks emailed to individual RR and classroom teachers, teacher leaders, and site coordinators. Several precautions were taken to avoid problems related to self-reported data. When measuring standards and guidelines, respondents were asked to report objective facts about RR implementation in their context, not to make evaluative judgments about themselves or their own contributions. And, to the extent possible, questions were carefully worded so as to not suggest that one response was “correct” or more appropriate.

Site Coordinators Survey

The nature of the RR scale-up requires that many of the activities noted in the Scale-Up Logic Model be replicated by regional and school-level stakeholders to ensure adherent implementation of the RR model across contexts. At the regional level, the administrative expertise provided by the RR site coordinator is a primary input that supports the oversight and

coordination of RR implementation. Though not specified in the RR Logic Model (see Figure 2), the site coordinators' responsibilities are considered implementation activities. Surveying site coordinators provides the opportunity to quantitatively overview the regional level components of RR implementation.

An online survey containing both fixed-response and open-ended items was distributed to each site coordinator in June 2012. The survey collected information regarding oversight of on-going implementation activities, such as:

- » their efforts to secure appropriate training facilities, training materials, and leveled texts;
- » their participation in the recruitment of personnel and assignment of teachers to schools;
- » how they monitor teacher caseloads to assure appropriate support for sites;
- » the nature and frequency of their meetings with teacher leaders to troubleshoot site issues and seek support for adaptations; and,
- » monitoring and tracking expenditures for reporting.

Survey links were emailed to 164 site coordinators; 105 responded (64 percent response rate).

Teacher Leader Survey

The teacher leader survey was designed as a compliment to the focus groups, and addressed systems and patterns of support and oversight of the i3 scale-up. In particular, the survey examined respondents' experiences with key scale-up activities, including:

- » the content and standards reflected in the training provided to RR teachers by the UTCs;
- » the exchange of resources and the patterns of communication between UTCs and teacher leaders, within the cohort of teacher leaders, and from the teacher leaders to the RR teachers working in schools; and,
- » teacher leader use of IDEC data, and the utility of such data.

Teacher leaders were also asked to provide information on attendance at training sessions for each of the RR teachers they trained, and on the content and pedagogy of the training sessions. These data are key to measuring fidelity to the standards for RR teacher training.

The survey included fixed-response and open-ended items, and was distributed to all 189 teacher leaders involved in the i3 scale-up. 169 teacher leaders responded to the survey (89 percent response rate).

RR Teacher Survey

In May, all RR teachers were asked to complete an online survey about their experiences during the previous year. The survey included a mix of fixed-response and open-ended items, and was designed to explore multiple components of the i3 scale-up and RR implementation. These included:

- » fidelity of instruction to the Standards and Guidelines;
- » the training RR teachers received;
- » use of curricular materials;
- » implementation of RR in one-on-one lessons with students; and,
- » nature and frequency of communication with RR teacher leaders, school and district administrators, students' parents, other teachers at the school, and other RR teachers.

Survey links were emailed to 1,018 RR teachers¹; of these, 799 responded to the survey (78 percent response rate).

Classroom Teacher Survey

Certain 1st-grade teachers were also asked to complete an online survey at the end of the 2011-12 school year. The survey was administered to the regular homeroom/literacy teachers of all students receiving the RR treatment as well as all students in the control group. The survey focused on both RR-related and regular classroom activities, such as:

- » the nature and frequency of interactions between the classroom teachers and RR teachers;
- » the content of regular classroom literacy instruction;
- » classroom teachers' understanding of RR philosophy and procedures; and,
- » classroom teachers' perceptions of the impact of RR on participating students' motivation and language skills.

This survey included a \$10 online gift card incentive. In May 2012, survey links were emailed to 1,347 1st-grade teachers. Of these, 587 responded to the survey (44 percent response rate).²

¹ The list of email addresses for RR teachers was updated several times during the school year. Therefore, the total number of teachers surveyed does not match the number of teachers who were sent activity logs during the course of the year.

² Unfortunately, no additional data are available for 1st-grade teachers who did not respond to the survey, thus limiting our ability to conduct non-response analyses.

RR Teacher Activity Logs

The full population of 1,244 i3-supported RR teachers³ was asked to complete daily activity logs on randomly-selected days from January through April 2012. Teachers were randomly split into two groups, and each group received a total of four activity log invitations to be completed on four assigned days. The survey began with the question “Did you work today?”; those RR teachers who responded that they had worked (87 percent across all logs completed, n=2,229) were asked to complete the remainder of the survey.⁴ Response rates for a single log ranged from 48 percent to 55 percent, with 71 percent of RR teachers (n=882) completing at least one log. Of these 882 respondents, 16 percent completed only one log, 17 percent completed two logs, 23 percent completed three logs, and 44 percent completed all four logs. In total, RR teachers completed 2,623 logs documenting their daily activities from January through April 2012.

Analyses of the log data use conditional calculations to account for these response rates, and to recognize that a substantial part of the work of RR teachers is expected to vary from teacher to teacher and day to day. We first calculate the percentage of RR teachers who engaged in each activity on at least one of their logged days. This provides an indication of how many RR teachers engage in this kind of activity frequently enough to be observed in a sample of one to four days. Second, for those teachers who reported engaging in each activity on at least one day, we calculate the percentage of logged days on which they engaged in that activity. This provides an indication of how regularly these teachers engage in each activity, given that they do the activity at least some of the time. Lastly, we calculate the average amount of time per day teachers engage each activity for those days on which the activity occurs. This provides an indication of how much time per day RR teachers tend to devote to each activity on those days that the activity occurs.

Reading Recovery Intervention Document Review

The Standards and Guidelines of Reading Recovery in the United States (5th Edition) (RRCNA, 2009) describes numerous documents that should be created and retained for each lesson and each student in Reading Recovery. As such, CPRE selected a random sample of 50 Reading Recovery students, and contacted the RR teacher for each student via mail and email. The RR

3 The list of email addresses for RR teachers was updated several times during the school year. Therefore, the total number of teachers surveyed does not match the number of teachers who were sent activity logs during the course of the year.

4 Although the instructions for the activity log stated that the teacher was to complete the log for a specific date, some teachers interpreted the first question on the log as asking whether they worked on the day they chose to complete the log (i.e., not the intended log day). Responses to an open-ended follow-up question asking why a teacher didn't work included statements such as, “Because it's Saturday.” This suggests that some teachers who completed their log late may have logged the wrong day, and that the rate of 87 percent of teachers working on the intended log day is likely an underestimate. The logs for 2012-13 have been revised to include the intended date in the questions (e.g., “Did you work on January 21st?”).

teacher was asked to submit copies of RR documentation with student names redacted. These documents included, but were not limited to, the following: OS documents, running records, predictions of progress, lesson records, records of reading/writing vocabulary and text levels, and documentation related to discontinuation or referral. Each submission was reviewed and inventoried in order to record the nature and frequency of documents collected for each sampled student, and benchmark against the expectations for documentation as specified in the RR Standards and Guidelines.

Case Studies

The goal of the case study component of this evaluation was to explore how RR is being implemented in varying real-world contexts. The aim of each individual case was to understand the relationship between school context and RR implementation; that is, how school context shapes, and is shaped by: a) the school-level decisions made about RR, 2) the school-level processes used to implement RR, and 3) the school-level outcomes of RR. Included in "school context" were such elements as the school's overarching approach to literacy, existing and emergent relationships between stakeholders, and the way the work of RR was organized and executed.

These case studies were guided by the following research questions:

- » How is RR being implemented in varying real world contexts? What inputs are required? What are the costs associated with implementation? What are the varying outcomes of implementation?
- » If and how does a school context (e.g. school's overarching approach to literacy, relationships between stakeholders, organization of RR work) shape the way RR is implemented?
- » If and how does the implementation of RR shape a school's context?

In 2012, nine cases were completed on individual schools. These study sites were selected from a representative sample of i3-funded sites across nine of the 19 UTC regions—one site per UTC region. The schools were distributed across rural, suburban, and urban settings, and ranged in size, year of implementation, and demographic makeup. Each case was assigned to a single researcher, who was responsible for all aspects of gaining access to the site, collecting and analyzing data, and writing a comprehensive case study report.

Gaining Access

The first point of contact for each site was the RR teacher leader assigned to the sampled school. If the teacher leader agreed to participate, she was then asked to initiate contact between the researcher and the principal, RR teacher(s), 1st-grade classroom teachers, and the site coordinator.

Only i3-funded RR teachers were selected for data collection; in cases where a school had more than one i3-funded RR teacher, one was randomly chosen for both interviews and observations, though attempts were made to observe all RR teachers at the school.

Case Study Data Collection

A majority of the data for each case were collected during three-day site visits conducted by the assigned researcher in the spring of 2012. Site visits included observation of at least one RR lesson conducted by the RR teacher with each of her current students. Observations noted the sequence of activities through which the RR teacher guided the student, when and how the RR teacher prompted students, techniques the RR teacher used to press students on specific skills, and the proportion of the lesson dedicated to each activity. Each observation was followed by a short sequence of questions asking the RR teacher to explain why the lesson proceeded in the particular way it had, what she noticed about the student during the lesson, what choices she made in response to what she saw in the student, what about the lesson was successful, and what challenges she encountered during the lesson.

Following the observation of RR lessons, the researcher shadowed the RR teacher as she went about her other duties for the rest of the day. Observed duties included activities such as small-group pullout instruction, administering assessments, and consulting with other teachers. The goal of shadowing was to understand the RR teacher's daily routine and interactions with students and adults at the school, both related and unrelated to RR. Researchers noted the RR teacher's process of preparing for lessons, the content and tone of conversations she had with teachers and administrators regarding literacy instruction, and with whom in the building contact was most frequently made. When possible, the researcher also shadowed the RR teacher as she attended her weekly professional development session conducted by her teacher leader.

A longer interview was also conducted with the RR teacher independent of shadowing and observation. The goal of this interview was to probe the RR teacher's understanding of her role in the school, her perception of how RR was impacting students and the school overall, and to map her RR support network both inside and outside the school. More specifically, the interview included questions about:

- » the RR teacher's background as a teacher, and her choice to be trained in RR;
- » the training she receives;
- » the response she, and RR in general, have received from the school community;
- » perceptions of institutional change at the school related to the implementation of RR;
- » challenges to implementing RR with fidelity (as defined by the Standards and Guidelines);

- » how well RR aligns with the school's overall approach to literacy instruction;
- » the school- and district-level supports and oversights in place for RR, and those needed;
- » prospects for the future of RR in the school.

Researchers also conducted interviews with the principals of each assigned school. This interview focused on the history of RR adoption and implementation in the school; the influence of i3 funds on adoption and implementation; the attitudes toward RR throughout the building; the RR support structures in place; the alignment of RR with the literacy goals of the school; and the perceived "value" of RR as an investment (e.g. whether or not the significant financial commitment is worth the level of impact on student achievement).

The teacher leader responsible for the site was interviewed to learn not only about their role in the implementation of RR in the school and district, but also to bring a new perspective to comments from the RR teacher and principal. From their unique position, being both involved in the school and removed from it at different moments, a teacher leader is well positioned to verify or question the structures and interactions the RR teacher and principal describe. In addition to questions about history of RR adoption and student achievement, the teacher leader was asked questions about how well the principal of the school understands RR, what level of collaboration exists between the RR teacher and other teachers, and to what extent RR has become part of the school's approach to teaching and learning. When possible, the site coordinator who oversees the school was also interviewed. This interview generally covered the same topics as the principal interview, but from the perspective of a district-focused rather than building-focused stakeholder. The site coordinator interview also included questions regarding the broader district commitment to RR and plans for the program going forward.

The final data collection priorities during the site visit were to observe literacy instruction in two 1st-grade general education classrooms, and to interview 1st-grade classroom teachers. Two general education classrooms and teachers were chosen for data collection at each school, based on the presence of at least one current RR student in their class. Observation of literacy instruction in the general education classroom looked for activities related to specific skills (concepts about print, phonemic awareness, phonics, vocabulary, fluency, comprehension, writing) and noted in what instructional setting those skills were practiced (whole group instruction, small group instruction, individual instruction, or classwork). Particular attention was paid to the way current RR students participated in the instructional settings, and how the teacher interacted with those students compared to how they interacted with other, non-RR students.

Interviews with general education teachers focused first on their overall familiarity with RR, such as how students were chosen to participate in the program, and what happens during an RR lesson. These interviews also included questions about structures and relationships surrounding RR in the building, including how the teacher thought about her individual role

in RR implementation, her interactions with the RR teacher, what RR supports were available, how RR was generally regarded, and the level of understanding and support of RR provided by the principal. Finally, 1st- grade teachers were asked to comment on what impacts on pedagogy and student achievement they saw resulting from RR, particularly whether their own literacy instructional practices had shifted with their increasing familiarity with RR, and how the RR students in their class were progressing in comparison to other students.

Case Study Data Analysis

After interviews were transcribed, each researcher analyzed the qualitative data collected from his or her assigned site. The overarching questions used to guide this analysis were:

- » What are the different ways people are thinking and talking about RR?
- » What contextual characteristics influence the way RR is being implemented in this school, and are they affecting the school's ability to implement RR with fidelity?
- » What constitutes the RR support network at the school?

The choice of specific method of analysis (e.g. open coding, analytical memoing) was left to each researcher, but all were instructed to probe for particular themes within these data in order to answer the questions above. These themes were:

- » History of RR implementation: the impetus behind adoption, why RR was seen as an appropriate intervention to meet the needs of the school and district, who championed the adoption of RR in the district and how people responded;
- » RR team composition: who contributes to the practice of RR at the school and in what ways, their perception/characterization of their role in RR;
- » RR network structure: key relationships that shape the implementation of RR at the school;
- » Creating institutional fit: the process of integrating RR into the school learning culture, factors that supported or hindered integration;
- » Response of school community to presence of RR: the range of perceptions and attitudes in the school about the effectiveness of RR;
- » Range of impacts: impacts on students (achievement), teachers (pedagogy), and the school (overall vision, approach to literacy, data utilization practices); and,
- » Plans for sustaining RR going forward: growing or reducing the program.

Each researcher used their analysis to write a detailed case study memo of approximately 7,500

words addressing these key themes. The case memos were then analyzed as a group, looking for similarities and differences across cases that might explain variations in implementation and success.

Early Impacts on Student Achievement

This chapter presents the statistical analysis of ITBS reading test scores in order to estimate the impacts of Reading Recovery on students participating in the i3 scale-up randomized experiment during the 2011-12 school year. As described in Chapter II, HLM was used to analyze differences between the ITBS reading scores of students in treatment and control groups. The primary analyses focused on the ITBS Total Reading Scale scores, with additional exploratory analyses of impacts for the Reading Words and Reading Comprehension subscales, and also for subgroup effects on Total Reading for ELL students and students in rural schools.

School Participation and Data Availability

Of the 209 i3 schools randomly selected to participate in the RCT in the 2011-12 school year, 158 schools (76 percent) actually carried out the random assignment process. The 51 schools that did not carry out random assignment are not included in the impact analyses presented here.

Although several modes of direct and indirect communication were used to inform schools of their expected participation in the random assignment study, a substantial number of schools (i.e., 51) did not randomize. The communications included direct emails from IDEC to individual teacher leaders and RR teachers, distribution of documents describing the evaluation design to UTCs and teacher leaders, and inclusion of a video on the IDEC website describing the evaluation design and procedures. Follow-up with non-participating schools suggested that these broad communications were not sufficient to closely monitor compliance of schools selected for the random assignment study. Furthermore, the lack of continued close monitoring of randomization compliance left us with no data from 2011-12 that clearly explain why schools did not comply. As a result of this issue, additional monitoring of timely random assignment was implemented for the 2012-13 school year. Results from more recent data collection suggest that the schools that did not comply with the random assignment protocol often had legitimate reasons that were beyond the school's control. These included school closure, discontinuation

of Reading Recovery, staffing changes, data errors (e.g., duplication of schools in the list of participating schools), and miscommunication (e.g., undelivered emails). This suggests that the number of schools making a deliberate decision to not comply with the random assignment protocol is relatively small.

Within the 158 schools that implemented random assignment, a total of 1,253 students were randomly assigned to treatment (N=628) and control (N=625) conditions.¹ Of these students, a total of 1,241 (622 treatment, 619 control) were successfully matched to data recorded in IDEC. Of these students, a total of 1,002 had ITBS scores recorded in IDEC (530 treatment, 472 control). After linking treatment students to their matched controls (i.e., via the matched pairs used during random assignment) a total of 433 matched pairs in 147 schools included a treatment and a control student with ITBS data. This sample of 866 students (433 matched pairs) represented 69 percent of the students from matched pairs in schools that carried out the random assignment. The missing data at the student level primarily resulted from student mobility or other factors that led to the inability or failure to administer the ITBS tests to both treatment and control students in a pair. Matched pairs without complete data are not included in the impact analyses presented here.

The multi-site, matched-pairs design of this random assignment study means that each school and each pair is an independent mini-experiment, and that the ability to calculate valid causal impacts is less prone to problems associated with school non-participation or missing data. Although the sample size is reduced and may be less representative of the target population, the impact estimates for the reduced sample are still valid indicators of causal impacts for those schools and students that actually participated in the experiment.

Of course, the ability to generalize results to the entire population of i3 schools is a critical goal for this impact study given that the key focus is the overall impacts of the intervention on the eligible population of students when implemented at scale. As such, we performed statistical tests of differences in student demographics for students included in the impact analyses and those dropped due to incomplete data. Analyses of differences in student characteristics for those students included and excluded from the analytic sample suggest no significant differences in pretest OS Text Reading Levels ($p = .13$), gender ($p = .91$), race ($p = .67$), or ELL status ($p = .52$). As additional cohorts of students and schools participate in the RCT each year, we will collect additional data to further explore the reasons why schools and students are missing data, and the degree to which this may or may not limit generalizability of results.

¹ There are a few extra students assigned to the treatment condition in schools with less than 8 eligible students. Any eligible student without a matched pair was automatically assigned to the treatment condition, although their data is not included in any impact analyses due to the absence of a matched control.

Baseline Balance Tests

Baseline balance tests were performed in order to examine whether the treatment and control groups were equivalent on observed characteristics after random assignment. Table 3 presents results for baseline balance tests for student demographics of the final analytic sample of 866 students in 147 schools.

Table 3. Baseline Balance Tests for Student Demographics

Pretreatment Variable	Treatment Group	Control Group	p-value for Difference
Gender (n=862)			
Male	61%	61%	0.96
Female	39%	39%	
ELL Status (n=860)			
ELL	17%	18%	0.47
Non-ELL	83%	81%	
Race (n=856)			
Black	18%	19%	0.93
Hispanic	22%	20%	
White	57%	56%	
Other	3%	5%	

Source: IDEC student demographic data.

No significant differences were found between treatment and control groups on gender, ELL status, or race.² The percentages in each column match up well between the treatment and control groups, suggesting that random assignment produced treatment and control groups that were well balanced immediately prior to implementation of RR for the group of treatment students.

Table 4 shows a baseline balance test for prior reading performance of the final analytic sample of 866 students in 147 schools. Again, no significant differences were found between the treatment and control groups, and the percentages in the two columns match up well. This confirms that the treatment and control groups had initial reading performance that was nearly identical immediately prior to implementation of RR for the treatment group students.

² Free and reduced price lunch status data were not available for students included in the 2011-12 RCT. These data are being collected beginning in 2012-13.

Table 4. Baseline Balance Test for Student Pretest Reading Performance

Pretreatment Variable	Treatment Group	Control Group	p-value for Difference
Text Reading Level			
0	51%	52%	0.87
1	21%	19%	
2	17%	20%	
3+	11%	10%	

Pretest text reading level based on the Observation Survey of Early Literacy Achievement (OS).

Overall Impacts on ITBS Reading Scores

Table 5 shows simple descriptive statistics for the treatment and control groups on raw scores and scale scores from the reading sections of the ITBS. For both sets of scores, the means are over one-half of a standard deviation larger in the treatment group. Differences in percentile ranks are +16 for Reading Words, +20 for Reading Comprehension, and +18 overall.

Table 5. Descriptive Statistics for ITBS Scores for Treatment and Control Groups

Mid-Year Outcomes	Treatment Group (n=433)	Control Group (n=433)
ITBS Reading Words Raw Scores		
Mean	21.1	18.2
(Standard Deviation)	(5.1)	(5.2)
ITBS Comprehension Raw Scores		
Mean	9.9	7.8
(Standard Deviation)	(3.9)	(3.2)
ITBS Reading Words Scale Scores		
Mean	141.2	136.7
(Standard Deviation)	(9.0)	(7.6)
Mean Percentile Rank ^a	43	27
ITBS Comprehension Scale Scores		
Mean	140.0	135.5
(Standard Deviation)	(8.9)	(7.4)
Mean Percentile Rank ^a	39	19
ITBS Total Scale Scores		
Mean	139.2	135.0
(Standard Deviation)	(7.6)	(6.2)
Mean Percentile Rank ^a	36	18

^a Percentile ranks based on ITBS Grade 1 midyear norms (Hoover et al., 2006).

Statistical tests of significance of differences in ITBS scale scores between the treatment and control groups were performed using HLM models as described in Chapter II and Appendix A. Results from those analyses are presented in Tables 6, 7, and 8 below.

Analyses of impacts on ITBS Total Reading scores showed a highly significant positive effect of Reading Recovery overall. The point estimate for the difference between treatment and control students' expected Total Reading Scores on the ITBS was 4.24 with a p-value significant at greater than 99 percent confidence. Dividing that point estimate by the standard deviation of the control group yields a Glass' D effect size of 0.68 standard deviations. This effect estimate reflects the impact of RR relative to the population of struggling readers eligible for RR in participating schools. Alternatively, dividing the point estimate by the standard deviation from the ITBS 2005 national norming sample of first graders (i.e., $s = 9.1$) yields a Cohen's D effect size of 0.47 standard deviations. This effect estimate reflects the impact of RR relative to the full population of all first graders across the nation.

Table 6. HLM Analysis of Overall Treatment Effects of Reading Recovery on ITBS Composite Reading Scores

Dependent Variable: Mid-Year ITBS Total Reading Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	135.13	0.34	<.0001
Pretest (β_1)	1.85	0.21	<.0001
Treatment Effect (β_2)	4.24	0.46	<.0001
Random Effects			
Matched Pair Variance (ω^2)	3.83	1.99	0.027
School Intercept Variance (τ^2)	4.78	2.04	0.010
School Treatment Impact Variance (ξ^2)	9.64	4.02	0.008
School Intercept/Impact Correlation (ρ)	0.15	0.31	0.635
Student-Level Residual Variance (σ^2)	27.77	2.45	<.0001

Note: See Table 5 for raw means and standard deviations for the outcome variable.

Analyses of impacts on the ITBS Reading Words subscale showed similar results. The model estimates are shown in Table 7. The point estimate for the difference between treatment and control students' expected Reading Words scores on the ITBS was 4.56 points ($p < .0001$).

Dividing that point estimate by the standard deviation of the control group yields a Glass' D effect size of 0.60 standard deviations. Alternatively, dividing the point estimate by the standard deviation from the ITBS 2005 national norming sample of first graders (i.e., $s = 10.2$) yields a Cohen's D effect size of 0.45 standard deviations.

Table 7. HLM Analysis of Overall Treatment Effects of Reading Recovery on ITBS Reading Words Subscale Scores

Dependent Variable: Mid-Year ITBS Reading Words Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	136.84	0.43	<.0001
Pretest (β_1)	1.94	0.26	<.0001
Treatment Effect (β_2)	4.56	0.57	<.0001
Random Effects			
Matched Pair Variance (ω^2)	2.77	3.05	0.182
School Intercept Variance (τ^2)	8.20	3.15	0.005
School Treatment Impact Variance (ξ^2)	12.39	5.86	0.017
School Intercept/Impact Correlation (ρ)	-0.06	0.33	0.854
Student-Level Residual Variance (σ^2)	46.67	4.02	<.0001

Note: See Table 5 for raw means and standard deviations for the outcome variable.

Analyses of impacts on the ITBS Reading Comprehension subscale also showed similar results. The point estimate for the difference between treatment and control students' expected Reading Comprehension scores on the ITBS was 4.51 with a p-value significant at greater than 99 percent confidence. Dividing that point estimate by the standard deviation of the control group yields a Glass' D effect size of 0.61 standard deviations. Alternatively, dividing the point estimate by the standard deviation from the ITBS 2005 national norming sample of first graders (i.e., $s = 10.2$) yields a Cohen's D effect size of 0.44 standard deviations.

Table 8. HLM Analysis of Overall Treatment Effects of Reading Recovery on ITBS Reading Comprehension Subscale Scores

Dependent Variable: Mid-Year ITBS Reading Comprehension Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	135.68	0.38	<.0001
Pretest (β_1)	2.31	0.24	<.0001
Treatment Effect (β_2)	4.51	0.58	<.0001
Random Effects			
Matched Pair Variance (ω^2)	2.95	3.11	0.171
School Intercept Variance (τ^2)	2.02	2.08	0.167
School Treatment Impact Variance (ξ^2)	11.84	4.57	0.005
School Intercept/Impact Correlation (ρ)	0.00	n/a	n/a
Student-Level Residual Variance (σ^2)	48.72	3.81	<.0001

Note: See Table 5 for raw means and standard deviations for the outcome variable.

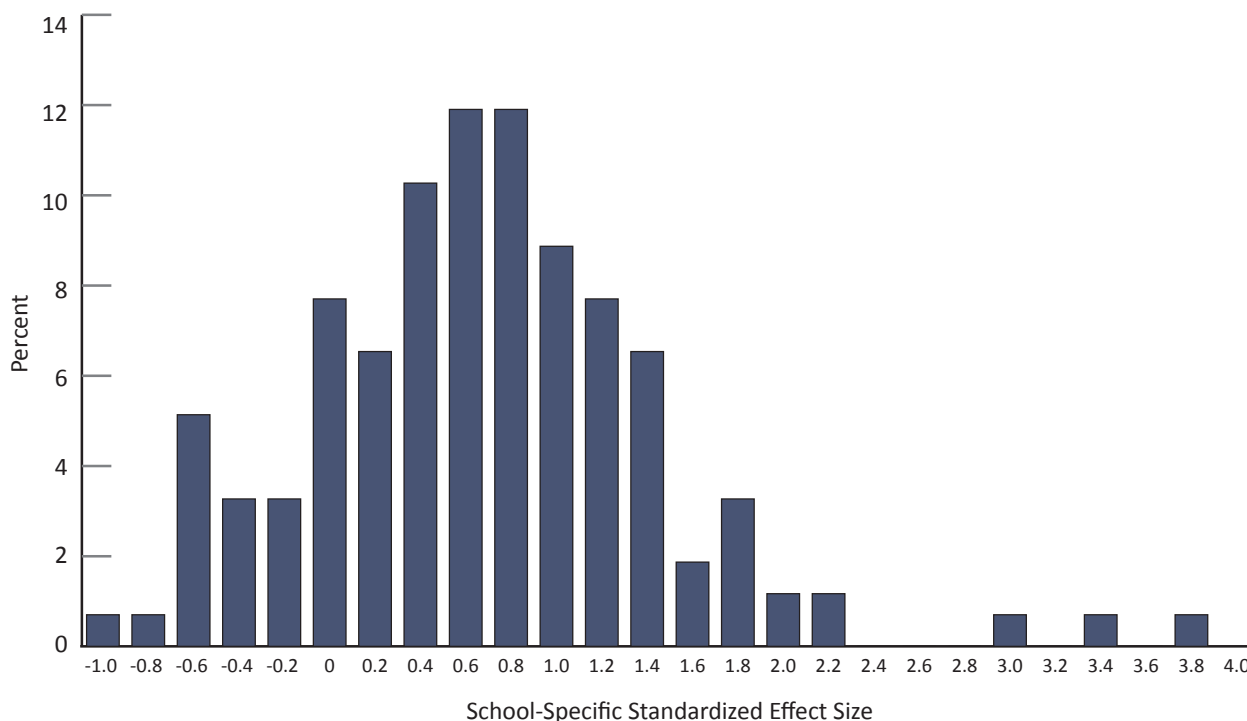
The standardized effect sizes between .60 and .68 standard deviations are large relative to typical effect sizes found in educational evaluations. For example, the impacts of Reading Recovery here were 5.7 times larger than the average effects of Title I programs reviewed by Borman and D’Agostino (1996). Even when benchmarking the impacts on ITBS scores relative to the full population of first graders in the nation, the standardized effect sizes between .44 and .47 standard deviations are still large (3.9 times the average effect of Title I programs).

It is also helpful to benchmark the treatment effects against expected gains on the ITBS for the national sample of students used to norm the ITBS tests. This allows calculation of impacts as an increase in growth rate during the study period. From the start of first grade through the fifth month (i.e., the period during which the treatment students received RR instruction), ITBS Reading Total scale scores are expected to increase from 133 to 144 for the average student in the US (Hoover et al, 2003). This increase of 11 points over a 5-month period suggests that the additional gains of 4.2 points experienced by Reading Recovery students is equivalent to an additional 1.9 months of learning. Alternatively, the additional 4.2 points translates to a growth rate that is 38 percent greater than the national average growth rate for beginning first graders.

School-Level Variation in Overall Impacts Estimates

The highly significant variance components for random effects in the HLM models of impacts on ITBS scores suggests that the magnitude of the impact estimates of RR varies substantially across schools. The results for the overall impact model in Table 6 show an average effect of +4.24 points, with a random effect covariance estimate for the school-level impacts of 9.64 points. Taking the square-root of this covariance estimate yields a standard deviation of 3.1 points. Figure 6 below shows a histogram of adjusted empirical Bayes estimates³ of school-specific impact estimates for the 147 schools included in the overall analysis. Eighty-one percent of the schools have positive impact estimates, the modal impact is .7 standard deviations, and 29 percent of schools have impact estimates greater than one full standard deviation. There are also a few positive outliers, suggesting that a few schools have very large positive impact estimates in excess of two full standard deviations.

Figure 6. Adjusted Empirical Bayes Estimates⁵ of School-Specific Impacts of Reading Recovery



³ Empirical Bayes (EB) estimates are adjusted to address over-shrinkage (see Bloom, Raudenbush, and Weiss, 2012; Raudenbush & Bryk, 2002, p. 88) by multiplying each EB estimate by the ratio of the estimated true impact variance (i.e., the random effect covariance estimate for the school-specific treatment effect) to the variance of the raw empirical Bayes estimates.

Impacts on ITBS Reading Scores in Rural Schools and for English Language Learners

The i3 Scale-Up of Reading Recovery included a specific focus on rural schools and ELLs. Exploratory analyses of treatment effects on ITBS Total Reading Scores for these two subgroups are presented below.

Table 9 shows simple descriptive statistics for ITBS Total Reading Scores of students in rural schools by treatment and control groups. Once again, the means are over one-half of a standard deviation larger in the treatment group.

Table 9. Descriptive Statistics for ITBS Scores for Treatment and Control Groups in Rural Schools

Mid-Year Outcomes	Treatment Group N=73	Control Group N=73
ITBS Total Scale Scores		
Mean	140.9	137.4
(Standard Deviation)	(7.1)	(6.8)

Statistical tests of significance of differences in ITBS Total Reading scores between the treatment and control groups in rural schools were performed using HLM models as described in Chapter II and Appendix A. Results from that analysis are presented in Table 10.

Table 10. HLM Analysis of Overall Treatment Effects of Reading Recovery on ITBS Composite Reading Scores in Rural Schools⁴

Dependent Variable: Mid-Year ITBS Total Reading Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	136.75	0.72	<.0001
Pretest (β_1)	2.55	0.38	<.0001
Treatment Effect (β_2)	4.01	0.97	<.001
Random Effects			
Matched Pair Variance (ω^2)	6.20	4.49	0.083
School Intercept Variance (τ^2)	1.92	3.12	0.269
School Treatment Impact Variance (ξ^2)	5.62	5.07	0.134
School Intercept/Impact Correlation (ρ)	0.00	n/a	n/a
Student-Level Residual Variance (σ^2)	23.45	4.61	<.0001

Note: See Table 9 for raw means and standard deviations for the outcome variable.

4 Based on 146 students in 24 schools.

The results for rural schools were very similar to the overall results. Analyses of impacts on ITBS Total Reading scores showed a highly significant positive effect of Reading Recovery in rural schools. The point estimate for the difference between rural treatment and control students' expected Total Reading scores on the ITBS was 4.01 with a p-value significant at greater than 99 percent confidence. Dividing that point estimate by the standard deviation of the control group yields a Glass' D effect size of 0.59 standard deviations.

Because of the reduced sample size, the estimates for all of the covariance parameters are no longer statistically significant. This is not surprising given that only 24 rural schools out of the 152 rural schools in the i3 scale-up in the 2011-12 were sampled for the random assignment study. As more data are collected and pooled across cohorts in subsequent years, we expect to see significant variance in effects for this subpopulation as well, and we plan moderation analyses to explore school and contextual factors that explain variation in impacts across rural schools.

Table 11 shows simple descriptive statistics for ITBS Total Reading Scores of ELL students by treatment and control groups. Once again, the means are approximately one-half of a standard deviation larger in the treatment group.

Table 11. Descriptive Statistics for ITBS Scores for ELL Students by Treatment and Control Groups

Mid-Year Outcomes	Treatment Group N=105	Control Group N=105
ITBS Total Scale Scores		
Mean	137.9	134.5
(Standard Deviation)	(7.3)	(5.6)

Statistical tests of significance of differences in ITBS Total Reading scores of ELL students in the treatment and control groups were performed using HLM models as described in Chapter II and Appendix A. Results from that analysis are presented in Table 12.

Table 12. HLM Analysis of Treatment Effects of Reading Recovery on ITBS Composite Reading Scores for ELL Students⁵

Dependent Variable: Mid-Year ITBS Total Reading Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	135.48	0.61	<.0001
Pretest (β_1)	2.54	0.49	<.0001
Treatment Effect (β_2)	2.88	0.85	0.001
Random Effects			
Matched Pair Variance (ω^2)	2.50	3.72	0.251
School Intercept Variance (τ^2)	4.18	3.11	0.090
School Treatment Impact Variance (ξ^2)	12.25	6.17	0.024
School Intercept/Impact Correlation (ρ)	0.00	n/a	n/a
Student-Level Residual Variance (σ^2)	22.87	4.34	<.0001

Note: See Table 11 for raw means and standard deviations for the outcome variable.

The results for ELL students were also very similar to the overall results. Analyses of impacts on ITBS Total Reading scores of ELL students showed a highly significant positive effect of Reading Recovery. The point estimate for the difference between rural treatment and control students' expected Total Reading Scores on the ITBS was 2.88 with a p-value significant at greater than 99 percent confidence. Dividing that point estimate by the standard deviation of the control group yields a Glass' D effect size of 0.51 standard deviations.

Despite the reduced sample size, the estimates for the estimate of school-level variability in treatment effects is still statistically significant at the 95 percent confidence level and the variance here is even larger than the effect variance in the overall analysis. This suggests that while the vast majority of schools' Reading recovery programs have positive impacts on ELL students' reading performance, they vary greatly in their ability to produce sizable impact estimates for ELL students.

⁵ Based on 191 students in 63 schools.

Fidelity to Reading Recovery Standards and Guidelines

.....

The successful implementation of Reading Recovery and the success of the i3 scale-up depend on fidelity to the Reading Recovery model. A major research objective in the scale-up evaluation is to measure the amount and types of variation in implementation fidelity. To do this, CPRE evaluated the fidelity of the core components defined by the established Reading Recovery model, in the i3 supported scale-up sites. This component of the study is designed to produce objective quantification of implementation fidelity.

Fidelity data were collected via online surveys with RR teachers, teacher leaders, and site coordinators, as described in Chapter II Evaluation Study Design and Methods. Implementation fidelity results are presented in Table 17. First reported are the number of respondents and response rates calculated based on the population data that CPRE received from IDEC. Response rates for RR teachers, teacher leaders, and site coordinators were 73 percent, 88 percent and 58 percent respectively. The low response rate for Site Coordinators is likely due to the fact that their involvement in Reading Recovery is much less intense than that of teacher leaders and RR teachers (hence a lower incentive to respond), and that no monetary response incentives were provided to any of these groups. Table 17 also presents the percent of standards and guidelines that were met by respondents, overall and by subdomain. We have grouped standards into domains based on the inherent structure of the Standards and Guidelines, with some further consolidation by CPRE to guide interpretation of variation.

It is clear from the average percent of standards met that the Reading Recovery model is being implemented with high fidelity by RR teachers, teacher leaders, and site coordinators. On average, RR teachers met 95 percent of the standards, while teacher leaders and site coordinators met 87 percent and 88 percent of the standards respectively. This overall finding suggests that fidelity to the model is high, although it is difficult to reach absolute fidelity. Similarly, the guidelines were found to have high fidelity across the three roles, with a similar pattern of relative strengths across the domains (discussed below).

Overall, CPRE found that essential implementation and instructional standards were met with the highest fidelity. Specifically, those standards that pertained to site management, recruitment, and student instruction were implemented according to the Reading Recovery model. However, those standards and guidelines that pertain to growth and sustainability had relatively lower fidelity rates. For example the preparation of annual reports by teacher leaders and the nurturing of communication among district players by site coordinators tended to have lower fidelity rates. Future work will explore these trends and investigate barriers to full adherence.

Table 17: Implementation Fidelity of i3 National Scale-up of Reading Recovery

	RR teachers	Teacher Leaders	Site Coordinators
# Respondents	742	168	95
Response rate	73%	88%	58%
% Standards Met			
Overall (percent of standards met) ^a	95%	87%	88%
Selection	100%	100%	
Training	81%	73%	
Reading Recovery Lessons	98%		
Data and Monitoring	99%	100%	
Leadership		94%	
Research		68%	
Professional Development		98%	68%
Background			81%
Site Preparation & Maintenance			96%
Communication			79%
% Guidelines Met			
Overall	82%	68%	82%
Selection	73%	99%	
Training	84%	95%	
Reading Recovery Lessons	95%		
Data and Monitoring	78%	83%	
Leadership		88%	
Professional Development		56%	64%
Background			90%
Site Preparation & Maintenance			87%
Communication			72%

Source: Surveys of RR Teachers, Teacher Leaders, and Site Coordinators.

^a Overall percentages are calculated as unweighted averages across all standards or guidelines. Given that the number of standards and guidelines in each domain varies, these percentages are not a simple average across domains.

The selection of RR teachers and teacher leaders were found to have complete fidelity (i.e., 100 percent of teacher selection standards met). Additionally, near complete fidelity was reported for standards that pertain to conducting Reading Recovery lessons, suggesting that, as reported by RR teachers, the 1-on-1 lessons are being implemented as required. This finding is consistent with data from the RR teacher activity logs which indicate that RR teachers are largely meeting their daily scheduling expectations for instruction and the activities during the one-on-one lessons.

Across the ten categories of standards, training was the category for which RR teachers reported the lowest level of fidelity, with 81 percent of standards met. Teacher leaders also reported relatively lower fidelity in meeting the training standards (73 percent). Previously reported data on logistical impediments to some training requirements (discussed in Chapter 6) may help explain this numbers. For instance, some teachers must travel long distances to attend training classes, and the use of hybrid classes that meet in person only every other week limits the frequency of behind-the-glass training sessions for some teachers.

Teacher leaders also reported relatively lower fidelity to standards (68 percent) in the area of research, which largely pertains to the role of trained teacher leaders in preparing annual reports. CPRE will further explore this issue going forward, though current data lead us to suspect time constraints may be the primary factor.

Site coordinators reported high fidelity when performing site preparation and maintenance; 96 percent of standards were met in this category. These responsibilities relate to establishing and maintaining an RR site, which includes attending to fiscal matters, program monitoring, tasks related to training and oversight of personnel, and resource provisioning. While the site coordinators reported high fidelity to standards, many teacher leaders reported that they did not see their site coordinator as playing a key role in implementation. CPRE will continue to explore the site coordinator's role and the importance of their tasks to overall implementation.

The areas with lowest fidelity for site coordinators were professional development and communication; in these categories, 68 and 79 percent of standards were met, respectively. For site coordinators, professional development standards require ongoing learning about Reading Recovery. Their low compliance with these standards suggests that some site coordinators may not be fully informed about the intervention. Also, given that a key role of the site coordinator is advocating on behalf of Reading Recovery, the relatively low fidelity in the area of communication may be noteworthy.

Early Implementation Findings: Recruitment

Efforts to expand are not new to the RR network; over a nearly thirty-year history in the U.S., RR has relied on collaboration, advocacy, and outreach to facilitate proliferation of RR implementations nationally. The program began with 110 students in a single Ohio school district and had expanded to 152,000 school districts by the program's peak, in the 2000-2001 school year, without mass-marketing efforts or specialized sales staff.

During the past decade, the Reading Recovery community has experienced a severe decline in the number of school-level implementations and students served nationally. Consequently, receipt of the i3 award has been embraced by the RR community as an opportunity to increase the visibility of RR in the U.S. and grow the number of RR implementations.

During Year 1 of the i3 evaluation, questions regarding quality of RR implementation and student impacts were coupled with an exploration of the systems and methods leveraged by the RR community to meet the expansion objective of the scale-up. Key to scale-up is recruitment of new RR personnel, a process we explored through Interviews with UTC and focus groups with teacher.

Key findings regarding recruitment under the i3 scale-up include the following:

- » Over Years 1 and 2 of the evaluation, the Reading Recovery system successfully recruited and trained 1,121 i3 RR teachers (76 percent of the intended goal) in 847 i3 schools (144 percent of the intended goal), who provided RR instruction to 15,525 students (88 percent of the intended goal).
- » Each UTC region is a largely independent system with its own norms, policies and standardized processes. UTC directors are at liberty to design and carry out the recruitment strategy they decide is appropriate for their regional context. These strategies often leverage recruitment activities undertaken in the past, and are intended to respond to each region's unique characteristics.

- » UTC directors have taken a collaborative approach to the regional management of the i3 scale-up; yet, their efforts have primarily drawn on regional resources and have been supported by regional stakeholders.
- » UTC directors share strategies and cooperate with teacher leaders to increase outreach. Teacher leaders typically play a major role in recruiting schools and teachers in local contexts.
- » UTC directors experienced varied challenges to recruitment as a result of each region's unique context. The most common and persistent challenge was identified as the weak economic climate.
- » UTC directors reported their intentions to evolve traditional recruitment efforts by placing additional emphasis on recruitment strategies that involve personal contacts.

i3 Scale-up Recruitment Outcomes

Along with exploration of the processes supporting recruitment for the i3 scale-up, CPRE examined the outcomes of recruitment efforts. Over Years 1 and 2 of the evaluation, the Reading Recovery system recruited and trained 1,121 RR teachers in 847 schools, who provided RR instruction to 15,525 students. Comparing these numbers to the original recruitment goals for the project – 1,470 teachers in 588 schools, reaching 17,640 students over the first two years – reveals that recruitment of RR teachers was 24 percent below target, recruitment of schools was 44 percent above target, and instruction of students was 12 percent below target.

Of the 847 schools that received i3 funds in 2011-12, 536 used the funds to train new RR teachers in order to increase capacity to provide Reading Recovery to additional students in need. The remaining 311 schools used i3 funds to implement Reading Recovery for the first time. The largest number of i3 schools are found in suburban areas (237), followed by rural areas (225), cities (205), and towns (133).

In addition to locale, i3 schools also vary on characteristics such as participation in Title 1, percentage of ELL students, and percentage of free and reduced lunch participants. For example, 660 i3 schools were eligible for Title I funds, 245 schools had five percent or more of their students participating in ELL programs, and 639 i3 schools had at least 20 percent of their students take part in the free or reduced lunch program.

Although CPRE cannot speak to the specific differences between i3 and non-i3 schools with the data available (i.e., non-i3 schools are not included in this study), we examined the differences between i3 schools supporting new RR implementations and i3 schools that expanded their RR implementations; expanded implementations are defined as i3 schools that were supporting RR prior to using i3 grant funds to train RR teachers. Within the population of i3 schools, there is a small but significant difference in the average percentage of students receiving free or

reduced-price lunch in schools that had RR prior to the grant versus schools that were using the grant funds to implement RR for the first time; 46.6 percent of students in schools with existing RR programs participated in the free- or reduced-price lunch program, compared to 50.7 percent of students in new-to-RR schools ($p < 0.03$). This suggests that the expansion of i3 into new schools is reaching a more economically disadvantaged segment of the population.

Recruitment Roles

In an effort to meet the system-wide commitment to recruit 735 new RR teachers each year, and 15 new teacher leaders over the course of i3's five-year grant term, UTC directors have intensified efforts at expansion. As managers of the regional hubs that are central to the national RR network, UTC directors are solely responsible for leading regional efforts to cultivate awareness of the i3 scale-up and secure the participation of districts, schools, and teachers. CPRE hypothesized that UTC directors would work closely with teacher leaders to recruit new teachers, schools, and districts to the program. In reality, the allocation of recruitment roles for the i3 scale-up varied widely from region to region.

In some regions, the UTC director set the strategy and tightly orchestrated recruitment activities. These directors often reported spending significant time traveling and devoting considerable effort to personal contact with prospective participants. A number of UTC directors concluded that their own deep involvement was the key to successful recruitment. "It's slow going because it's so personal," one director reported. "I do find the best way is to go in there and teach a class myself. So I spend a lot of time on the road and in hotels." In other regions, teacher leaders assumed primary responsibility for recruitment, with some identifying their own prospects and developing their own recruitment strategies. As one university director explained:

We just said [to the teacher leaders], "OK, here is an opportunity and this is up to you if you want this opportunity to work." And so they took the bull by the horns and they went after it. And they would say "Do we have a certain number that we need to stop at, or whatever?" and I would say "No, whatever you can do, you go for it." And so most of them did.

The extent to which directors assumed personal responsibility for recruitment versus delegating to teacher leaders was a function of personal preference and their other competing responsibilities. Additionally, the fact that teacher leaders were positioned within school districts often led UTC directors to view recruitment activities as a natural extension of a teacher leader's responsibilities. In some regions, however, teacher leader shortages made director involvement in recruitment a practical necessity. "The teacher leaders and I, and the site coordinators, kind-of do [recruitment activities] together," reported one director. "But there's really a very small resource of teacher leaders, so it's quite challenging."

As the first point of RR contact for schools and districts, many teacher leaders reported tremendous, and tremendously varied, changes in their daily work lives as a result of i3 recruitment. Some teacher leaders reported exerting considerable effort to little avail; many struggled to recruit enough i3 teachers to fill a training class. Others were inundated with i3 requests. One teacher leader recalled:

I was the main contact person for superintendents, principals, and other administrators to contact. Sometimes they contacted the secretary at the university and she would immediately send them to me. I explained things like the Memorandum of Agreement, the responsibilities, the three years responsibilities, and so on. And then brainstormed lots of ways for them to find ways to pay for the positions that they were having to create. I was the main person that kept track of all of those [i3] phone calls. And I have to say they numbered in the ten thousands I would say, probably. I mean I was on the phone every day. I mean because they were calling back! [It] was a lot.

Management and Oversight of the Recruitment Process

As part of the i3 scale-up, each UTC is working to reach pre-defined, annual recruitment targets, and UTC directors are keenly aware of their own regions' progress toward those goals. Despite these benchmarks, UTC directors have avoided implementing formal processes for managing or monitoring recruitment activities in their regions. For most, coordinating recruitment is an informal process characterized by frequent, mostly unscheduled phone and email communication with teacher leaders, who have significant autonomy in managing their own recruitment efforts.

While most UTC directors report using conference calls and professional development meetings as opportunities to touch base about recruitment progress and strategy, they express little interest in tightly monitoring teacher leaders' activities, or in implementing strict accountability measures around recruitment. As one UTC director reported:

There really isn't any oversight. I don't know if that is a term that would apply to the way it works in our state. You know, we meet twice a year. At the beginning of the year we have a conference call. We talked about recruitment effort, and all I'm asking them to do is to make the contact in the district... First of all, they're busy, too. And that's not their job as far as I'm concerned.

UTC directors' approach to managing recruitment therefore seems more collaborative than directive, characterized primarily by sharing strategies with teacher leaders, and cooperating around outreach. To the extent that university trainers provide feedback to teacher leaders related to their recruitment efforts, it comes mostly in the form of problem-solving support and encouragement. As one trainer explained: "I'll tell the teacher leaders, 'Well, don't be dispirited.

Let's go back in another year. Let's look at their data.'" This approach is reflective of a particularly striking feature of the Reading Recovery system. Generally speaking, it values faith in individual professionalism and judgment over hierarchical monitoring and accountability systems. The presumption, UTC directors explained, is that teacher leaders are doing everything they can to recruit new schools.

Recruitment Strategies

During Years 1-2 of the i3 scale-up, most UTC directors reported that the strategies they used for recruitment closely resembled outreach activities they had undertaken in the past. UTC directors also reported that, although they mostly drew upon a shared set of established outreach activities, they each developed their own i3 recruitment programs unique to their region.

UTC directors layered and combined recruitment strategies in distinct ways and, to varying degrees, supplemented them with other efforts. Some regions relied primarily on mass outreach, for instance, while others emphasized personal contacts within existing networks. As a result, recruitment efforts differed vastly from region to region. One UTC director described her recruitment efforts as heavily reliant on flyers and brochures created by RRCNA, and her own role as primarily responsive to schools' inquiries. Others reported doing significantly more direct outreach: "I probably start with phone calls telling people about the new initiative, follow up with an email and in that email try to make an appointment for a visit," explained one UTC director. "And then if there was no response, I would follow up with a telephone call."

Targeted Recruitment Efforts

Many UTC directors focused their recruitment efforts on existing sites they regarded as having expansion potential; districts and schools with existing implementations were encouraged to use i3 funds to add more RR teachers. Another common strategy was to use i3 funds to reinvigorate flagging implementations, or to reinstitute Reading Recovery in schools where it had existed previously. "Out of [the new sites] that we recruited for this year, I would say that we are just in a handful of new districts," reported one UTC director. "But what we were able to do is go back to some of the other schools where they have lost implementation one way or another and go back and train some new people." This was deemed an acceptable mode of scaling up by OSU and OII, given that Reading Recovery would be provided to students in need who would otherwise not have received the intervention.

The i3 grant's prioritization of low-performing schools, rural schools, and schools with significant numbers of ELLs led many UTC directors to narrow the focus of their recruitment outreach more than in the past. Directors reported combing through state data to identify schools that met the grant's priorities, and targeting those schools first. One director reflected:

I'm reaching out more proactively in a more targeted way, through analyzing the data off of the websites and that sort of thing to decide "This would be a good area. They've got a lot of low performing schools. They've got a lot of ESL kids here. I'm going to reach out to them with some specifics about reasons why they might want to participate."

Many directors also reported intensifying efforts to craft recruitment messages that address the challenging current economic environment. For several UTC directors this meant brainstorming realistic solutions to schools' individual fiscal quandaries. One trainer explained:

One thing we've had to talk [about] with just about every district we've recruited is how do you staff Reading Recovery with the existing personnel and what would that look like, and what are the different ways you can do that?

Across the board, UTC directors described a long-range perspective on recruitment; schools or districts that seemed interested might not adopt Reading Recovery this year, but would be regarded as prospects for the future. In some regions, much of the work of recruitment amounted to simply educating prospective schools and districts about the program with the hope that small efforts would yield future growth.

Leveraging Existing Resources

Conferences, meetings, networking events, and professional affiliations became important recruitment avenues for some UTC directors who used them as opportunities to increase awareness of the i3 scale-up. One UTC director shared i3 information via a statewide email listserv. Another noted: "Not every state has this, but we have a state Reading Recovery conference that's held in the city, and we draw from across the state... So that's another way that we're recruiting."

A few UTC directors benefited from prior linkages between RR and other instructional initiatives when recruiting districts and schools. As one trainer explained, "There's another literacy initiative... that requires a school have Reading Recovery in order to implement it. And [one district] is training three [RR] teachers this year because they want to be a part of that." In several regions, university-based comprehensive literacy initiatives – for instance, Fountas & Pinnell's Literacy Collaborative and Dorn's Partnerships in Comprehensive Literacy – were important platforms for RR implementation. Often well-known and highly regarded, these comprehensive literacy initiatives include broad-based professional development and instructional programming designed to unite teachers and administrators behind a shared philosophy of literacy instruction that is similar to Reading Recovery's approach.

One notable difference across regions was the extent to which i3 recruitment efforts leveraged political and funding resources. This variability was in large part due to the significant differences in regional early-literacy infrastructures. CPRE found that there are several regions that enjoy

significant state-level political support for Reading Recovery and early-literacy programming. While much less prevalent, some regions are also afforded access to funding through state-level budget allocations. In regions where political support exists, UTC directors report devoting considerable energy to nurturing it. One director described making annual visits with her teacher leaders to the Washington, D.C. office of a supportive politician.

Recruitment efforts also varied in the extent to which they were informed by long-range strategic planning; some trainers were able to leverage relationships with influential supporters that they had spent years developing. One UTC director recalled:

Our school has a consortium with connections to a couple of big urban districts in the state, and before I even applied to participate in the [i3] grant, I talked with the director of that, who is at the state university, explaining what the grant was and asking if he thought this would work for those schools. And he did. So we've been working through [that relationship] to have entrée to some of the big urban schools in our state.

Despite the regional differences in their recruitment efforts, all UTCs emphasized the opportunity the i3 grant provides to train RR teachers at no cost. Mass mailings to new prospects and, in some cases, specially designed websites, were used to inform schools and districts of this opportunity. Some directors found that the i3 funding stream provided them access to decision-makers who might not have otherwise been receptive. Other directors found success in emphasizing the prestige of the federal program itself. "The tuition, coursework, the materials, the books—the teachers love all of this; we sold them right away," one director recalled. "But it's getting the administrators on [board] and that's why I always use this piece about 'This is an i3 grant. This is a grant from the federal government,' and all of a sudden they start to listen."

In future years of this evaluation CPRE will explore the relationships between recruitment strategies, context, and regional recruitment outcomes. Findings will not be used to compare UTC regions, but will be reported thematically and used to inform our understanding of effective recruitment strategies and contextual variation across the entire RR system.

i3 Recruitment Challenges

The i3 scale-up has been described by many UTC directors as an infusion of new life for Reading Recovery; however, it has also had unintended negative consequences. Receipt of the i3 grant and efforts to bring Reading Recovery to scale have intensified existing challenges, and to varying degrees undermined regional recruitment efforts. The challenges to recruitment under the i3 scale-up during Years 1 and 2 were an outgrowth of each region's unique fiscal, socio-political, and educational context.

Challenges Related to the i3 Grant

UTC directors reported multiple challenges directly connected to the receipt of i3 funding. During Year 1, recruitment efforts were constrained by the timing of the i3 award and questions regarding the school selection criteria set forth by USDOE. Confirmation of receipt of the i3 grant in the fall of 2010 occurred two months after training for new RR teachers typically begins. As a consequence, UTC directors were precluded from using i3 funds to attract districts and schools during the height of the recruitment season in late spring and early summer 2010. One UTC director discussed the consequences of this timing:

Year 1 was an extremely late start. Things were not finalized until October. In my state, that is one month after school begins, and school districts have their budgets, personnel, etcetera set. People are assigned to positions. It's difficult for schools to shift gears after a month of school.

A full month into the school year, many UTC directors were unable to secure the participation of schools and districts, or were reluctant to adapt the model for training RR teachers by initiating training classes mid-year. In either case, the only option available to directors was to retroactively apply the i3 label to RR teachers in existing training classes. In many instances, these RR teachers-in-training were not educators in schools that met USDOE's priority criteria - another impediment to Year 1 recruitment. A UTC director explained:

The issue is that I had more teachers in training than could qualify for i3. In Year 1, the third priority of any school didn't exist. When I could, I did retroactively bring people in at the very last minute. If we would've had a little more flex on [the priority criteria], I would have hit a higher number in Year 1.

Other, less frequently noted challenges related to i3 logistical issues, or what one UTC director referred to as "teething problems related to the protocols, paperwork, policies and a number of changes occurring." These challenges included confusion over research requirements for the impact evaluation and the graduate admissions process for RR teachers; however, often these problems were resolved before Year 2 began.

Lack of Capacity

During Year 2 of the i3 scale-up, the much anticipated reinvigoration of Reading Recovery was constrained by issues of capacity, including time and workforce availability. "I do think that this feels like a huge opportunity for us," one UTC director explained. "It's just finding all of the time and energy that you need to meet the opportunity." The continuous recruitment process, which is characterized by a rigorous phase of direct outreach and personalized problem-solving, required significant time and presented an obstacle to the i3 scale-up. "There is one thing I learned," one UTC director shared. "This whole notion of recruiting is very time consuming because you have to spend a lot of face-to-face time."

Many other UTC directors echoed this characterization of the recruitment process as time-consuming. They reported feeling challenged by the experience of balancing their once-primary roles as university faculty and RR trainer of trainers with the responsibilities of their new role as a regional administrator of the i3 grant:

I'm doing this solo...and I'm stretched thin. I have to go over all the details with the reimbursement. In addition to my university faculty role, I'm also training. I'm trying to cover several large sites because there is no school district that would support them. I've taken them under my wing and am serving as the site coordinator...I'm wearing several hats.

The issue of feeling "stretched thin" was echoed by other UTC directors, but was heard most often from directors in regions where teacher leaders were limited in their ability to engage in recruitment activities because of restrictions imposed by a teacher leader's sponsoring school district or an already over-saturated workload:

[Recruitment] is challenging because we had such a personal approach. It was challenging to get to that many schools because there are so few [teacher leaders], and they all have very high workloads. We don't just train teachers in Reading Recovery. My teacher leaders are teaching Reading Recovery classes, small group classes, literacy lessons classes, literacy processing classes. They are teaching a lot of courses, and all of that has ongoing professional development.

Another UTC director described the root of teacher leader constraints in her region this way:

We have administrators who do not want teacher leaders to recruit, which is really interesting since that's always been their job. It really has always been their job...This particular site that I'm thinking of has had Reading Recovery for twenty years. They just completed the twentieth year. The teacher leader has done a lot of recruiting over the years...I was called in to speak to the General Education Directors, presumably about another initiative that I'm starting up, and when I got there, I got all these questions about the i3 grant...the Site Coordinator from this site was new in that role. She had been in that system for a long time but had not been a Site Coordinator. She very clearly said, "Well we can't expect our teacher leaders to recruit."

While many UTC directors reported playing central roles in generating interest for RR among districts and schools, it was universally acknowledged that the ability to move forward with securing their commitment was contingent on the availability of teacher leaders to monitor and support site-level implementations. A dearth of teacher leaders in reasonable proximity to recruitment hotspots, and the over-saturation of teacher leader workloads, led many UTC

directors to limit the geographical areas in which they attempted to recruit or to turn away inquiring systems. One director explained:

I would say there have been limitations to what I've been able to do...In the southern part of the state we've lost our Reading Recovery training center. It was an economic decision. A new superintendent closed the center overnight. So, my closest teacher leaders are in the western part of the state and their hands are absolutely full. Their plates are overflowing with their own big population of clients.

Difficult Fiscal Climate

The downturn of the U.S. economy has also had resounding consequences for the recruitment efforts of UTC directors across the country. The loss of previously established regional training sites, the inability to hire staff at the university level, and the limited ability to train additional teacher leaders to fuel recruitment were issues directly attributed to fiscal constraints at the local, regional, and state levels. The most widely articulated impacts of the fiscal climate were the resulting cuts in staff, and the subsequent resistance of school and district-level administrators to the idea of implementing a new program or hiring new staff. A UTC director commented:

I think that the biggest issue is just the funding. We can offer training, which everyone is so excited about, but with the economy being the way it is—so tight. The teachers are in very tight positions now. Before, we used to have several reading teachers in a school, and now we're lucky if we just have maybe one. Even though we have this great funding with the training, administrators will say, "Can we have that cost for the salary?," and we have to come back and say, "No, this is just about training."

Another director described the implications of financial uncertainty for recruitment efforts in her state this way:

Last year we thought that we did pretty well. We had a lot of interest, but we have a state that's in financial difficulty, so, in the spring of this year, there was no state budget...They didn't know, even in June when the school year ended. I was trying to keep in touch with them and confirm that yes, they were going to sign a memorandum of agreement, participate, and send their teachers, but a school that was going to send six backed out. A school that was going to send four backed out. A school that we talked with that had interest and had three elementary buildings and wanted to work in all three buildings, they backed out because no one knew what their budget was.

Many UTC directors' recruitment efforts included working with districts and schools to problem-solve systemic implementation issues. Against the backdrop of an anemic fiscal climate, directors reported spending significant time negotiating with districts and schools to organize personnel and allocate their time to facilitate implementation of RR. With varying levels of success, directors battled against the commonly-articulated belief that Reading Recovery is "too expensive", and worked hard to convince schools and districts that hiring additional staff is not a requirement for adopting RR. A UTC director explained:

There is one thing that we've had to discuss with just about every district that we've recruited. That is how do you staff Reading Recovery with existing personnel, what would that look like, and what are the different ways you can do that.

Gaining Access, Combating Misconceptions

The importance of influencing decision-makers was another key theme that arose during discussions of the challenges to regional recruitment efforts. UTC directors worked thoughtfully to identify and connect with decision-makers such as principals and curriculum directors within targeted systems – an often arduous task. One UTC director lamented that she was unable to gain access to contact information from a centralized source, "I can't get e-mails, principals' e-mails or anything like that." While others found that their solicitations were simply ignored:

We sent e-mail blasts to superintendents, to all the districts in the state. We pulled from the state list of curriculum directors, Title One directors, superintendents and that kind of thing, but we didn't get many hits from that. Interest mostly came from our sites where Reading Recovery was going on and from schools that knew about it.

In those cases where directors were able to successfully communicate with decision-makers, many reported working diligently to address several recurrent issues: lack of awareness of Reading Recovery or misconceptions about the model, preference for alternate intervention programs, and changes in district-level leadership and philosophy. One director commented:

A new superintendent was hired, and it was an internal hire.... It's unfortunate because [the previous administrator] was a huge advocate and that's a big district that had two teacher leaders years ago; it was ripe for scaling up. I haven't given up, but that was kind of a blow.

Many UTC directors reported encountering situations that required them to formulate counterarguments to address misconceptions and biases:

Quite honestly, I think Reading First hurt us a lot. Because there were people in the Reading First initiative from the State Department who would go out and talk to administrators and say that Reading Recovery didn't work. It was a whisper campaign. I mean, I think everybody had some experience with that. It was very hard to uncover this but then you'd hear from someone that this had been said...Some of the people who worked in the Department of Education when there was money for Reading First are now in key positions in districts. They don't have the same impact because you don't have the department head putting that kind of information out there. We [think], oh, that's just history. Yet, that history still lives. There are still people that do believe that Reading Recovery doesn't work because that's what they heard. You show them What Works Clearinghouse and they'll say, "Oh, Accelerated Reader got good marks!" So, they make all kinds of excuses.

Scaling-up with Sustainability

To receive i3 funds, districts, schools, and teachers were asked to make at least a three-year commitment to implementing Reading Recovery. Most UTC directors reported that many districts were wary of assuming responsibility for funding Reading Recovery for three years, and therefore declined the i3 opportunity.¹ Willing districts were asked to sign a Memorandum of Agreement (MOA) to formally document their commitment to the program. Some UTC directors also reported using additional measures, formal and informal, to judge a system's potential for implementing RR over the long-term. For instance, one director used a school's response to the lengthy application process to measure their level of commitment:

One of the things we also put in place around i3 is that we ask our schools to fill out a five or six page application. It's just for us to look at; it's just to insure that the school doesn't just say, "Oh yeah! I'm going to train a teacher. Give me this money." [We want to ensure that] they put some thought into it...It's a positive thing but at the same time it could be impacting the number of schools who are making investments. We had one school say, 'I'm not filling that thing out.' So, we didn't get them.

Although not directly asked about issues of attrition, more than half of UTC directors interviewed mentioned RR teacher attrition as a primary challenge to their recruitment efforts:

There were 7 [teachers trained] in Year 1, but before they finished their training the emergency financial manager and the district staff pulled them out of their

¹ I3 funds are only available to support a system's training of RR teachers. Additional funds are not available to support the continuing cost of providing a trained RR teacher with ongoing support by a teacher leader.

training. They said the teachers had to be assigned to do other things. So those people were left in limbo. This year, by working hard with the teacher leader and the current site coordinator, I was able to get 6 of the 7 [teachers] back in to finish their i3 training under the grant, but one of them was not permitted to continue the training. So, hopefully they will finish by the end of this year.

In situations where RR teachers moved, took another job, or administrators rearranged staff, UTC directors reported having little recourse for holding a district accountable to their three-year commitment to RR:

I just found out last week that one of the teachers who was trained last year moved to another school and took another position. I mean, it's one out of fourteen but now, with a larger number in Year 2, I'm not sure. [The potential for attrition] is an unknown factor...I'm really concerned about this. I mean, as exciting as it sounds that we have over eighty teachers participating in Year 2, I really don't know. Even though we have signed MOAs, in the end, we can't force them.

Over the first two years of the i3 grant, 230 teachers that were recruited to participate in the i3 scale-up failed to initiate the RR teacher training or moved out of their role as RR teachers - an attrition rate of 19 percent. CPRE will further explore issues of attrition amongst i3 schools with future data collection.

Support for Recruitment

With few exceptions, UTC directors and teacher leaders alike reported that recruiting schools and teachers for the i3 scale-up has been demanding and time-consuming. Many in each role spoke of the significant impact on their personal workloads, and expressed appreciation for the support they received in some areas, while lamenting its absence in others.

Support to UTC Directors

UTC directors expressed unanimous enthusiasm for the support they received for recruitment from the i3 office at OSU. In particular, they expressed gratitude to the project director and fiscal administrator for what they described as "amazing" responsiveness, communication, and problem-solving support. One director reflected:

I don't feel as though it's been that difficult, considering I only have a million little questions, really. It's just knowing that there's somebody that will get back to you quickly about things, that's on top of stuff, that is picking up errors which I've made, in terms of how the schools are being labeled or something. That's very helpful to know that there's another safety net.

Other directors concurred. "When there is a question, we're able to get an answer, and certainly there's been an ear to listen, which has been helpful, because there are all kinds of questions," said one UTC director. "So [the i3 office staff has] been very responsive and their responses have been timely so that we can move forward."

In addition to the support provided for their recruitment efforts, UTC directors spoke highly of the marketing materials developed by RRCNA to support outreach: "The i3 grant website - they try and keep a lot of information there including letters or brochures or whatever," explained one director. "Those products, so to speak, are helpful to see and either replicate or work from."

Along with the structured technical assistance and resources provided through the i3 grant office at OSU, UTC directors found meaningful support for their recruitment efforts in their network of fellow university trainers. While formal gatherings (e.g., the annual meeting of the NATG) provided opportunities for UTC directors to support one another, many found at least as much value in less formal contacts:

Well, we just network. We talk, you know, "Did you do a round of letters?" "Yeah, I did a round of letters." "Are you gonna do another round?" "Yes, I think I'm gonna do another round." "Well, if you're gonna do it I'm gonna do it." I mean, we have that... it's not very formal, it's just very informal and we're friends, so it's like... It's a close-knit group. It's not a business conversation.

Gaps in the Support Network

UTC directors consistently reported frustration and insufficient support for several other issues. Specifically, many described confusion surrounding the i3 recruitment guidelines, and initial questions about which teachers were eligible for i3-funded training. Others mentioned concerns over the evaluation design and its recruitment implications. For instance, some principals felt that randomly assigning students to receive Reading Recovery in a subset of schools—as is required by the RCT—was counter to the design of the intervention. Most UTC directors reported, however, that these issues had been largely resolved within the first few months of the scale-up.

Some UTC directors encountered additional frustrations as a result of less-than-supportive environments within their own universities. Though some reported finding great support and excitement about the i3 grant in their home institution, other directors described challenges ranging from copious paperwork and administrative hurdles, to a lack of support from university staff. For instance, in response to a request for assistance with recruitment, one trainer recalled being told "You're on your own" by a university official. "I won't go to him for help," she added.

Staffing Shortages

Staffing problems were among the most consistent themes across all regions with respect to recruitment. A majority of UTC directors reported having too few trainers and, especially, teacher

leaders to share the considerable recruitment workload. One director lamented her failure to anticipate her increased staffing needs at the outset of the grant period:

I definitely need another trainer. I mean it's, truly, it is a ten- to twelve-hour day for me nearly, six days a week. Keeping up with the i3 stuff is massive for us ... it's just, it's a lot of work. You know, we put this i3 grant together so fast, so quickly, we had this short window of time. So when you talk about hindsight, I mean, I can look back and think about what I didn't put in the budget that I know now I should have and I can't. So it's like, oh my gosh, I should have put another staff person in there just to help us deal with it!

As described in the section detailing recruitment challenges, capacity issues have affected most regions to some extent.

Support to Teacher Leaders

Like UTC directors, teacher leaders reported a range of experiences with regard to support for their recruitment efforts. A majority of teacher leaders described a support system of one kind or another, but the primary source of support varied from one teacher leader to the next. Some found support within the Reading Recovery community, particularly from university trainers or other teacher leaders. Others reported strong supportive relationships with site coordinators or other school- or district-level administrators. One teacher leader explained:

When we have a site coordinator that's very knowledgeable about Reading Recovery and funding sources and understands the state and federal government, then they can serve as a liaison for us so we can do the work that we need to do, which is working with teachers and children.

Since UTC directors serve as regional liaisons with the i3 office, teacher leaders generally did not experience support from the i3 staff directly, or recognize the office as a source of support for recruitment efforts in particular. Instead, nearly all teacher leaders received some degree of support from their UTC directors around the logistics of i3 recruitment, including issues like the development of MOAs for participating schools, and the processes for distributing i3 funds.

A UTC director described the changing nature of her interactions with teacher leaders since the start of i3 scale-up this way:

My email correspondence with those that are this year training i3 teachers has increased quite a bit. Yeah, there's a lot of questions they have that are coming in now about "Where are the books? I didn't get my books, or didn't get the ITBS set, or I'm missing this, I'm missing that. When can I expect to get this kind of stuff?"... Now we have a tuition waiver instead of [our previous system], so we have to change a bit in how we register people for the courses. You know, things like that. So there's absolutely increased communication around troubleshooting, mostly.

For some teacher leaders, the scale-up brought significant shifts not only in the type of support received from UTC, but also the volume. "I saw nothing from the university before the i3," said one teacher leader. "As far as recruitment went it was all on the teacher leaders. But then when the i3 came around I saw the university take a huge role, and I think it's because their name was on the line: 'We're invested, we're going to do this.'"

For other teacher leaders, the support from their UTC left something to be desired:

I don't think there's a plan. I think the last time we had a teacher leader PD (professional development) it was pretty much put on our plate that it's our responsibility to get these schools. I mean and rightly so... we know the school districts. And then if we need help to talk to the training center. But as [far as] a plan, I wish there was one for this whole State. I mean I wish there was more of a push of getting this out to public knowledge, you know you all did a great job of getting all that out to the different districts. I wish we would do something like that.

In some regions, teacher leaders turned to one another for support around recruitment. This was particularly true in regions where the teacher leaders bore the primary responsibility for recruitment, where they tended to work across multiple districts, or where they were widespread geographically. In these cases, moral support or idea-sharing sometimes gave way to the redistribution of teacher caseloads. One teacher leader reported: "Just because of sheer geography [another TL] called me and said, 'There is this new system that's interested in training next year so will you take them and support them?' So that's how I gained the, the new system that I have."

Future Plans for Recruitment

Although all UTC directors expressed a commitment to recruitment, many also expressed concern at the prospect of accruing schools at the expense of high-quality, long-term implementations:

It's one thing to recruit teachers and have them trained because the training is free and all the expenses are covered in Year 1, but a lot of schools are asking what will happen beyond Year 1. I think this is one issue, I mean, getting schools to commit to the long-term plan. Not just getting people trained for one year, you know, not having any sense of commitment to longer-term implementation of Reading Recovery. And frankly, I'm concerned about this. I think we're all concerned about this. We don't know how this is going to play out.

As a consequence of this concern, directors reported their intentions to support sustainable implementations by focusing on building capacity and communication. Many emphasized the

benefits that their regional systems would experience beginning in Year Three of the scale-up as a result of the infusion of additional university and teacher leader personnel. “We’ve had meetings with site coordinators, offered the support and all that sort of thing, but we haven’t done as much as we will be able to do next year with two trainers,” explained one UTC director. Another excitedly detailed the future impact of gaining several new teacher leaders:

I’ve got these two new teacher leaders I’m training. That will expand our capacity, and if I can get another outside grant, then I’ll try [to train] another person, or two, next year. That’s all we’re limited by – truly – the workload of the teacher leaders and myself; how many people can we get to. The 24 [RR teachers in training] have everyone stretched to the limit this year.

Several other UTC directors also overviewed their plans to secure the resources that would allow them to train more teacher leaders:

The big goal – and this is what we want to think about for recruitment next year – is to think about where we need to establish a couple of hubs, a couple of training sites. Putting them in the areas that are dark right now. We are just thinking about supporting [rural areas] that way. It seems like they need that sort of structure. [They need] that training site.

Although much less frequently noted than plans to incorporate and train new staff, a few directors shared plans to increase communication with teacher leaders and potential sites in order to foster a deeper understanding of the RR model and ensure stronger implementations.

Talking more with my teacher leader groups—trying to really impress upon them how important it is that they council schools and districts that come to them wanting to implement Reading Recovery, just Reading Recovery in general, regardless of i3... I’m telling the teacher leaders and the sites that are saying, “Yay, rah, we have all these teachers to train, more money for us to operate our center and all of that kind of stuff.” Wait a minute! You still have to have an implementation that’s good, with fidelity. So now we’re going to back up a bit and really dig in and say, “Let’s really ask the districts some questions. Let’s find out what their need is for Reading Recovery.”

The emphasis on strong implementation and fidelity by many UTC directors was couched in a willingness to compromise. In light of limitations imposed by an increasingly constrained teacher leader workforce, large geographic areas, and the need to attract and train as many new teachers as possible, directors reported their intentions to deviate from practices codified in the Reading Recovery Standards and Guidelines via a formal waiver process. Most frequently, they described deviating from the standards pertaining to class-size requirements for RR teachers in training – which can be an impediment to establishing training classes – or the transportation

of children for RR teacher training. "We need to address this issue of small classes," explained one UTC director. "We have a mission to get this going because when the marketing material comes out, we want to be able to say 'Yes, we can train you and this is how we can do it.'" Another UTC director described her idea for addressing the transportation of children to training in more remote areas:

I think we're going to have to rely more on technology in the future because in our states for teachers to bring their kids [to training] it's a two hour drive on snowy roads. It's a big issue. A two hour or three hour class and then another two hour drive home. So those kids are on the road for seven hours...We've been looking into things like Tanberg, but really good systems. Technology has come a long way, and it's improving almost daily. However the Standards and Guidelines require a face to face. So we're constantly juggling that.

As the effort to bring the Reading Recovery program to scale is just beginning, UTC directors' plans for Year Three of the i3 grant primarily centered on strengthening and building upon existing regional recruitment efforts. At the core of most directors' prospective strategies was the plan to increase contact, specifically more personal contacts, with targeted schools and districts:

Once again, we will be working together, putting a grant letter together, but also following up with a phone call is really important. Even if they don't talk to us, a good solid e-mail and trying to follow up. We'll be speaking to their curriculum directors. That will be helpful too. Planting the seed, I think, is the most important thing, and I'm going to focus on districts that have real concerns.

Those directors who had previously incorporated personal contact as part of their recruitment strategy reported their intentions to continue leveraging established relationships, as well as their professional networks. Additionally, they reported intentions to explore more creative venues for interacting with potential RR implementers:

We have a huge number of Reading Recovery professionals who've become principals, superintendents and everything in between. I'm compiling a list of people, who themselves were trained in Reading Recovery, who do not now have it. I'm going to draw upon their positive, strong memories of Reading Recovery, and that's going to be a new recruitment effort. We're going to host a reception on campus to recognize existing Reading Recovery professionals in leadership roles...and invite them to celebrate the newest successes of Reading Recovery.

Conclusion: Recruitment

Two years after the allocation of the i3 award, UTC directors are excited about the opportunity to expand the reach of Reading Recovery, and they are generally optimistic about their regional system's ability to meet i3's recruitment benchmarks. Drawing upon their historical experience with the expansion process, the RR network has remained committed to regional autonomy. This commitment has allowed stakeholders across each UTC region to formulate, manage, and carry out a comprehensive recruitment approach that honors directors' and teacher leaders' insider perspective on the strategy best suited to scaling-up Reading Recovery in each context.

While Year 1 presented challenges to recruitment that were, for the most part, related to the requirements of the i3 grant, more persistent, ongoing obstacles also exist. While many regions have successfully met and even exceeded their recruitment benchmarks, others have faced significant barriers to recruitment, such as manpower shortages, a difficult fiscal climate, and even resistance to the philosophy of literacy instruction underpinning Reading Recovery.

Although UTC directors praised the support of national Reading Recovery stakeholders such as the i3 office and RRCNA, recruitment efforts drew almost entirely upon internal, regional resources. As the RR network's overall ability to meet i3 recruitment benchmarks is contingent upon the success of individual UTC regions, historical predilections towards autonomy may need to be weighed against the broader goals of the nationwide scale-up.

Early Implementation Findings: Training

The i3 scale-up grant was intended, in part, to fund training of both RR teacher leaders and teachers. Because very few teacher leaders were trained with i3 funds in 2011-12, however, CPRE did not focus on teacher leader training in the first year of the evaluation. The findings about training presented in this report pertain only to the training of RR teachers.

Although Reading Recovery's model assumes a close connection between the training of RR teachers and the implementation of the intervention at the school site, for purposes of this report we have separated the findings related to these two components of the program. This section provides evidence of RR teacher training experiences from interviews conducted at the beginning of the 2011-12 school year, teacher leader focus groups, and RR teacher and teacher leader surveys administered near the end of the school year. It focuses on RR teacher training experience; continuing contact for trained RR teachers; RR teachers' perceptions of their preparedness; and the broader implications of the RR training experience. The following overview pertains to only those teachers whose training was funded through the i3 grant.

Key findings regarding the training of RR teachers participating in the i3 scale-up include:

- » RR teachers found behind-the-glass sessions to be one of the most valuable aspects of their training experience.
- » RR teachers appreciated the immediate feedback and support provided by teacher leaders through one-on-one interactions during school visits. They felt this feedback helped them make concrete changes to their teaching.
- » RR teachers felt that their training experience prepared them well for RR implementation and working with RR students.
- » RR teachers viewed teacher leaders as important and accessible sources of information about literacy instruction, and as supports for their own professional growth.

- » RR teachers reported that RR training was transformative in terms of their own understanding of literacy. Their new skills in observation, assessment, and instruction were useful in both their RR and non-RR instructional roles.

RR teacher Training: Year 1

Training for RR teachers took place at designated training facilities – RTCs – and at the schools where the teachers are assigned. During their training, RR teachers are expected to learn to design and deliver individual daily lessons; to document lessons; and to collect and utilize various types of student progress data. This learning requires intense effort from teachers, and is supported several ways; first, all new RR teachers-in-training complete a week-long summer introduction focused on the administration, scoring, and interpretation of the Observation Survey of Early Literacy Achievement (OS). Second, they complete a year-long academic course taught by a teacher leader, for which they receive a grade and 8 to 10 university credits. This coursework uses as its primary text Marie Clay's *Literacy Lessons Designed for Individuals, Parts I and II* (2005), and incorporates regular opportunities to observe instruction through behind-the-glass lessons. As they complete this course, RR teachers-in-training provide one-on-one daily lessons to four RR students; they also attend weekly three-hour training sessions conducted by their teacher leader. Finally, both trained RR teachers and those in training receive school-based visits from their teacher leader, who observes them in Reading Recovery lessons and provides immediate feedback and suggestions.

The Behind-the-Glass Session: A Valuable Teaching Opportunity

RR teachers reported that behind-the-glass sessions played an important role in their training. During training classes, RR teachers observe full RR lessons taught by other teachers-in-training, and are also required to be observed by their peers and their teacher leader. These sessions provide opportunities for RR teachers-in-training to both give and receive feedback on RR instruction. The RR Standards and Guidelines require that each RR teacher-in-training observe approximately 30-36 behind-the-glass sessions taught by other teachers-in-training, and deliver a lesson behind the glass themselves at least three times a year.

Survey responses indicated variability in the number of behind-the-glass sessions RR teachers-in-training observed: 19 percent reported observing 10 or fewer behind-the-glass sessions; 26 percent observed between 11-20 sessions; 28 percent observed between 21-30 sessions; and approximately 25 percent observed more than 30 behind-the-glass sessions. This suggests that a majority of the RR teachers training in 2011-12 fell short of the required 30-36 observations. However, a majority of RR teachers in training during the scale-up were successfully meeting the minimum requirement for lessons taught behind the glass; survey responses indicate that 87 percent taught behind the glass three or more times. Thirteen percent of RR teachers-in-training taught behind the glass only once or twice.

The variability in first-year teachers' behind-the-glass observations may be due to the size of the training classes; smaller classes afford fewer opportunities to observe lessons. Another factor may be that some RR teachers were trained through a new initiative being piloted by several UTCs - a hybrid training class that alternates online and in-person class sessions. Meeting only on alternate weeks may have affected the ability of the RR teachers involved with this pilot to participate in behind-the-glass sessions. In future years, CPRE will monitor participation in long-distance training classes.

Behind-the-glass sessions were perceived as a particularly valuable means of training for work with students, for both the RR teacher behind the glass and those teachers observing. Teachers who go behind the glass receive immediate feedback and suggestions on an actual lesson with a student, and are able to make adjustments to their lessons the very next day. One RR teacher explained how the behind-the-glass sessions are different from more typical teacher observations:

When you're observed, it's not a contrived, really perfect lesson, but it's actually what you do every day, and you're getting really honest feedback on the things that others see that are working well for the student you were working with. And there are things you may want to try to continue to accelerate their progress... But they're really positive experiences. And I think those conversations that we had as colleagues just were very in-depth in regards to what the students have as far as skills and what they need to accomplish. So I appreciated it very much.

For those teachers who are observing, the behind-the-glass session offers an opportunity to provide constructive feedback to the teacher they observe, while learning new skills or strategies they can use themselves. One RR teacher described how observing a behind-the-glass session helped give her new ideas:

Sometimes you do your Reading Recovery lesson, you always assume you have to be perfect. So, when you watch someone doing [something different], you think "Oh, I will try this. Oh, what they're trying is great! I should try it in my next lesson."

Some RR teachers-in-training suggested they would especially benefit from opportunities to observe more seasoned RR teachers or teacher leaders behind the glass:

You know we watch each other behind the glass and that's very helpful, but all of us are in training. It might be kind of nice to watch more experienced teachers behind the glass do lessons.

Teacher Leader Supports for RR Teachers

Survey responses by teacher leaders indicated considerable variability in the number of RR teachers they supported and/or monitored during the 2011-2012 school year. The RR Standards and Guidelines note that operating sites should “limit the number of teachers supported and monitored by the teacher leader to 42, or considerably fewer, depending on factors such as distance, the number of teachers per school, and the number of districts” (2.24). Survey data indicated that 46 percent of teacher leaders supported 1-20 RR teachers, and 28 percent supported 21-42 RR teachers. Nine percent of teacher leaders exceeded the recommended caseload; seven percent supported 43-59 RR teachers, and two percent supported over 60 RR teachers. It is important to note that these survey responses include the total number of RR teachers - those in training as well as those previously trained - being supported by a teacher leader.

The number of RR teachers they support, combined with the distance a teacher leader needs to travel to visit each teacher, influences how much one-on-one time a teacher leader can spend with each teacher and still meet her other RR responsibilities. “I singlehandedly support a site of 50 teachers in a region of 3,000 square miles,” explained one teacher leader. When schedules were tight and the travel distance was long, teacher leaders sometimes had to adjust the amount of time they spent with the different RR teachers at their school sites. One RR teacher explained that her teacher leader “probably visits some of the other [RR teachers] a little bit more because I am a little bit farther away. About an hour and 15 minute drive. I’d say she’s been here at least six times.” When visits must be limited, teacher leaders are not able to give their RR teachers the same amount of in-person feedback.

Despite some variation in availability of teacher leaders and support provided by them, RR teachers were overall very positive about the role their teacher leader played in helping them become comfortable with RR instruction. One RR teacher summed up her experience working with her teacher leader this way:

My teacher leader is extremely supportive. She’s always available by phone or by email. She is there often for site visits, and, I just feel that anything that I need help with there’s always someone to answer and assist me. I feel that they’re really guiding me through this year and helping me with what I need to do.

Many RR teachers described their teacher leader as an instructional coach who helped develop their individual skills by modeling lessons and providing immediate feedback following observations – two types of support that allowed teachers to make concrete improvements to their teaching. Another RR teacher commented:

[My teacher leader] is very personable. She's very positive. She's a lot like a coach, where she kind of coaches us along. Sometime the Reading Recovery itself is all about having a constructive approach. Which means, you kind of let the learner come to the learning themselves. You kind of just lead them there. And so with that in mind, that's kind of how they approach it with us as well. They want us to discover these things and, you know, lead ourselves into understanding.

Benefits of Training Cohorts

RR teacher training occurred in a cohort model, with groups of RR teachers training together at the same pace. Many RR teachers indicated in interviews that they felt a sense of community and camaraderie develop among their colleagues during the training process. One RR teacher explained that "just being in a community of people who are all struggling with the same issues, and having someone to guide us around figuring them out, has been great."

At school sites where multiple RR teachers worked and trained together, RR teachers remarked on the benefit of having opportunities to learn from frequent demonstrations and observations. One RR teacher commented: "I'm liking the fact that my teaching partner and I are in the same classroom so we hear each other and that informs our teaching and we ask each other questions." Overall, RR teachers felt supported by one another, and drew on each other's experiences interacting with students.

This experience was not limited to first year teachers; trained RR teachers also expressed the benefits of working with their colleagues. By exchanging feedback, RR teachers at the same school were able to address their questions and concerns frequently. Some RR teachers noted that having a supportive network within their school facilitated instruction and collaboration, because they were able to learn from one another and provide resources or support as needed. Teacher leaders reiterated this idea that having a RR community at the school can be helpful. One teacher leader explained:

The one school that I had my teacher at this year, she was very lucky in that she had two ideal RR teachers with her. One used to be a teacher leader... I still made my coaching visits but it was just nice to have three teachers in a room. For a new teacher-in-training, it was very effective to have that support for her and for me.

Challenges to Training

Distance and Class-size Requirements

RR training entailed a considerable amount of time outside of regular work hours for both RR teachers and teacher leaders. Although the timing of the weekly training varied, most sessions

were scheduled in the evening. For most teacher leaders and RR teachers, the training site was located nearby; however, some participants reported having to travel several hours in order to complete their training requirements. One RR teacher indicated that her training site was four hours away, and her teacher leader also drives that distance. Another RR teacher explained, "My training is actually wonderful. The only thing I don't like is actually having to drive over an hour or two to my classes." She also mentioned that the length of the drive deterred other teachers from taking part in Reading Recovery.

Some UTCs have attempted to address the challenge of distance in different ways. As noted, a small number of UTCs are piloting a hybrid model of training that involves real-time instruction provided by a teacher leader using teleconferencing technology. Other UTCs are making allowances for in-person training classes to be held every other week, with a distance learning component (e.g., an online lecture/webinar or additional reading or assignments) for the 'off' weeks. This does ease the burden of travel for many teachers and teacher leaders; however, because course content is often directly related to next steps in the RR classroom, the two-week gap can be difficult for RR teachers-in-training who must wait until the next in-person training class to address problems and make adjustments to their practices. One RR teacher explained:

Because I'm so far away from campus, we've been meeting every other week, which I appreciate gas-wise. But I also see it as a disadvantage because sometimes I was learning information after the fact, which has its pro and cons.

The teacher continued to explain that one of the benefits of meeting every other week was that she could apply what she learned during a meeting to the lesson that week. A challenge to meeting every other week was that often she had no guidance during those weeks she was not meeting. Those RR teachers who had weekly classes often spoke about the value of being able to immediately apply what they learned in their training class to their RR lessons the same week, which allowed them to practice their new skills before the next week's training session. One RR teacher explained: "We'll learn something or study an article and then we'll try to implement something within our teaching strategies for that upcoming week." She, and other RR teachers, commented that the opportunity to learn new content and apply it to the RR lessons each week was important as they tried to hone their skills while continuously working with students.

Requirements for training class sizes can be a significant challenge to RR teacher training under certain circumstances. First, class sizes must be large enough to allow RR teachers-in-training to observe the 30-36 behind-the-glass lessons they are required to observe during their training year. Additionally, some of the universities where RR teachers-in-training take courses for credit have requirements regarding minimum student enrollment. One teacher leader mentioned that she anticipated being unable to have a training class in 2013 because, despite the scale-up, she was not able to recruit six potential RR teachers (the minimum number required by her university). In order to meet class-size requirements, some teacher leaders reported combining

new teacher from multiple sites and districts. While this may help address the problem of undersized classes, it could compound the problem of distance for teachers who must travel to attend each class.

The Challenges and Benefits of Pacing

Nearly 70 percent of all RR teachers started working with RR students within two weeks of the first day of school. Many expressed that their first few months implementing RR were particularly challenging since they were simultaneously teaching RR students while learning RR instructional methods. The pace of training meant that sometimes teachers encountered instructional tasks or issues that had not yet been covered in class. One teacher explained that each component of Reading Recovery was gradually introduced to trainees; she learned about the Observation Survey first, and later how to take a running record. Though most of the time the pacing worked, sometimes the training objectives did not necessarily correspond with her students' progress:

You now are adding this part and now are adding this part. So it started off we were just reading with them. Then we were reading and writing. Then we were reading off the books, writing, and doing a running record. You know each couple weeks we'd add a new component, and I think for me, I needed to see all that as a whole before I jumped into it. So if we were able to maybe start the class in August versus September, go all through August to learn all that stuff, to see what it really looks like so that when I get into it, I really have an idea of what it should be... That's the kind of learner I am though.

It was not until the teacher saw a behind-the-glass session much later in her training that she was able to put everything together and see how the full program is designed to be implemented:

It was very hard for me until I saw people teaching behind the glass, which obviously didn't start right away. When I saw that then it all pulled together. I was like, "Oh ok, I get it now." But I never had that model. To see what a full lesson would have looked like. And it's different reading it in a book versus really seeing that kind of interaction.

While most RR teachers said the simultaneous learning and applying of RR strategy felt like "trial by fire," they still conceded it was the best way to cement their understanding of the program. As one teacher commented:

I definitely think, because you are a practicing RR teacher during the training year, the first couple of months are a little rough, because you're trying to learn the process as you are simultaneously teaching. However I don't think there's another way to do it because you have to do it while learning it in order for things to cement and for you to learn them.

RR Teacher Perceptions of Preparation

Overall, RR teachers described their experience with, and perception of, the training they received in positive terms, and many felt the training was of high quality. A vast majority of RR teachers reported feeling very well prepared for their RR responsibilities, such as planning and conducting lessons, and reporting data. One RR teacher explained:

I feel really well prepared. This course is really intense and I would say I learned more in this one class than I have in my entire master's program to become a Reading Specialist... I feel like I've learned more about the reading and writing process and what kids might be doing and what they might not be doing and where they might be getting stuck and how I can push their learning forward.

This sentiment was reiterated throughout much of the interview data, and substantiated through the RR teacher survey data highlighted in Table 13. These data show that RR teachers' feeling of preparedness after training varied by task. For example, the RR teachers surveyed felt most prepared to administer the Observation Survey at the beginning of the year; virtually all (99 percent) of RR teachers reported feeling either adequately, well, or very well prepared to complete this task. Overall, a majority of RR teachers also felt that their first-year training prepared them to conduct the tasks that are related to conducting a lesson. The RR teachers surveyed felt least prepared to develop students' comprehension; just less than 91 percent reported feeling adequately, well, or very well prepared to do so.

Table 13: Reported Levels of Preparation to Conduct Various Tasks Expected of RR Teachers (N=747)

	Less than Adequately Prepared	Adequately Prepared	Well Prepared	Very Well Prepared
Administering the Observation Survey	1.2	6.3	32.7	59.8
Knowing when to transition students to the next reading level	3.5	24.2	46.7	25.6
Successfully discontinuing students	4.7	28.7	42.6	24.1
Choosing texts that accelerate student progress	4.8	26.1	44.9	24.2
Developing students' fluency	4.5	26.9	43.5	25.0
Developing students' comprehension	9.2	30.4	39.5	20.9
Improving students' independent reading skills	4	22.1	46.6	27.3
Improving students' writing skills	3.5	21.7	48.7	26.1
Roaming Around the Known	6.2	25.0	40.4	28.4

Source: Surveys administered to RR teachers.

Though not an explicit focus of RR teacher training, CPRE investigated the extent to which teachers felt prepared to coordinate and collaborate with district and school stakeholders. Though the overall percentages of RR teachers who reported feeling prepared are fairly high, Table 14 shows distinct variation in preparedness to coordinate and collaborate across different groups.

Table 14: Reported Levels of Preparation of First-year RR Teachers to Coordinate and Collaborate with Reading Recovery Stakeholders (N=747)

	Less Than Adequately Prepared	Adequately Prepared	Well Prepared	Very Well Prepared
District leaders	23.6	29.9	32.7	13.9
Classroom teacher	10.9	24.2	41.2	23.7
School leaders	12.4	29.5	39.2	18.9
Students' parents	10.0	24.6	41.8	23.6
School leaders	12.4	29.5	39.2	18.9

Source: Surveys administered to RR teachers.

Feelings of under-preparedness were higher for developing collaborations with school leaders (12 percent) as compared to parents (10 percent) and classroom teachers (11 percent). Twenty-four percent of RR teachers did not feel prepared, or felt minimally prepared, to coordinate and collaborate with district leaders with respect to Reading Recovery.

The ability to coordinate and collaborate with other stakeholders is an important skill for RR teachers to develop as they are expected to work with other teachers and school staff to implement RR and monitor student progress. Nearly 60 percent of RR teachers indicated that they were part of a group or team that monitored the progress of Reading Recovery in the school; feeling less prepared to work with other RR stakeholders may have an effect on those RR teachers and the effectiveness of their monitoring team. Also, if RR teachers are reticent to work with others, their reticence may limit buy-in from people otherwise interested in the success of Reading Recovery. Issues of collaboration are addressed in the next chapter, and will continue to be explored through future data collection.

RR Teacher Growth in the First Year

Survey responses indicated that, over the course of their training year, RR teachers felt they gained confidence and developed skills through their work with students. Many of the RR teachers reported having very different experiences with their first and second cohort of students. "Now that I'm working with my second-round students," commented one teacher, "I feel like I'm doing a better job with them than I did with my first round just because with the first round you're trying to learn so much." Another RR teacher saw the second cohort of students as an opportunity to apply all of the knowledge she gained by working with the first:

So now just looking back, I'm going to do things differently this time. After being in training since September, I know what works better. Now I get a second chance with this other group of kids.

RR teachers reported that, with their second group of students, they had a better sense of what the overall program and individual lessons should entail. Some RR teachers even expressed an interest in keeping their first cohort of students for longer than the allotted time, maybe even the whole year, to compensate for the amount of learning they had to do at the beginning of their training. This way the first cohort of students is not "short-changed" by a teacher's developing skill. Another RR teacher suggested changing the structure of the training year so that a teacher sees only one or two students at the beginning of the year, then eases into a full load of students once the teacher is more comfortable.

Continuing Contact for Trained RR Teachers

A theoretical underpinning of the RR model is that teachers should be lifelong learners; as such, RR teacher training is an ongoing and extensive process. RR teachers who successfully complete their training year enter a phase of training called "continuing contact," during which they attend professional development, maintain communication with their teacher leader, and occasionally receive one-on-one visits from their teacher leader. In interviews, teachers who had successfully completed their first year reported that the continuing contact aspect of their training was crucial for maintaining their skills going forward.

Behind-the-Glass for Trained Teachers

As part of their continuing contact, trained RR teachers are expected to attend "at least six continuing professional development sessions each year, including a minimum of four behind-the-glass sessions with two lessons per session" (Standards and Guidelines of Reading Recovery in the United States, 5th Edition, p.15). Seventy-six percent of teacher leaders surveyed reported that they provide six to eight continuing contact hours for trained teachers per year; 78 percent reported that the continuing contact sessions they provided included at least four behind-the-glass sessions, with at least two lessons per session. Six percent of teacher leaders surveyed

reported only overseeing four behind-the-glass sessions, and three percent reported that the continuing contact sessions did not include a behind-the-glass component.

As discussed in the previous section, RR teachers-in-training noted that behind-the-glass sessions were valuable experiences during their first year; a similar sentiment was reported by trained RR teachers. Ninety-two percent of trained teachers reported teaching behind the glass once or twice; eight percent reported that they did not teach any behind-the-glass sessions.

Teacher Leader Support for Trained Teachers

Trained RR teachers are visited by their teacher leader less frequently than RR teachers-in-training. The Standards and Guidelines state that trained teachers are to “receive at least one school visit from a teacher leader annually” (3.42). According to the survey of teacher leaders, 36 percent of those working with trained RR teachers reported visiting them once in 2011-12; 58 percent visited their trained RR teachers two to four times, and six percent visited them more than six times. Many RR teachers in continuing contact recalled how valuable immediate feedback was for their development, and noted that they do not receive it as often as when they were in training. “I wish that we could keep getting the same kind of training,” commented one trained RR teacher. “And even though we still have continuing contact, it’s not intense, so you do not have that constant feedback that you get when you’re in the training.”

Though they wish it was more frequent, trained RR teachers did speak positively about their continuing contact professional development, commenting that it further developed their understanding of Reading Recovery. One teacher remarked:

I just feel like I’ve been able to take what I learned last year, which I felt I did a very good job last year. I’m very proud of how it went. But with the continuing contact this year, it takes it to a whole new level of deeper understanding.

Broader Implications of Training Experience

The RR teachers interviewed were overwhelmingly positive about their training experience, and many also expressed enthusiasm for the way their RR-specific training affected their understanding and instruction of literacy. “What I’ve have learned so far,” explained a RR teacher, “will change my teaching forever.”

A New Perspective on Literacy

Many RR teachers were surprised by how much knowledge they gained during their training, even those with prior experience or specialized training in literacy, reading, or special education. These teachers had expected “to know most of it,” as one respondent noted. However, many found that the training provided a fresh perspective on teaching reading skills. One teacher commented that “[the training] was a complete shift in thinking,” that led her to reconceptualize

the way children learn to read. RR teachers talked about their new perspective on literacy differently; for some, the training offered them tools to help “hone in” on exactly where students required assistance:

I think the big thing is the time to sit back and observe children and really closely observe and take notes. And then take what you see happening and then decide, OK, what can I do specifically to take that child from what they know and help remediate that specific thing...and then just the really honing in on exactly where the confusion is.

Other teachers spoke about how their RR training helped them develop a deeper understanding of the reading process:

I have been doing this, working with literacy and struggling readers for four years, five years. But going through this training I just think it really made me even better, the way that I can speak to teachers, and really explain what’s going on and how things need to shift or move in their classroom. It just really gave me a much deeper understanding.

The ability to change the way RR teachers think about teaching and learning appears to be one of the most powerful elements of RR training.

Translatable Skills

One of the reported benefits of RR training was how much of what RR teachers learned could be applied to other literacy-focused teaching activities. Nearly 72 percent of all RR teachers who responded to the survey reported that implementing RR improved their general literacy instruction to a great extent (i.e., the top category on a four-point Likert scale), and many reported feeling reassured about their skills as reading teachers overall. RR teachers who were interviewed noted that the use of RR strategies and language particularly enhanced their literacy instruction in small group settings:

I found that it had really enhanced my small group instruction even though I had a reading endorsement and I was in the classroom for eight years, I found that is helped me go deeper as a reading teacher.

Another RR teacher mentioned how she was able to apply her RR training elsewhere:

I’m using some of the same [Reading Recovery] strategies with my 2nd-3rd graders. Just some of the ways of getting the kids to hone in with the language and the comprehension. Those strategies that I’m learning there I’m also implementing elsewhere.

A few RR teachers indicated that they would have appreciated more formal training focused on how to apply RR strategies to non-RR aspects of their work. One teacher explained, "I would love some training on maybe how to generalize it more and use it with small groups, but I realize it wasn't designed for that purpose." When RR teachers are able to apply their training across other literacy activities, the number of students impacted by Reading Recovery can move beyond the students receiving one-on-one support.

Conclusion: Training

In general, RR teachers trained with i3 funds reported positive training experiences. Most felt that the training provided them with the knowledge and skills they needed to conduct RR lessons with students. An important facet of the training was the continuous improvement that took place as a result of behind-the-glass sessions and teacher leader observations; both experiences provided opportunities for RR teachers to receive immediate feedback, and allowed teachers to self-correct throughout the year. Teacher leaders were considered a vital part of RR teacher training. Even with limited time and resources, teacher leaders had a significant positive impact on RR teachers.

Early Implementation Findings: School-Level Implementation

CPRE conducted extensive research into the implementation of Reading Recovery in the schools participating in the i3 scale-up during the 2011-2012 school year. This research yielded insights about the activities and processes underlying implementation of the intervention in context. It also helped to clarify what questions require further exploration if the relationship between the implementation and impact of Reading Recovery is to be fully understood.

Although the heart of Reading Recovery is a one-on-one intervention involving only a teacher and a small number of individual students, CPRE has observed that the “instructional core” (Elmore, 2005; 2007; 2009) of the intervention is supported by a system of people and processes that span schools and districts. Indeed, this view of Reading Recovery as a system-level intervention is elaborated in Clay’s writings (Clay, 1987). CPRE’s discussion of the implementation of Reading Recovery therefore focuses on two overlapping and, to varying degrees, integrated tiers. The first of these comprises the roles and relationships that establish and define Reading Recovery as a contextual feature of a school. The second tier includes the processes and activities that constitute the intervention’s instructional core. The following discussion will address both what CPRE found, and what it has yet to learn, about each of these tiers.

Key findings regarding the implementation of RR in schools participating in the i3 scale-up include:

- » Most RR teachers provided one-on-one lessons to RR students for half the day and fulfilled other roles in their schools during the other half.
- » RR teachers encountered obstacles to adhering to the schedule of daily one-on-one lessons with each student. However, most found ways to surmount these obstacles.
- » RR teachers communicated frequently with 1st-grade teachers, mostly about RR students’ progress. Some shared instructional strategies with their colleagues, but a minority had explicit instructional leadership roles.

- » Many RR teachers participated in their schools' Response to Intervention processes. The sophistication of these processes and the extent to which RR data was utilized varied.
- » Teacher leaders were reported to be critical supports for RR teachers, though their roles and responsibilities varied widely from site to site. Some had significant influence over decisions affecting RR implementation at the school level, or even over literacy instruction generally. Others had very little influence.
- » CPRE observed inconsistencies in the ways schools and districts selected students to participate in RR. However, there was strong consistency in the execution of RR lessons.
- » There was variation in the extent to which students completing RR received ongoing supportive services at their schools. A student's participation in RR, and the data generated in the process, were given weight in decisions about special education.
- » RR teachers, 1st-grade teachers, and parents were enthusiastic about the intervention's impacts on student achievement, motivation, independence, and confidence.
- » There was considerable variability across schools in the degree of alignment between RR and the school's existing literacy curricula and interventions.

School-Level Implementation: Findings from Case Studies

In some ways the nine cases selected from nine UTC regions were remarkably similar – all were following the Standards & Guidelines for student instruction with few modifications to the prescribed RR routines. When looking at the structures that underlie RR, however, the nine cases revealed a range of experiences in the way RR came to the district and school, the level of commitment to RR by different stakeholders, and the extent to which RR has been integrated into the fabric of the schools.

In some districts, the decision to adopt RR emerged from a top-down process, where the original advocate was a high-level administrator such as an Assistant Superintendent or Director of Literacy Curriculum. One consequence of having support from a district-level advocate was a slight alteration to the role of the teacher leaders in the district; with high-level support, they were often no longer solely responsible for recruiting schools. In other districts, interest in the program was built on a much more bottom-up process; in these cases the original advocate was often a teacher who had some familiarity with RR, and saw it as of potential benefit to the literacy program at their school or across the district. For some advocates, the primary draw of RR was its emphasis on data and research; for others it was attraction to RR's instructional philosophy. Districts also took different routes to pushing RR out to their schools. In some sites, district administrators appealed to principals first, generating enthusiasm and commitment at that level before agreeing to dedicate RR resources to a school. Securing principal buy-in for RR in a school seemed to make a positive impact on school-level understanding of RR and support of RR teachers.

In general, principals and general education teachers varied in their understanding and support of RR, and this variation was often associated with experience. For instance, newer teachers more often expressed openness to learning new vocabulary and techniques to use with RR students in their classrooms, and more willingness to release students for RR instruction. On the other hand, some more experienced teachers expressed their preference for their long-held routines, which did not easily accommodate RR. In other cases the variation in support for RR across general education teachers was due to level of comfort with the pull-out model of intervention, with some teachers preferring for students to remain in their classroom. A frequently expressed objection to RR, even among teachers who recognized the program's positive impact, was that RR serves too few students.

A clear theme across the cases is that these types of tensions between RR and general education teachers are best managed at the school level by the principal—rather than the teacher leader who works across schools—as the principal is most familiar with the school context. The case study schools showed different levels of principal involvement in managing the implementation of RR; some were outspoken advocates for RR who took great pains to protect RR teacher time and generate buy-in from classroom teachers. Other principals were less involved with the management of RR in the building; in these cases RR teachers sometimes expressed feeling less supported in their schools, and felt that general education teachers misunderstood (or even resented) their work. Such comments suggested a clear role for the principal in managing teacher understanding and expectations of RR that was not being fulfilled.

A distinct, if somewhat unexpected, benefit of RR revealed by the case studies was how it helped some schools and districts as they moved to streamline and create coherence in their literacy programs. Many RR teachers, principals, general education teachers, and administrators commented that the presence of RR revealed issues with their previous approach to literacy, which was often a patchwork of many interventions that all targeted different skills.

A major takeaway from the first nine cases was that the mechanics of RR instruction are quite uniform across sites, but the dynamics of school-level implementation vary widely. Overall, the cases indicated that RR should not be addressed solely as a set of practices with clear boundaries that can be inserted or removed from a building without consequence. Instead, there are distinct, significant organizational moves and cultural shifts associated with RR; some are results of RR implementation, others must be proactively created prior to, or concurrent with, implementation in order for RR to have the best chance of success. These data suggest that two important areas for future research are: a) the varying dynamics of, and structures in place to support, RR implementation in different contexts; and, b) the cultural aspects of legitimation and institutionalization of RR across sites. Future case study work will focus on these two primary issues.

School-level Implementation: Roles & Relationships

The implementation of Reading Recovery requires the cooperation and investment of many key players at both the school and district levels. In the first full year of the scale-up, CPRE examined each of these roles.

The RR teacher

As previously described, the instructional expectations of RR teachers are that they will “teach at least four 1st-grade students per day individually for 30-minute daily sessions in a school setting throughout the year” (Standards and Guidelines of Reading Recovery in the United States, 5th Edition, p. 14). Additional responsibilities directly related to student instruction include student assessment, data collection and dissemination, and communication. In addition to acting as the RR interventionist, the RR teacher is envisioned by advocates of the program as a literacy resource in her school, though this role is not elaborated in the Standards and Guidelines. In this idealized role, the RR teacher acts as an instructional expert whose day-to-day interactions with students, staff, and administrators naturally build understanding and receptivity to the practices of Reading Recovery throughout the school.

Over the 2011-2012 school year, CPRE observed that the majority of i3-trained RR teachers were in compliance with the instructional expectations of their role – those related to the delivery of the intervention itself. CPRE also found that RR teachers are generally viewed by their colleagues within the school as experts in literacy instruction. However, there was considerably more variation in the extent to which i3-trained RR teachers actually served as instructional resources to others in their schools, and the extent to which they influenced their colleagues’ literacy beliefs or practices.

Wearing multiple hats

The Standards and Guidelines recommend that RR teachers play dual roles in their schools: teaching RR students in one-on-one sessions for roughly half the day, and assuming other responsibilities during the other half. The results of the daily activity log (see Table 15) completed by 838 RR teachers (2,229 work days) suggest that virtually all RR teachers worked with students in one-on-one lessons nearly every day. On average, RR teachers spent just over two hours per day conducting these lessons. In the other half of the school day, most RR teachers worked as interventionists serving individual students or small groups. Some spent the second half of their day working as classroom teachers, English as a Second Language (ESL) or special education teachers, or as instructional coaches.

Table 15 provides a detailed breakdown of RR teachers’ days, showing the prevalence and time spent on specific activities on the days when log respondents indicated they had worked (see footnote 1 for an explanation of how percentages in this table were calculated). After providing one-on-one RR lessons, the second and third most common activity RR teachers reported engaging in were “planning” and “small-group instruction.”

Table 15: Major Daily Activities of RR teachers¹

Major Activities	Percent of RR Teachers	Percent of Logged Days	Conditional ⁹ Average Time in Minutes
1-on-1 Reading Recovery Lessons	94%	94%	143
Planning	94%	93%	135
Small Group Instruction	74%	83%	121
Administrative Work	56%	69%	77
Classroom Teaching/Co-Teaching	46%	69%	119
Communicating About Reading Recovery	36%	54%	37
Receiving Professional Development	35%	40%	141
Administering Assessments	24%	36%	92
Instructional Coaching	16%	46%	64
ELL Instruction	6%	63%	97
Other	87%	82%	81

Source: RR Teacher Activity Logs. n=2,229 logged workdays from 838 RR Teachers.

Note: Activities are ordered by prevalence across teachers.

Among the 838 teachers who completed at least one log, 94 percent of teachers recorded time conducting 1-on-1 lessons and doing planning on at least one of their one to four logged days, while 74 percent recorded time doing small-group instruction on at least one of their logged days.² The remaining activities were far less consistent in their prevalence across teachers and

¹ Because a substantial part of the work of RR teachers is expected to vary from teacher to teacher and day to day, the data in this table is presented as follows. The first column shows the percentage of RR teachers who engaged in each activity on at least one of their logged days. This provides an indication of how many RR teachers engage in this kind of activity frequently enough to be observed in a random sample of one to four days. The second column shows the percentage of logged days on which the teachers identified in the first column reported engaging in each activity. This provides an indication of how regularly these teachers engage in each activity, given that they do the activity at least some of the time. The third column in the table shows the average amount of time per day these teachers engage in each activity on those days on which the activity occurs. This provides an indication of how much time per day RR teachers tend to devote to each activity on those days that the activity occurs.

² This should not be taken to suggest that 6% of RR Teachers never do 1-on-1 lessons or planning. Although 6% of teachers did not record time conducting 1-on-1 lessons in their activity log, a follow-up question asking "How many students did you conduct 1-on-1 Reading Recovery lessons with today?" revealed that only one teacher on one logged day responded, "I did not conduct any 1-on-1 Reading Recovery lessons today." This suggests that the log data may be underestimating prevalence of activities due to inaccurate reporting on a small number of logs.

days, as reflected in the lower percentages of logged days and RR teachers. For example, only 16 percent of RR teachers engaged in instructional coaching on only 46 percent of their logged days. On these days, these teachers spent an average of one hour engaged in instructional coaching. Classroom teaching or co-teaching was more common, occurring for 46 percent of RR teachers on 69 percent of logged days, with an average of approximately two hours spent teaching on those days. Very few teachers engaged in ELL instruction, but those who did engaged fairly regularly, devoting over 1.5 hours per day on 63 percent of their logged days. Very few teachers (i.e., 24 percent) reported administering assessments on any of their logged days. For those who did, they spent an average of about 1.5 hours on 36 percent of their logged days. Administrative work was much more prevalent, with 56 percent of teachers spending an average of just over 1 hour on 69 percent of their logged days. Just over one-third of RR teachers reported participating in professional development for an average of almost 2.5 hours per day on 35 percent of their logged days.

Thirty percent of RR teachers reported perceiving conflicts between their RR roles and their other duties, or that it was difficult to plan and conduct their RR lessons in addition to their other job responsibilities. "I would love to do [Reading Recovery] full time and not be the instructional coach," one RR teacher explained. "But we don't have any positions like that in our district. I feel very split in having [to be] Reading Recovery and an instructional coach."

Less frequently, RR teachers described staffing arrangements that stretched them across multiple buildings:

I think there's only two of us in the district who are split between two schools and it's just difficult because each school is a little bit different with their administrator and with how they do things. Not necessarily with Reading Recovery - that's pretty straightforward with both schools. But in terms of everything else and how meetings are handled and how small groups and all that stuff. So it has been challenging, and it is kind of like I have double meetings because I have to do all the meetings for both schools because I work with both schools and the kids in those grade levels. So you know, it's a lot of meetings.

In part because of the multiple hats they wear, a majority of RR teachers reported being extraordinarily busy, and the daily activity log confirmed that they worked long hours. Responses indicate that on the vast majority of days (>75 percent), RR teachers started working at or before 8:00am and continued working until at least 4:00pm. However, it was not uncommon for RR teachers to work even longer hours, with 25 percent of logged days beginning at 7:15am or earlier and ending at 7:45pm or later. The median workday for an RR teacher involved 9 hours of work-related activities, and 25 percent of workdays lasted 10 or more hours.

Protecting Time for RR Lessons

Despite the difficulties stemming from their busy schedules and multiple roles, RR teachers in i3 schools fulfilled their daily RR lesson responsibilities adherently. On 73 percent of the days on which RR teachers reported conducting one-on-one lessons, they reported working with four students. The activity logs indicated that RR teachers worked with only three students on 14 percent of the logged days on which they conducted lessons, two students on six percent of those logged days, and only one student on three percent those of logged days. Very infrequently (i.e., on < three percent of logged days) did RR teachers report working with more than four students in one-on-one lessons.

The most common reason RR teachers cited for their work with fewer than four students was that one or more students were absent from school that day. Other common reasons included mandatory testing, field trips, conflicts with the RR teacher's own schedule and responsibilities. One teacher described disruptions of this sort:

There were a couple times when I wasn't able to see kids... I was at school but I couldn't see the kids because I had to be in team planning. There was one other time that I couldn't see the kids because they had a concert at my time.

While logistical conflicts like these were commonplace, they were also regarded as largely surmountable with the cooperation of RR teachers' 1st-grade colleagues. i3 RR teachers almost universally approached their work with an attitude of flexibility and commitment to doing whatever was required to make up missed lessons. As one teacher explained:

If we're pulled out of the building for a meeting, a district meeting or something, I try to make a lesson up another day. I'll do two, one in the morning, then one in the afternoon if I know I'm going to miss one. For example, last week we had a conference, so what I did is I didn't go to the conference in the morning time... So I really try to make sure that I meet with [RR students] every day. Sometimes it's difficult, like I said, once, when we had the Reading Recovery conference that we went to, I made up an extra day during the week. I took them twice, like once in the morning and then I pulled them again in the afternoon.

Communication, Leadership, and Influence

The Standards and Guidelines require that RR teachers "communicate periodically with 1st-grade teachers and other school personnel," and suggest that they "contribute to the development of a school team to monitor program progress" (p. 11). The time RR teachers have available for communication is quite limited—36 percent of RR teachers reported spending just over half an hour on 54 percent of their logged days communicating with parents, teachers, their principal, RR staff, and other individuals about Reading Recovery.

While most communication was focused on the progress of RR students, many RR teachers also shared more general information about literacy with their colleagues—68 percent of RR teachers indicated that they had opportunities to do so. Many RR teachers also reported participating on school-wide data or Response to Intervention (RTI) teams (see section entitled “Follow-up Support” for more information on RTI). A smaller number had explicit leadership roles in their schools, in which they were specifically charged with sharing their knowledge about literacy instruction with others.

Collaboration with 1st-Grade Teachers

Communication between RR teachers and other individuals in their schools varied substantially depending on the stakeholder. As would be expected, RR teachers reported being most closely connected to their 1st-grade teacher colleagues. RR teachers are generally considered a part of the 1st-grade instructional team, and many described regular communication with the classroom teachers of RR students. Eighty-two percent of RR teacher survey respondents reported communicating weekly, if not daily, with a 1st-grade teacher about Reading Recovery. These interactions occurred mostly through brief, informal interactions throughout the day, rather than through scheduled meetings.

RR teachers reported that the majority of their communication with 1st-grade teachers was focused on specific students' progress, and ways to support them in the regular 1st-grade classroom. One teacher described a typical interaction:

I can tell them “This is what I’m working on.” I see that one of my students is a stronger writer than he is a reader. So I said “Well, make sure that he’s reading everything that he’s writing.”

RR teachers also consistently reported communicating with their students' classroom teachers as the students complete the program. First-grade teachers confirmed through the survey that this communication occurred; 91 percent reported that the information they received from the RR teacher helped determine ways to support that student in class. Sixty-seven percent of 1st-grade teachers also reported sharing the information with the student's parent or guardian.

RR teachers also reported finding occasional opportunities to share more general information about instructional philosophy or strategies with classroom teachers. Most often, these opportunities were informal, and arose when classroom teachers began to notice progress in students who are participating in Reading Recovery. One RR teacher explained:

[B]ecause they were showing such growth [the other teachers] kind-of said “Okay, if you’re making progress, what are you doing?” And so then we just got a hold of our year’s lessons and “Here’s what our lessons look like and here’s what we do.” They came and watched lessons and then we kind-of helped them formulate how to do folder construction in a similar way.

This informal peer coaching, like much of the interaction between RR teachers and their classroom colleagues, tended to happen in short bursts, and to be focused around one or two very specific topics or strategies. For instance, in interviews RR teachers spoke of showing 1st-grade teachers how to teach students to break words into parts rather than relying solely on traditional phonics techniques, how to use sound boxes with their lower readers, or how to take and analyze a running record. One teacher commented:

I have pulled my 1st-grade teachers into my office and I have showed them a few things, just in the few minutes of free time I have in a day. I have showed them some things that are exciting to me and that I know are really, really working and that will work well in the classroom, and I know that they have started implementing those.

And though it was not typical, some RR teachers reported that the influence of RR was spreading beyond 1st-grade, and that other teachers in their school were using RR techniques in their classrooms:

What [special educators who are exposed to RR are] finding is that the way we teach reading makes a special ed teacher better at teaching reading than what they were doing. So they want to use the same strategies. So “Hey, I want to be a better special ed teacher. What if I use Reading Recovery strategies for the reading part for my special ed students?” So then they do it ... and actually when you’re finished you have the most incredible documentation, your OS surveys.

As will be discussed later in this section, RR teachers collect significant amounts of data in the course of their one-on-one lessons, and many share data with 1st-grade teachers as a way of informing them of students’ progress and guiding classroom teachers’ work with students. Sixty-seven percent of RR teachers surveyed reported that the data they collected during RR lessons was used “moderately” or “extensively” to set literacy learning goals for RR students in their general education classrooms, and 57 percent indicated that it was used to tailor classroom teachers’ instruction to their RR students’ needs. One RR teacher explained:

So [I find what] I can copy off the charts [I create in RR] to keep track of the words they’re able to write and the words they’re able to read. [I give the teachers] copies of those charts that [tell] which text level they’re at. And I try to use that information in all of my running records. Sometimes I’ll take a running record and say “This is what they’re doing, and this is what I do to help them.”

Although some RR teachers were certain that their colleagues were applying the information they shared about literacy instruction in their own 1st-grade classrooms, many reported that it was difficult to know whether the classroom teachers were actually incorporating these strategies

into their instruction. Just 22 percent of RR teachers surveyed felt that they had influenced other teachers' classroom instruction "to a great extent," with 29 percent reporting moderate influence. Approximately 41 percent of RR teachers indicated that they had little or no effect on the classroom instruction of other teachers. However, 75 percent of 1st-grade teachers surveyed indicated that they had learned about literacy from the RR teacher and the program, suggesting that in some cases the RR teacher is influencing the instructional practices of her colleagues more than she knows.

Many 1st-grade teachers also reported having a solid understanding of Reading Recovery itself. Roughly 92 percent reported that they understood the goals of Reading Recovery, and 89 percent indicated that they understood the theory of literacy instruction that underlies the intervention. Notably, only 40 percent indicated that they had observed a Reading Recovery lesson this year, and 30 percent stated that they have never observed one. This gap between 1st-grade teachers' reported understanding of RR and their actual exposure to RR instruction suggests, consistent with reports from some RR teachers, that many 1st-grade teachers may believe they understand the philosophy and approach of RR better than they actually do. In the future, this evaluation will continue to explore the ways RR teachers interact with and influence their classroom teacher colleagues.

Participation on School-level Teams

Many RR teachers reported participating on teams that worked collaboratively to monitor students' achievement and specify instructional interventions. These teams went by different names - intervention teams, data teams, student-support teams, Response to Intervention (RtI) teams - and provided the primary formal structure through which RR teachers shared information about RR and students' progress with others in their buildings. In many schools involved in the i3 scale-up, these teams were tasked with evaluating individual students' progress and specifying appropriate interventions where needed; they generally included the RR teacher, one or more regular classroom teachers, special education and ELL teacher, and the school psychologist. At times, the principal was also involved. In other instances, 1st-grade team meetings were the primary forum through which RR teachers participated in discussion and decision-making about students' progress and needs.

The RtI process, and RR's place in the RtI structure generally, will be discussed in more detail in the section entitled "Follow-up Support."

Formal Instructional Leadership

Although instructional leadership is not a requirement of the RR teacher role as defined by the Standards and Guidelines, some RR teachers have leadership roles in their schools, and some principals spoke of their efforts to tap into the RR teachers' literacy expertise. One principal explained that:

When there's a question about literacy or there's a question about reading or writing instruction, [the RR teachers] are going to be the leaders in the school for literacy. They're going to be the people that our other colleagues are asking questions to... we're trying to match up our teachers who are weaker in guided reading instruction with those RR teachers so they can see what the best practices look like.

A few RR teachers reported that they, or their teacher leaders, had been asked to lead professional development sessions related to a literacy topic. Some RR teachers described inviting classroom teachers, at the request of their principal, to observe behind-the-glass teaching sessions involving their students. In a number of cases, new RR teachers stated that they anticipated increased leadership roles in the future:

Next year ... we are implementing Fountas and Pinnell³ and my principal really wants me to, you know, I'm going to be helping all the teachers with this. I think the Reading Recovery is just a very much more intense form of Fountas, I mean, it's basically they're going to be doing literacy groups, they're going to be doing running records, they're going to have to keep logs and all that, so I'll be very prepared as an instructional coach to help them because I've been trained as a RR teacher.

Thirty-three percent of RR teachers' surveyed also believed their data was used "moderately" or "extensively" to determine topics for school-wide professional development, whether or not they were involved in providing the training.

CPRE will continue to study RR teacher roles in their schools through the lens of communication, leadership, and influence as the scale-up matures and RR teachers become more settled in their roles.

The Teacher Leader

Along with the RR teacher, the teacher leader plays a central role in the implementation of Reading Recovery in schools and districts. In addition to providing training and ongoing support for RR teachers, teacher leaders also administer one-on-one RR lessons to two to four students.⁴ Though this responsibility is less elaborated in the Standards and Guidelines, teacher leaders are also expected to raise awareness of Reading Recovery and work toward strong implementation at both the school and district levels. Through this combination of roles, the teacher leader is

³ Fountas & Pinnell's (Heinemann) collection of literacy resources includes a book-leveling system and a range of other materials and systems for assessing and instructing students. Fountas & Pinnell materials are aligned with Reading Recovery.

⁴ Teacher leaders work with four students during their training year, and at least two students per year thereafter.

often described as “a change agent” (Fullan, 1993) or, in Marie Clay’s terminology, a “redirecting system” (Clay, 1987, p. 46). In the RR model, the teacher leader is envisioned as a key ground-level advocate for the program, and is considered critical not just to ongoing implementation, but to the institution of the program as a sustainable component of a school.

In examining the roles of teacher leaders involved with the scale-up, CPRE found strong fidelity to the training and teacher-support functions detailed in the Standards and Guidelines. The extent to which teacher leaders’ roles included a focus on the advancement of Reading Recovery within schools and districts varied by how they conceptualized this aspect of their role, and by the specific circumstances of their positions within the districts or other entities that employed them.

A School-level Advocate

Nearly all RR teachers reported that the role of the teacher leader was critical to RR implementation, and that the training and instructional support they received from their teacher leader was consistent with the expectations detailed in the Standards and Guidelines. According to both RR teachers and teacher leaders themselves, there was significant variation in the roles played by teacher leaders across school contexts. In most cases, RR teachers reported that school visits by their teacher leader were fairly infrequent, and included limited contact with the principal or other teachers in the building. Sometimes this was due the principals’ inaccessibility; in other cases, teacher leaders did not have the time to track down every principal when an RR teacher was unable to. One teacher leader explained:

I’m in a large city. [It takes] half an hour to an hour . . . just to get to each teacher. And so if I’m going to do a couple of teachers a day to visit and then do class at night, I don’t have time to see the principals. I had to tell the trainees: “If you could possibly get that principal standing near your room when I walk in, at least I can say hello.” Some of them were good at it. Some of those teachers believed that, “Yeah, my principal needs to know you and all that.” Others just didn’t help me out at all and so I didn’t even have time to hunt the principal down, stick my head in the door, because of the travel.

In other cases, RR teachers regarded their teacher leaders as important allies and advocates who helped ensure that the intervention was implemented as designed, and who helped educate administrators and teachers alike about Reading Recovery. One teacher noted:

[My teacher leader] worked with our principal, helping my principal understand the process and some of the things we were needing. Like, for instance, we needed a schedule change. . . . But she has also met with and worked with some of the teachers in our building. You know, kind of helping them understand Reading Recovery and kind of helping them understand our language.

A Crucial Support for Instruction

The pivotal nature of the teacher leader's role as a support for the daily work of RR teachers emerged as a strong theme in the first full year of research. Most RR teachers looked forward to both group training activities and their teacher leaders' site visits as opportunities to learn more about instruction. They described teacher leaders as having extensive expertise, and as skilled at helping teachers build their problem-solving and analytical abilities. Many RR teachers particularly valued sitting side-by-side with their teacher leader during a lesson, and the discussions that followed. One teacher recalled:

Once I think she spent three hours with me on one student. You know, after she leaves or the student leaves, then we have usually quite a lengthy discussion on what I tried what she'd seen and what the student's reading behaviors are... She's able to pick up on things that she sees the students doing that I'm not picking up on. I always walk away with a ton of information and things to try and kind-of open my eyes to certain behaviors that the student's showing.

In general, RR teachers found their teacher leaders accessible, responsive, and deeply committed to both their own success and that of their students:

She will go beyond for each person that's in her class. I mean, she loves what she does, she believes in what she does and she wants to continue and I feel very comfortable talking with her about needs. She's non-threatening yet she makes sure you get your work done. I can't say enough positive things about her. She's been fabulous. She's interesting to listen to. She's very deep in her thinking and analyzing of Reading Recovery and I feel like I've learned a lot from her in that regard. And I'm a pretty deep thinker but it just, you know, the way she has helped us analyze and really read students has been amazing to me.

The high regard in which RR teachers held their teacher leaders was a particularly strong finding in this year of evaluation. Negative comments about teacher leaders surfaced occasionally—a few teachers found their teacher leaders less than responsive by email, for instance. However, these complaints were very much the exception.

A Context-dependent Role

While their basic tasks and responsibilities to RR teachers were consistent across sites, the specifics of other aspects of teacher leaders' roles varied significantly from one context to the next. In some regions, teacher leaders had primary responsibility for recruiting new teachers and schools to participate in the scale-up; these teacher leaders spent significant time meeting with administrators and "pitching" Reading Recovery. In other regions, teacher leaders worked as a classroom teacher part of the day while supporting a small caseload of RR teachers.

Teacher leaders' positions also differed significantly in terms of funding. Most teacher leaders are employees of one of the districts they serve and, like RR teachers, many have additional responsibilities as instructional coaches or teachers. Unlike most RR teachers, however, teacher leaders frequently have responsibilities stretching across multiple schools or even multiple districts; districts that don't have full Reading Recovery sites often purchase support for their RR teachers from teacher leaders based in other districts. In a few cases, teacher leaders involved with the scale-up were employed by partner universities, or by state agencies involved with the provision of early literacy services.

The extent to which teacher leaders were able to execute tasks laid out for them in the Standards and Guidelines varied widely by context. In some cases, teacher leaders were able to address the broader vision for their roles by positioning themselves as key experts and instructional consultants for school and district leaders. In other cases, they were union employees with primary obligations to their own schools, and lacked both the time and access required to influence district or school policy.

The RR teacher selection process offers one example of the vast differences in teacher leaders' ability to meet the expectations of their role. The Standards and Guidelines recommend that teacher leaders "assist in recruiting and identifying appropriate teachers for the training class" (Standards and Guidelines of Reading Recovery in the U.S., 5th Edition, p. 23). Teacher leaders note that RR teacher selection is an important task, as the RR teacher position should not be offered to weak teachers, and some describe their direct input in the process. One teacher leader reflects:

When I had contact with the principals and superintendents and other administrators, any time I talked to them and the issue of training and who they might train was brought up, I would go over with them... I had buzz words that I used that would suggest the type of person: The knowledge-seeker, the well-qualified person who's already doing well at whatever they're doing now, those kinds of things.

Other teacher leaders reported having no input into who is selected for the RR teacher role in their sites; the choice is entirely at the discretion of principals with whom they rarely interact. Variations like these were common in accounts of teacher leaders' roles, which raises questions about the alignment of their intended role and the contextually enacted reality.

Principals

A vast literature documents both the importance and the complexity of principals' roles in the implementation of instructional programs (Hallinger, 2003, 2005; Supovitz, Sirinides, & May, 2010; Heck, 1992; Leithwood & Jantzi, 1992), and Reading Recovery appears to be no exception. While precise understandings about how principals impact RR implementation are

still being developed, and considerably more research is needed on this issue, a few preliminary trends were identified during the first year of research on the scale-up.

There were striking differences in the ways RR teachers and teacher leaders characterized their principals' involvement with RR implementation. Some principals were described as champions of RR who worked diligently to ensure that RR exists and is protected in their buildings; a few were even RR-trained themselves. It was often these highly-invested principals who identified leadership roles for RR teachers, and who found ways to harness the teacher leader as a resource.

Other RR teachers reported, however, that their principals were neither highly involved in the implementation of Reading Recovery, nor particularly helpful. Some mentioned that their principals were hesitant to allow them to devote adequate time to RR lessons; one teacher remarked:

Sometimes I think she sees it as "This is taking away from me having my instructional coach." And that's frustrating. Because when she would want me to do something, I would say "I have Reading Recovery students at that time." Things would come up and she would want me to do certain things and I would have to tell her "This is my time for Reading Recovery."

Principal prioritization of RR - that is, the extent to which they protected their RR teacher's time - appeared to be a key implementation factor in some schools. Some principals facilitated RR implementation by creating structural supports for student progress, communication, data-utilization, and buy-in at the school level. These supports were varied in nature and in the degree to which they were coordinated with other systems in the school.

District Leadership

According to the reports of teacher leaders, the extent to which district-level administrators were involved in the implementation of Reading Recovery was highly variable. In some districts, teacher leaders spoke of strong advocates for RR who held district leadership roles; these individuals often supported RR implementation by ensuring that district funding was available to support the program, and focusing on instructional alignment. The role of the site coordinator (typically a district-level administrator) in RR implementation at both the school and district level warrants particular attention going forward, as many teacher leaders described their site coordinators as relatively uninvolved with the program's implementation.

Implementing the Instructional Core: Identification to Discontinuation

Examining the roles and relationships that support and define Reading Recovery within a school is one part of understanding its implementation. A second level of investigation focuses on the processes and activities related to Reading Recovery instruction itself—those at the instructional

core of the intervention. CPRE has observed that the instructional core of Reading Recovery comprises several related components: selection of students to receive Reading Recovery; delivery of one-on-one lessons; monitoring of student progress; and programming of follow-up support for students who have completed the intervention. During the first full-year of evaluation, CPRE identified findings related to each of these processes.

Student Selection

Reading Recovery strives to serve the students at greatest risk of lifelong low literacy, an aspiration that is operationalized in the Standards and Guidelines by the suggestion that schools provide the intervention to at least the lowest-achieving 15 to 20 percent of a 1st-grade cohort. The belief that early, expert, targeted instruction can accelerate the literacy progress of any child lies at the heart of the Reading Recovery philosophy; the Standards and Guidelines therefore indicate that the selection of students to participate in the intervention should be guided by Marie Clay's assertion that "all kinds of children with all kinds of difficulties can be included." Specifically, Clay's recommendation states that:

...exceptions are not made for children of lower intelligence, for second-language children, for children with low language skills, for children with poor motor coordination, for children who seem immature, for children who score poorly on readiness measures, or for children who have been categorized by someone else as learning disabled. (Standards and Guidelines of Reading Recovery in the United States, 6th Edition, p. 6).

To facilitate this inclusive selection process, the Standards and Guidelines require schools to use only the OS to select the lowest-achieving 1st-grade students, and to serve the lowest scorers first.

When investigating the student-selection practices of schools participating in the i3 scale-up, CPRE examined the systems used to identify a preliminary pool of students who may qualify for Reading Recovery, and the processes by which students were ultimately selected from within that larger group to receive the intervention. This examination revealed considerable variability in the ways schools determined which students receive RR, and considerable deviation from the stated goal of targeting the lowest-scoring students without exception.

Identifying Potential Candidates

CPRE found that most schools in the scale-up used a two-tiered process to select students for Reading Recovery. While a small number of schools elected to administer the OS to all 1st-grade students, most assessed only students who had already been identified as potential candidates for RR. Typically, a preliminary pool of low-achieving first graders was assembled, and the students from that group who scored lowest on the OS were assigned to the intervention. Much

of the variability observed around student selection occurred in this first-round identification of prospective participants. Variations at this stage were present in which school staff members were involved in the nomination of students to the preliminary pool, the criteria on which the nominations were based, and the level (classroom, school, or district) at which the 1st-grade cohort was defined.

In most schools, the first-round student identification process was collaborative, informed by input from first grade and, often, kindergarten teachers. In many cases, the student identification teams also included interventionists, ELL and special education teachers, instructional coaches, and school administrators. Typically, schools nominated students to the preliminary pool on the basis of reading levels determined using Fountas & Pinnell, Accelerated Reader⁵ or other leveling systems, or assessment data from teacher-administered tests. However, in a number of schools, students were nominated to the pool based on classroom teachers' general observations or "gut" impressions of their needs. In a few cases, a RR teacher was uncertain how the preliminary pool of students had been identified. When asked to name the assessment used to identify the lowest achieving students in her school, one RR teacher responded: "Let me check... I don't have that. I just have the baseline test evaluation chart. I don't have the actual name of what they're using."

An additional source of variation in student selection was inconsistency in the level at which the 1st-grade cohort was defined. In some schools, teams looked across all 1st-grade classrooms to identify the lowest first graders in the whole school, even if they were distributed unequally across classrooms. In other cases, each classroom teacher was asked to nominate a pre-determined number of students, regardless of whether they were among the lowest 1st-grade students overall. One teacher explained:

At the end of the year I will give the kindergarten teachers an evaluation sheet. They will go through and they will list their students from top to the bottom. Coming next year, I will go and I will test the bottom twelve. I actually did twelve. And the way that they asked me to do it was do four in each classroom, because I have three 1st-grade teachers.

In a few cases, students were nominated to the preliminary pool at the district level. This led to more students being identified at some schools than at others.

Making Final Selections

CPRE found somewhat more consistency in the processes used in the second tier of the selection process, where students are selected to receive RR from the preliminary pool of prospective candidates. However, significant departures from the Standards and Guidelines were evident at this stage as well. In keeping with the Standards and Guidelines, RR teachers were generally

⁵ Accelerated Reader is a product of Renaissance Learning, Inc.

responsible for selecting students from the preliminary pool, and generally selected those who scored lowest on the OS. Occasionally other data was considered as well, and in these instances it was usually at the insistence of a school or district administrator; RR teachers sometimes reported that these decisions were beyond their control.

Many RR teachers described looking to their teacher leaders for help in interpreting students' Observation Survey scores during the selection process, or in choosing between students with very similar scores. A few RR teachers reported that the final selection of students rested entirely in the hands of their teacher leaders. One RR teacher commented:

I can't tell you if [the selected students] were the lowest performing out of [the] 20 percent [the school identified for screening]. I gave the Observation Survey selection sheet results to my teacher leader. Based on the information that is on that sheet, she is the one who makes the decision as to who receives the services.

In a few instances, someone else at the school made the final selection:

One of our risk teachers—she is running [the Reading Recovery student selection process]. So she put it on a computer and she printed out a list. And then we go from that list....I think the computer kind of spit out, you know, from lowest to highest. After we enter our Observation Survey information and it kicks out . . . the range from low to high, and that's where we get [the final list of RR students] from.

As was the case with first-round nominations, final selection decisions were sometimes impacted by the level at which schools or districts compared prospective Reading Recovery participants with their peers. Most selection occurred at the school level, such that each student's OS score was compared only with those of other students in the same building. However, in a few instances selection happened at the district level, and each student's OS score was compared with those of other students across the entire district. An RR teacher in a district that used this approach recalled a student she would have considered eligible for RR because he was among the lowest readers at his school; this student was not selected for RR because he scored above the district-wide cutoff on the OS.

Schools also differed in the extent to which students who qualified for Reading Recovery based on their OS scores were excluded for other reasons. A significant number of teachers reported that their school or district policies, or their understanding of Reading Recovery policy, required the exclusion of students who had Individualized Education Plans (IEPs) for special education services, or who were repeating first grade. One teacher explained:

If a student has an IEP with reading goals on it, they can't be in Reading Recovery. And we have one little girl that was too low to even fit in an LLI [Leveled Literacy Intervention] group, so she was a prime candidate for Reading Recovery, but that's the case with her. She has an IEP and gets reading services so we couldn't take her, and we wanted to. And there was a boy that we had done the second-time-around Observation Survey on him and we were getting ready to take the second round, and my teaching partner went to go look in his file for something and found that he had been retained in first grade. He had come from a charter school I think, and did first grade last year and so he was in first grade again this year. And if someone is a retainee, you know, from first grade, they don't qualify for Reading Recovery.

Attendance and behavior problems were also cited by a significant number of teachers as reasons why students with very low OS scores were disqualified for Reading Recovery. In these instances, students with chronic problems were seen as unlikely to meet with success in the intervention, and were often excluded outright. Describing her school's selection process, one RR teacher noted that first graders who had had 10 or more absences in kindergarten were automatically excluded. This policy had resulted in the exclusion of four of the school's twelve lowest readers from Reading Recovery.

RR teachers often expressed disappointment over school- or district-level policies that prohibited certain students from receiving the intervention. Several noted that the policies were in conflict with Marie Clay's teachings, or that the selection process at their school was not aligned with teachers' perceptions of which students were most in need. "We have two students who are repeaters who really need it, but that's something we don't do," explained one teacher. "They're like the lowest in the whole grade now and we can't serve them."

CPRE also encountered mixed messages from within RR about the exclusion of 1st-grade repeaters and students with IEPs. A number of RR teachers indicated that their teacher leaders counseled them not to select students with IEPs, and some referenced efforts by special education teachers and supervisors to prevent IEP students from receiving Reading Recovery. Others, however, reported having teacher leaders who were adamant that students with IEPs should not be excluded. All in all, CPRE noted considerable confusion around this issue.

A number of teachers also expressed regrets about students who started school partway through the year, after the RR selection process was complete:

Unfortunately, because they either came in after our assessments were completed or ... you know some were admitted to the school later, we weren't able to pick them up. So there were a few that we did not get that we would've liked to have taken.

In only a handful of cases did RR teachers express the wish for more restrictive student selection criteria. These teachers were concerned that certain students selected for the intervention would be less successful in RR than others with slightly higher OS scores. One teacher whose RR students were selected at the district level reflected:

I know I probably don't agree with what Reading Recovery has to say about it, because you always take the lowest kids even if they have an IEP and I get that philosophy... [but] two of mine missed most of kindergarten because of behavior. One of them was on a significant behavior IEP... Two of mine are ELL, and one of them has a cleft palate with 90 percent unintelligible speech and it was never fixed... At the building level I think we would have chosen another student that would really fly and benefit from it.

Because of the inconsistencies across schools, and the conflicting understandings about whether certain students should be excluded, CPRE identified the student selection process as a site of considerable deviation from the Standards and Guidelines. This deviation is potentially significant in terms of the alignment of the intervention as it is implemented with the stated philosophy of serving all children. This issue warrants deeper investigation going forward - is Reading Recovery's positioning as an RtI intervention fundamentally at odds with its mission of serving all children⁶?

Reading Recovery Instruction

No component of the Reading Recovery model is more vital to the intervention's instructional core than the enactment of the RR lesson itself. Reading Recovery instruction is grounded in the specific set of activities identified and detailed by Marie Clay, and one lesson has much in common with another in terms of sequence, language, and even rhythm and tone. Within that basic structure, instructional moves vary based on teachers' nuanced observations of individual students' needs. Controlling the sequence and timing of RR lessons in order to deliver targeted instruction within the short 30-minute session is therefore a key to successful implementation. Helping teachers master this lesson format is a central role of the teacher leader.

Reading Recovery is designed to be a highly data-informed intervention; observations made and documented during each day's lesson should shape the content and direction of future lessons. The Standards and Guidelines require teachers to maintain records on every student—specifically OS documents and summaries; predictions of progress; lesson records; running records; and records of reading vocabulary, writing vocabulary, and text reading level.

⁶ Students with IEPs would, by definition, be beyond Tier II and ineligible for Reading Recovery under many schools' RtI models.

Lesson Structure and Execution

RR teachers overwhelmingly reported fidelity to the lesson structure developed by Marie Clay and endorsed by Reading Recovery; CPRE's investigation of the delivery of RR lessons in i3 scale-up schools confirmed this finding; strong fidelity to the standard format of lessons was observed, even in schools that struggled with other aspects of implementation.

Each lesson began with re-reading familiar books and a running record; then word or letter work on the wallboard; story composition; assembling a cut-up sentence; and finally previewing and reading a new book. Data from the daily activity logs confirm the RR teachers' adherence to this structure. Each RR teacher who reported conducting one-on-one lessons was asked to complete a set of questions about their lesson with one randomly sampled student from that logged day. In 17 percent of the cases, RR teachers reported that they spent that particular lesson "roaming around the known" as part of the initial sequence of work with a new RR student. In another five percent of the cases, RR teachers reported administering the OS in order to obtain baseline or post-intervention performance data for a student. In the other 78 percent of the cases, RR teachers reported conducting a standard one-on-one lesson with the randomly sampled student they were asked about.

Table 16 shows the prevalence of specific instructional activities during these lessons. The first seven activities, Reading Familiar Books through Reading a New Book, are essential elements in the instructional experience of RR students and are expected to occur in every one-on-one RR lesson. The high percentages in Table 16 confirm that these activities are included in virtually all one-on-one lessons. The activity most likely to be skipped was Assembly of a Cut-up Story, although it was skipped in only 10 percent of the lessons.

Many of the RR teachers-in-training interviewed during the 2011-2012 school year reported difficulty fitting their lessons into the recommended 30-minute timeframe. "It's hard for me," one teacher said. "I'm a teacher that likes to keep teaching and so I find it very difficult to stay within the 30 minutes." Teachers reported using timers to ensure that certain portions of the lesson did not take too long and consulting with their teacher leaders about ways to stay on schedule. The daily activity logs indicate that the vast majority of RR teachers spent just over two hours a day providing lessons to four students—an average of roughly 30 minutes per lesson (see Table 15). This suggests that RR teachers may be overestimating the extent to which their lessons tend to run over time, or that they may be running long on certain activities or lessons and making up the time on others.

Table 16: Frequency of Instructional Activities During Reading Recovery Lessons

Instructional Activities	% of Lessons
Reading Familiar Books	99%
Running Record Administered	95%
Word and/or Letter Work	96%
Story Writing	97%
Assembly of a Cut-up Story	90%
New Book Introduced	97%
Reading a New Book	96%
Created a Lesson Record	65%
Created a Record of Text Level	43%
Created a Prediction of Progress	9%
Created a Record of Reading Vocabulary	29%
Created a Record of Writing Vocabulary	54%

Source: RR Teacher Activity Logs. n=1,663 logged 1-on-1 lessons from 740 RR teachers.

Monitoring students’ progress

Reading Recovery regards the monitoring of student progress as integrally connected with the provision of instruction. Teachers are expected to maintain lesson records, including daily running records, and to collect data on students’ text reading levels and their progress in specific areas on a daily basis. Based on the review of intervention documents submitted for 50 randomly sampled students and the activity logs completed by RR Teachers, CPRE found that some components of this documentation are completed more consistently than others. For example, 100 percent of the randomly selected student files reviewed at the end of the 2011-12 school year included documentation of daily running records. However, the daily activity logs reveal less regularity in the completion of lesson records and records of text reading level (TRL). The lesson record and TRL are not part of the explicit instruction experienced by the student, but are used by the RR teacher to record the details of what was accomplished during the lesson. Lesson records are supposed to be created for each lesson, and each student’s TRL is supposed to be updated at least weekly. However, according to data from the daily activity logs, RR teachers reported creating lesson records for only 65 percent of their lessons, and recording TRLs for only

43 percent of their lessons; these data suggest that the details of many lessons are not being documented in a formal way.

The remaining three activities shown in Table 16—the creation or modification of a prediction of progress, record of reading vocabulary, or record of writing vocabulary—are not required in every lesson. The completion of these activities on any given day is at the discretion of each RR teacher. Of these activities, the one completed least frequently was the prediction of progress, which was created or modified in only nine percent of lessons, or roughly every one to two weeks. Records of reading and writing vocabulary were updated more frequently, in 29 percent and 54 percent of lessons, respectively.

Why many teachers document some aspects of students' progress more sporadically than others is unclear. It may be that teacher leaders are advising teachers that some documents need less frequent attention than formal policy suggests. It is also possible, given teachers' reported difficulties with completing their lessons on time, that new RR teachers tend to minimize documentation to allow more time for instruction. This issue will be investigated further in future years of the evaluation.

Data Use by Teacher Leaders

While the large majority of RR teachers spoke effusively about their teacher leaders' impact on their instruction, relatively few specifically mentioned teacher leaders' support of their data collection processes. However, a great deal of RR teacher-teacher leader collaboration was seen around the sharing and utilization of the student progress data that was collected. The Standards and Guidelines require teachers to periodically submit data to their teacher leaders and, indeed, the data pipeline between RR teachers and teacher leaders appears to be quite strong. Teachers share student data more regularly, and more comprehensively, with their teacher leaders than with anyone else in their schools or districts.

Student and lesson data are the primary means by which teacher leaders monitor the instructional core of RR implementation across multiple schools— it is their primary window into both students' and RR teachers' development. Missed lessons, unusual patterns in text reading level growth, and gaps in the data suggest trouble, and serve as indications to teacher leaders that a site visit is necessary. One teacher explained:

I know that my teacher gets really worried. Well, not worried, but she's looking at that data every week and if a student is not moving one level a week then she wants to come out a see what's going on. I guess that's kind-of that we should be expecting each student to go up once a week—like [one] level—and if they're not we need to look at that.

The procedures and timing of data-sharing between RR teachers and teacher leaders varied. Individual teacher leaders appeared to develop their own requirements and timeframes for collecting and analyzing data, and these systems differed significantly from one teacher leader to the next. Some teacher leaders collected hard copies of student reports or portfolios on a weekly basis, while others requested that certain data be submitted via email or entered into a database. Similarly, teacher leaders developed their own systems for debriefing with RR teachers about their data; many plan an annual conversation with each teacher, once the annual data have been collected for submission to IDEC. In between annual debriefs, some teacher leaders scheduled regular meetings with RR teachers; others waited for signs of trouble, or a request for help, before discussing data in-depth with individual teachers. While these differences are partly a function of teacher leaders' individual approaches to oversight and the size of their caseloads, they are also related to the RR teachers' skill levels and experience. In general, teachers-in-training receive more consistent and proactive supervision than those who are already trained. One teacher leader described her process for reviewing student data from her trained RR teachers this way:

"Five-week checks" [is what] we call them and they tell the level the child is on and how the writing's coming underway. And there's a place to write "help," there's a place to write "child getting on," and so we're constantly keeping track of where these kids are also.

Interestingly, CPRE observed that a small minority of teacher leaders thought very differently about their roles with respect to their teachers' data. A long-time teacher leader explained her perspective:

I don't collect stuff from them at all... I'm working really hard to build, I don't know what you'd really call it... a community of responsible teachers? And so when we come together, whether it's for weekly training class or whether it's a monthly ongoing professional development session, there's always conversation about student progress and there's always invitation both by me and my colleagues: "Do you need any help? Do you need help from each other? Do you need help from me? Does anybody need a visit?" And through that dialogue, I hope, I think, I believe I'm achieving the same kind of—I'm looking for the same kind of outcomes that [other teacher leaders] are looking for with paper. But I'm very loath to turn it into "I'm in charge of that data!"

While not common, statements like this from teacher leaders were surprising given Reading Recovery's focus on data-driven instruction, and RR advocates' heavy reliance on student data to demonstrate the intervention's effectiveness to school and district decision-makers. However, highly individualized practices are quite consistent with the ethos of Reading Recovery, and mirror its instructional belief that learning must be strengths-based and self-directed. Though

the assumption is more often implicit than overtly stated, Reading Recovery UTC directors and teacher leaders - especially some of those with long tenures - appear to believe that adults thrive under the same terms.

After the Intervention

Reading Recovery strives to produce accelerated learning in the short-term, and to equip students to maintain their growth in future years. With the understanding that not all students will reach their progress goals, Reading Recovery further strives to provide useful data to school-level decision-makers regarding the needs of students who fail to successfully “discontinue,” particularly those who are suspected of having disabilities. In CPRE’s conceptualization, the processes by which students are moved out of Reading Recovery, the types of supports they are given after the intervention, and the ways in which RR data informs instructional follow-up constitute the final component of Reading Recovery’s instructional core.

CPRE examined the frequency with which students met their short-term learning goals within 12 to 20 weeks and were discontinued, as well as the numbers of students who fell short of their goals, or left Reading Recovery before the conclusion of the intervention. CPRE also explored the processes around discontinuation, the management of situations in which students’ progress goals were not reached, the reasons why the intervention was sometimes disrupted before the end of the 12- to 20-week cycle, and the use of RR data in the special education referral process.

Discontinuation/Non-discontinuation

The successful discontinuation of students is obviously a key goal for RR teachers; however, it is well understood within the RR community that a significant number of students will not, and that those who do not discontinue may require continued support in order to maintain their progress. Of the 13,328 RR students for whom final intervention status was recorded in IDEC⁷, 52.4 percent of students in Reading Recovery in i3 schools during the 2011-12 school year successfully discontinued with no referral for additional intervention. Of the remaining students, 22.4 percent were referred for additional intervention, 4.7 percent changed schools, and 19.7 percent received less than 12 weeks of lessons (i.e., typically due to student absenteeism, the end of the school year, or the teacher leaving the school mid-year).

Particular students’ likelihood of discontinuing was a frequent topic of conversation for RR teachers; this is consistent with CPRE’s more general finding that RR teachers feel considerable personal responsibility for their students’ outcomes. Lamenting her school’s discontinuation rate for the year, one RR teacher said:

7 Final intervention status was recorded in IDEC for 86 percent of the full sample of 15,525 students in i3 schools during the 2011-12 school year. We have no information about the intervention status of the students with missing data, other than the fact that the RR teacher has not entered data for that student into IDEC.

Ugh. We had a few that discontinued. Not as many as we would have liked to have seen discontinue but we did have several that made a lot of progress and we did not have the full twenty weeks on the second-round students. So we feel that if we had had the full twenty weeks we possibly could have discontinued a few more.

Most RR teachers reported that they follow their discontinued students' progress once they leave the program, as the Standards and Guidelines mandate. A majority of those surveyed indicated that they check on former students' progress one to three times per month. One second-year teacher reflected:

I have discontinued most of my students. It's been good, and in addition to that most of the kids are doing well in 2nd grade. There are a few still who, you know, kind-of want that one-on-one and have dropped [in their literacy growth], and especially the ones who don't have quite as much support at home. They're still struggling with things moving on in the 2nd grade, but for the most part the outcomes have been really good.... I mean, I think for the most part my students have discontinued, but more importantly not only discontinued but like stayed on that path.

Some RR teachers described working closely with their teacher leader as students neared discontinuation, in the hopes that the extra support for their instruction would help the students make the final push. One first-year RR teacher noted:

I just talked to our teacher leader on Friday. I have a student that I think is close to discontinuing. But I've never had a kid discontinue before so I wanted her to come watch him to see what I need to do in the next six weeks of his program.

More questions were raised than answered in this year of the evaluation about how teacher leaders and RR teachers decide the appropriate text level at which a given student should be discontinued from Reading Recovery. While general guidelines exist and seem to be understood by many teachers, there appeared to be much variation. A number of teachers described collaborating with another RR teacher to obtain an independent opinion as to the child's final text reading level. Some teachers also submitted students' final OS data to their teacher leaders, who then made the final determination as to the child's discontinuation status.

Disruption of Service

While not particularly common, 18.6 percent⁸ of students who began Reading Recovery did

⁸ 2,407 of the 13,014 students with length of intervention data available in IDEC. We have no information about the length of intervention for students with missing data, other than the fact that the RR teacher has not entered data for that student into IDEC.

not complete the full 12- to 20-week instructional cycle. Many of the RR teachers involved with the scale-up seemed to regard attrition as an inevitable reality of the program. The reasons for students' failure to complete the intervention varied, but the most common factors appeared to be student mobility, attendance or behavior issues, and teachers' inability to complete two full 20-week cycles before the end of the school year.

Students with behavior problems were occasionally removed from Reading Recovery because they refused to participate or cooperate with the RR teacher. One teacher described an example of this situation:

The teacher couldn't get him to go with me. He would just sit there and really grab onto things not to go. He was also having challenges in the classroom. After several weeks of fighting this, he was unhappy, we couldn't get through lessons, so we decided take him out of Reading Recovery.

Other RR teachers mentioned losing students whose families moved away. "I had a child that was here on a Friday and gone on a Monday," one teacher recalled. Occasionally, students later returned to a school they had left; in those cases, some RR teachers described attempting to pick up where the student had left off.

Teachers who were unable to complete two full intervention cycles generally attributed that difficulty to a slow start at the beginning of the school year. Data collection in future years will include analyses to determine the prevalence of these delays and whether they occurred primarily in schools where RR was being implemented for the first time. Another issue for future investigation is the extent to which RR teachers were required to perform non-RR tasks in their schools at the beginning of the school year, thus delaying the start of Reading Recovery. Several teachers mentioned participating in the school-wide administration of DRAs or other assessments, for instance, prior to beginning RR lessons. The Standards and Guidelines stipulate that schools should administer the OS and begin serving students within two weeks of the start of the school year; however, it appears that this did not occur everywhere in 2011-12.

Follow-up Support

Schools varied in the extent to which they provided ongoing supplemental support to students following Reading Recovery. Most schools provided some manner of additional supports to students who did not successfully discontinue, and in a majority of cases the Rtl or intervention team was responsible for determining what those supports would be.

In many schools, RR is regarded by administrators and teachers as a component of the Rtl system, a more or less developed network of interventions designed to support struggling students and prevent unnecessary special-education placements. Rtl systems generally comprise a series of levels, or "tiers," of increasingly intensive instructional interventions, with regular classroom

instruction at Tier I. Students determined to require extra support beyond Tier I are advanced to higher tiers, where they receive more intensive support—often in a small-group setting initially, with individual services reserved for students who fail to progress in response to small-group instruction. In theory, the RtI system ensures that only those students who fail to respond multiple tiers of carefully monitored intervention receive psychological evaluations aimed at identifying possible disabilities; referral for a special education is therefore the culmination of the multi-tiered and often lengthy RtI process.

The supports available to students vary widely across schools participating in the scale-up, and in the extent to which they are organized as part of a formal RtI structure. Many schools identify Reading Recovery as either a Tier II or a Tier III intervention, and in some cases it is the most intensive literacy intervention offered to students before they are referred to special education. RR teachers involved with the scale-up frequently mentioned students who had failed to discontinue and were suspected to have a disability.

Some schools offer ongoing support to students who fail to discontinue from Reading Recovery. Frequently, this support takes the form of small groups provided by teachers funded through the federal Title I program or other interventionists. A number of RR teachers specifically mentioned that their students who discontinued from Reading Recovery were later served through Fountas & Pinnell's Leveled Literacy Intervention (LLI), a small-group intervention that is closely aligned with Reading Recovery.

Reading Recovery and Special Education Referral

The extensive data generated through Reading Recovery may help students move through the layers of the RtI processes and toward special education evaluation more quickly. While this assertion is based on preliminary findings and requires further investigation, it emerges from reports by RR and general education teachers that students' progress data from RR was an important source of information in the special education referral process. One teacher explained this benefit:

We've had a few kids who have not reached the expected progress and because of their time in Reading Recovery we know a whole lot more about them than just what they've had trouble with... We had two students who, at the end of their time, were recommended for further intervention. One has now been placed on an IEP. The other one has not, but because of their time in Reading Recovery they're getting much better services, because we knew a lot more about them.

First-grade teachers—who are generally responsible for completing the paperwork associated with a special-education referral—overwhelmingly reported that they found RR data helpful; 95 percent of 1st-grade teachers surveyed who had made a special-education referral involving a

RR student stated that the RR teacher had provided information that was useful in the referral process. A number of teachers also expressed the belief that the availability of RR data had streamlined the referral process in their school by reducing the steps that precede evaluation. One 1st-grade teacher noted:

We do additional interventions [with the Reading Recovery students who did not discontinue] just to see, but I think we're going to see their time in this whole [referral] process shortened through Reading Recovery. I am very adamant that if they are not making progress in Reading Recovery, there is something more going on.

Teacher Perceptions of Reading Recovery

The randomized experiment will continue to provide critical information about the extent of Reading Recovery's impact on students' literacy achievement. In addition to these experimental data, CPRE is interested in learning more about levels of buy-in to Reading Recovery among school-level stakeholders, particularly teachers. Teacher confidence in RR as an intervention may be an important contributor to level of support for the program in the school and district; level of support for RR has further implications for the possibility of the program becoming a sustained, integrated part of the school's approach to literacy instruction.

By interviewing and surveying RR teachers and 1st-grade teachers, we were able to develop a sense of how they perceive Reading Recovery to be affecting students' classroom habits and behaviors in addition to their literacy skills. In this first year of the evaluation, CPRE noted that RR teachers perceived tremendous progress in their students. Approximately 90 percent of RR teachers surveyed felt that RR had improved the literacy ability of the lowest achieving 15-20 percent of the 1st-grade population at their schools. One teacher noted that students "make so much growth. Whether or not they discontinue in the end, they've made a lot of growth."

The progress RR teachers cite includes growth in overall text reading levels and writing proficiency, but student gains do not end there. RR teachers also report striking improvements in students' oral language skills, attention, and ability to focus. RR teachers frequently commented that Reading Recovery gave struggling students an opportunity to succeed they would not have otherwise had. Comments like the following were typical:

I had three that were very hyperactive, attention deficit, and they would read a word, talk to me, look around the room, and it's just amazing as you do the program with them that their interest in reading develops to where they actually read through the book without stopping to talk or play or do something else.

[My student's] oral language was so limited and it was not an ESL child. It amazed me how strategic she got at reading because she didn't have that

background knowledge; she didn't have the oral language so she had to look at the print and work hard... She did discontinue and she went from, you know, she didn't even know how to make a sentence to write.

Without the support of Reading Recovery, teachers said, students like these might have faced very different prognoses. "Had these kids not gotten Reading Recovery," one teacher said, "they probably would have ended up in special ed because they would just keep falling behind and falling behind."

RR teachers also frequently commented on the ability of RR to help students problem-solve and use strategies while reading. As one teacher noted:

I'm watching children doing a lot of problem solving. Many of us teachers tend to be rescuers and jump in to help children solve problems during reading... I'm done with my first round of children and watched them come along and able to problem-solve on their own I'm thinking about that is really incredible to see. I haven't really seen that before.

RR students' 1st-grade classroom teachers also noticed improvements in their problem-solving capabilities. More than 85 percent of 1st-grade teachers surveyed noted that students who completed RR were able to problem-solve while reading as well as, or better than, the average student in their class. More than 80 percent of classroom teachers also noticed that when compared to other students, students who had completed Reading Recovery were better able to determine when they could solve a problem independently or needed to ask for help.

Teachers also report improvements in students' confidence, and their motivation and overall attitude about both reading and learning in general. One RR teacher recalled a student who was:

...very shy and quiet, and the teacher commented on how she came out of her shell. She goes to the front of the room now and does letter [work]. She puts her letters up on the magnetic board and explains to the kids in front of the class, like she's coming out of her shell. She's confident in herself.

This observation about changes in students' confidence was corroborated by RR students' 1st-grade teachers; 80 percent of those surveyed indicated that, after participating in the intervention, students demonstrated increased confidence in their ability to be successful.

Similarly, 81 percent of 1st-grade teachers agreed that, relative to other students in their classrooms, students who completed Reading Recovery demonstrated an average or above-average interest in learning after completing the intervention. Eighty-eight percent agreed that students who completed the program were more engaged during literacy instruction in their classrooms than they were prior to Reading Recovery. Ninety-two percent saw improvements in RR students' desire to succeed in school after they completed the intervention.

RR teachers' reports on their colleagues' perceptions about students growth were more mixed—a few teachers felt that their colleagues did not understand Reading Recovery, or did not believe it was effective. However, their accounts of feedback received from other teachers were predominantly positive. Echoing a common sentiment, one teacher recalled:

[The first grade teachers] think that it's been amazing for the students that we've served. They see the results from the students, they're seeing how they've grown through Reading Recovery and they're excited about that.

Another teacher commented:

They definitely have seen improvements and are asking me, can I take this one, can I take that one? And I tell them "No, I can only take four of them altogether." They want me to do them in groups. You know I can't!

Approximately 95 percent of RR teachers surveyed reported maintaining regular contact with their RR students' parents, and most used data to keep parents informed of their students' progress. RR teachers reported particular enthusiasm about the program's impact from students' parents. Many teachers had received positive comments from parents about a child's newfound love of reading, or even about a transformed attitude toward school in general. One teacher said:

One of the parents of the [student] that I picked up, she wrote me a letter just out of the blue, and said, thank you for helping [my child] with her reading and pinpointing her—I can't remember how she said it—but pinpointing her, you know, to where she needs extra help. She is reading so much better. She enjoys school now. I mean, you know, I was like "Wow, ok!"

While these perceptions of the wide-ranging impacts of RR on students are subjective, they reflect a level of confidence and buy-in to the program among RR and classroom teachers. The legitimization of a program in the eyes of important stakeholders is key to sustaining that program; comments like these suggest Reading Recovery is gaining legitimacy among 1st-grade teachers - and to some extent, parents - based on their observations of students. As the program is further legitimated and confidence in Reading Recovery grows, teachers and parents could reasonably be expected to become RR advocates, helping to fold RR into the fabric of the school or district's approach to literacy. If a goal is for RR to be sustained and supported to the point of institutionalization in a school or district, teacher buy-in will be crucial. Going forward, we will continue to collect data on teacher perceptions of impact to further explore the connection between buy-in, legitimacy, and the institutionalization of Reading Recovery in school contexts.

Reading Recovery and the Instructional Landscape

Some schools' understandings and practices around literacy instruction were highly aligned and consistent with RR's instructional philosophy, some were highly misaligned, and most schools fell somewhere in between. Interestingly, a majority of the RR teachers interviewed noted that their schools were in transition with respect to literacy instruction, and that in general these schools were transitioning toward closer alignment with RR.

Curricular Alignment

Reading Recovery is grounded in distinctive theoretical understandings—derived directly from the work of Marie Clay—about how children learn to read and write. These theoretical understandings emphasize, among other things, the individuality of knowledge and learning, the complexity and problem-solving nature of the reading process, the relationship between reading and writing, and the importance of authentic reading and writing experiences (Schmitt et al, 2005). CPRE explored the extent to which RR teachers, teacher leaders, and the 1st-grade teachers in schools involved with the scale-up believed their schools' overall approaches to literacy instruction aligned with these understandings and Reading Recovery's approach to instruction. RR teachers' reports point to a continuum of curricular alignment, with a pronounced lack of alignment with Reading Recovery in some schools and close alignment in a few. Most schools, according to RR teachers' reports, were moderately aligned.

In the schools that RR teachers perceived to have literacy approaches poorly aligned with RR, several characteristics were common: a predominance of whole-group literacy instruction; heavy reliance on basal readers; an emphasis on phonics instruction that focused on decoding words in isolation; explicit vocabulary and spelling instruction that was decontextualized from authentic reading experiences; and a lack of focus on writing. Some RR teachers in these schools noted widespread problems with fluency and comprehension, even among students considered to be strong readers, which they attributed to the schools' piecemeal approach to literacy instruction. One teacher commented:

The kids are real non-fluent over all, reading word for word almost, like you could almost be like showing them flash cards. When they come [to Reading Recovery], and even the supposed better readers—ones we didn't service the first time around, so these guys were not even amongst the lowest thirty-six kids—they stop dead and they try to sound it out or they ask. They don't have any strategic activity, you know, anything! You know, look at the picture, or what do you think it could be, or do you see a [familiar word] inside the word. None of this has been taught to them and it's very frustrating.

RR teachers in these schools sometimes described a fundamental disconnect between their own instruction and that of some of their classroom-teacher colleagues. "[The 1st-grade teachers]

don't use the same language that we do," one RR teacher said. "There's still a divide between what they do and what we do."

RR teachers observed misalignment not only with classroom instruction, but at times with RtI systems that included Reading Recovery as one intervention among several. In the survey of RR teachers, 62 percent indicated that the Reading Recovery model aligned with all other 1st-grade literacy interventions at their school. Thirty-two percent stated that Reading Recovery aligned with some but not all other literacy, and five percent reported that RR did not align with any of the other literacy interventions. One RR teacher described the district-mandated intervention that lies at the heart of her school's RtI program:

I see it's very scripted. It's not the same philosophy [as Reading Recovery]. Their assessment is how many words can you read in a minute. It's all about speed. And we've seen that our Reading Recovery kids sometimes will not score well on that and a lot of the time it's because they're not as fast. Or like, a kid will qualify [for this other program], but we look at that list as well when we're thinking about who we need to pull for Observation Survey, and on the Reading Recovery they'll do a lot better than they did in that. [The other one is] just not a consistent assessment.

At the other end of the continuum were a small number of schools whose instructional approaches were described by RR teachers as strongly aligned with the instructional approach and philosophy of Reading Recovery. These schools often used Fountas & Pinnell as the primary literacy program school-wide and, in many cases, were not first-time RR implementers; some had been implementing Reading Recovery for years and regarded the i3 grant as a chance to increase their capacity and serve more students by training additional teachers.

In some of these schools, administrators and classroom teachers had significant exposure to Reading Recovery or, in some cases, were RR-trained themselves. In other cases, the schools were participating in a comprehensive literacy program like the Partnerships in Comprehensive Literacy Model developed at the University of Arkansas at Little Rock. In a number of closely aligned schools, individuals had been trained, or were currently being trained, in Fountas & Pinnell's Leveled Literacy Intervention (LLI), a program designed to bring RR teaching philosophies to students through small groups rather than one-on-one instruction. One teacher described:

The district did try to pick [the small-group intervention] that was the closest match to [RR's] philosophy, which was the Leveled Literacy Instruction [sic] by Fountas & Pinnell—the LLI. Probably out of everything that was out there, that intervention program matches more of our philosophy than anything. It was the language that teachers knew, with leveled texts and, you know, going from whole-part-whole and working with words and things like that. So, a lot of what

we were doing balanced-literacy-wise had been built on the Reading Recovery philosophies and structures.

In most schools, however, RR teachers reported that the literacy curriculum was undergoing a transition, generally moving toward increased alignment with the approach and philosophy of Reading Recovery. In these schools, improving literacy instruction was considered a key institutional priority, and changes to the literacy programs were being discussed and/or actively undertaken. One teacher reflected:

In the past most of our reading instruction was done whole-group and they still do some whole group, but we've seen a change in more teachers spending more time in small-group reading instruction and less time in whole-group. You know, giving children more opportunities to read at their level. Plus just the whole idea of doing things differently, trying something different. Some of the phrasing that we use in Reading Recovery, you know, we've started to see our classroom teachers using [it]. Things like, you know, when you write [there is] that self-monitoring aspect. So I would say it has changed the way we look at our literacy instruction.

As in this example, curricular transitions frequently involved a shift away from whole-group instruction toward guided reading or other small-group arrangements, or a shift toward the use of workshop-model instruction in reading and/or writing. A number of the schools were newly implementing Boushey and Moser's CAFE (2009) system, often in combination with the related Daily 5 model. As one RR teacher explained, however, in most cases the transition was an on-going process:

We are definitely working to have a common vision but I would say that it's definitely not there right now. We've just started implementing readers' and writers' workshops, and some of the teachers still heavily use our basal reader. Some of the teachers have had prior [workshop-model] training, which really helps them feel comfortable teaching the readers' and writers' workshop model without so much guidance from the basal. So overall we're definitely moving towards readers' and writers' workshops, but we're not really consistent yet across all grade levels.

Interestingly, approximately 89 percent of 1st-grade teachers surveyed indicated that RR was aligned with their school's overall approach to literacy. This suggests that the curricular disconnect observed by RR was not evident to some 1st-grade teachers. Also notable was that, in some cases, RR teachers did not know what approaches or curricula the classroom teachers in their schools primarily used for literacy instruction.

CPRE is interested in the relationship between these alignment patterns and the adoption of

RR—in particular, the possible tendency of schools involved with the scale-up to be transitioning toward a more Reading Recovery-aligned literacy model. It remains to be seen how important curricular alignment is in terms of Reading Recovery students' short- or long-term outcomes: Does consistency in instruction impact results? These issues will be aspects of future study.

Conclusion: Changing students. Changing teachers. Changing schools?

These early results from the external evaluation of the Reading Recovery i3 scale-up suggest remarkable success for Reading Recovery. First, there was considerable success in scaling up the program with reasonable fidelity; the recruitment of teachers and schools came close to meeting project goals, and trends in recruitment suggest continued progress toward these goals. Second, RR training and implementation were done with high fidelity in schools participating in the scale-up, meeting the vast majority of the standards defined by Reading Recovery. Ultimately, there were large positive effects of Reading Recovery on student literacy performance overall, and these positive effects were also large for ELL students and students in rural schools.

Impacts on Student Reading Performance

The estimated effect of Reading Recovery on students' ITBS Total Reading Scores was .68 standard deviations relative to the study sample, and .47 standard deviations relative to the national population of first graders. These standardized effect sizes are large relative to typical effect sizes found in educational evaluations. For example, the impacts of Reading Recovery are up to 5.7 times larger than the average effects of Title I programs reviewed by Borman and D'Agostino (1996). Gains in percentile rank scores were also large, with treatment students outperforming control students by up to 20 percentile points. When compared to typical gains of first graders on the ITBS tests, Reading Total scale scores for the average student in the U.S. are expected to increase from 133 to 144 points from the start of first grade through the fifth month (i.e., the period during which the treatment students received RR instruction). This increase of 11 points over a 5-month period suggests that the additional gains of 4.2 points experienced by Reading Recovery students is equivalent to an additional 1.9 months of learning. Alternatively, the additional 4.2 points translates to a growth rate that is 38 percent greater than the national average growth rate for beginning first graders.

When broken down by ITBS subscale, effect estimates for Reading Recovery were similar on both the ITBS Reading Words subscale (i.e., .60 SD) and the ITBS Reading Comprehension subscale (i.e., .61 SD). Effect estimates for ELLs and students in rural schools were also similar (i.e., .51 and .59 SD, respectively).

The highly significant variance components for random effects in the HLM models of impacts on ITBS scores suggests that the magnitude of the impact estimates of Reading recovery varies substantially across schools. While none of the schools in the 2011-12 randomized experiment had negative impact estimates, and the modal impact estimate is .70 standard deviations, a few schools have very large positive impact estimates in excess of one full standard deviation. Future analyses will pool data from multiple cohorts of schools in order to explore school-level predictors of variation in impacts.

Program Oversight and Recruitment

The i3 scale-up grant required significant changes in responsibilities within the network of Reading Recovery UTCs. Most significantly, OSU, as the primary recipient of the i3 grant, was charged with a) overseeing recruitment and implementation across the network of 19 UTCs, and b) managing the overall budget for the project and dispensing funds through subcontracts to the 19 UTCs. Historically, UTCs have operated very independently; under the i3 project, it was a clear priority for OSU to support UTC autonomy while being responsive to the issues and concerns they reported. It was also important to the i3 Office to include UTC directors in discussions and decisions regarding large-scale changes in strategies or procedures for Reading Recovery implementation.

The recruitment goals of the i3 scale-up were to recruit 1,470 teachers in 588 schools, serving 17,640 students with Reading Recovery. Recruitment during the 2010-11 year was well below target, largely due to the late start of the project and the restrictive eligibility criteria for schools. However, recruitment was more successful during 2011-12, and at the end of the first two years of the scale-up the RR network had successfully recruited and trained 1,121 i3 RR teachers in 847 i3 schools, who provided RR instruction to 15,525 students. Overall, recruitment in the first two years of the scale-up was 24 percent below target for teachers, 44 percent above target for schools, and 12 percent below target for students.

Success in recruitment is largely attributed to the efforts of teacher leaders and UTC directors; individuals in both roles spent tremendous amounts of time reaching out to school principals and district staff. While each region had unique hurdles to overcome, the single biggest impediment to recruitment was a direct result of the economic downturn. District leaders and school principals were very reluctant to hire new staff and implement a new intervention while simultaneously facing budget cuts. As such, recruitment efforts often entailed developing creative ways to leverage existing staff and resources to support Reading Recovery in each site.

Training New RR Teachers

A critical element to the faithful implementation of Reading Recovery is the intensive training RR teachers receive during their first year. The behind-the-glass sessions were reported by RR teachers to be one of the most valuable aspects of their training experience. However, not all

Conclusion: Changing students. Changing teachers. Changing schools?

teachers are getting as much experience with behind-the-glass training as they should. While 87 percent of new RR teachers reported teaching behind the glass at least three times (as required by the RR Standards and Guidelines), the vast majority reported observing other teachers teach behind the glass fewer than the 36 required times. This may have been the result of smaller-than-usual training classes, or the great distances some RR teachers-in-training had to travel to attend behind-the-glass sessions.

RR teachers appreciated the immediate feedback and support given by teacher leaders during their site visits. RR teachers viewed teacher leaders as important and accessible sources of information about literacy instruction, and as supports for their own professional growth. RR teachers felt that the one-on-one interactions with the teacher leaders helped them make concrete changes to their RR teaching. In general, RR teachers felt that their training experience and work with their teacher leader prepared them for RR implementation and teaching RR students. Many RR teachers reported that their RR training was transformative in terms of their own instruction and understanding about literacy. Their new skills in assessment and observation, and new instructional strategies, were useful not only for their RR lessons but also in their other instructional roles.

Implementation Fidelity

School-level implementation of RR was, in most respects, faithful to the Reading Recovery Standards and Guidelines. RR teachers typically provided one-on-one lessons to RR students for half the day and fulfilled other roles in their schools in the remainder. There was strong fidelity to standards in the execution of RR lessons, but less fidelity to the requirements to formally document every lesson. RR teachers encountered challenges to adhering to the schedule of daily one-on-one lessons with each student, however most found ways to surmount those obstacles. Data from surveys, activity logs, interviews, and observations reveal that the job of an RR teacher is extremely demanding, and entails a very busy schedule with many long days.

RR teachers reported working very closely with teacher leaders and 1st-grade teachers, and identified teacher leaders as critical supports. Teacher leaders' roles and responsibilities varied widely from site to site. Some had significant influence over decisions affecting RR implementation, or even over a school's general approach to literacy instruction; others seemed to have very little influence on school-level operations outside the delivery of RR lessons. RR teachers communicated frequently with 1st-grade teachers. Most communications were about RR student progress, though some RR teachers shared instructional strategies with their colleagues. Many RR teachers participated in their schools' Rtl processes, but a minority held explicit instructional leadership positions at their school.

When identifying students eligible for RR intervention, schools and districts varied in their selection processes. While most schools used scores on the OS to identify low-performing students, the students ultimately selected to receive RR were not always the lowest performing

students in the first grade. In some schools, students were deemed ineligible for RR if they had an IEP or were chronically absent. By some accounts, these exclusions are not consistent with Marie Clay's intent for RR to reach all struggling readers—this was the most significant deviation from the RR Standards and Guidelines revealed in this evaluation. However, the student data generated through RR screening and instruction were often given weight in decisions about special education, suggesting RR can help students receive the supports they need even if they are not selected for RR lessons.

Reading Recovery as a Potential Lever for School Change

At this point, it is not clear what impact Reading Recovery has on schools' and systems' cultures of literacy instruction. The extent to which RR can influence "school change" seems to be dependent on factors that are often beyond the control of RR teachers and teacher leaders as they do their advocacy work, such as school climate, local leadership, and logistical issues. For example, the RR model includes the expectation that the teacher leader will play a leading role as a change agent, influencing school and district policy and practice. However, at the school level, most RR teachers were not aware of their teacher leaders having enough contact with others in the building to play such a role. It seems that the ability of teacher leaders to act as change agents varies widely, and that in some cases RR teachers are assuming this role.

While the literacy curricula and interventions are closely aligned with RR in some schools and markedly misaligned in others, most schools fall somewhere in between. A majority of RR teachers reported that the literacy programming at their school was in transition, generally toward closer alignment with RR. The extent to which the implementation of RR is catalyzing these changes, and the roles played by various stakeholders in this change process, will be topics of study in future years of this evaluation.

While it is still too early to say whether RR is inspiring systemic organizational changes, two things are clear. First, Reading Recovery is positively affecting student literacy outcomes, and second, RR teachers are sharing their new knowledge, strategies, and information with other teachers in their schools. There is much to learn through the remainder of this evaluation, but these conditions could reasonably be viewed as precursors to systemic change. It is the goal of the longitudinal components of this evaluation to provide insight into what school-wide changes are associated with RR implementation, and how the form of implementation influences such changes.

Appendix A: Statistical Model for Impacts of Reading Scores

The mathematical form of the primary impact model for the RCT study is:

$$Y_{ijk} = \beta_0 + \beta_1(\text{Pretest}) + \beta_2(\text{Trt}) + \gamma_j + \alpha_k + \varphi_k(\text{Trt}) + \varepsilon_{ijk}$$

With:

$$\gamma_j \sim N(0, \omega^2), \begin{pmatrix} \alpha_k \\ \varphi_k \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2 & \rho \\ \rho & \xi^2 \end{pmatrix} \right), \text{ and } \varepsilon_{ijk} \sim N(0, \sigma^2)$$

Where:

- Y_{ijk} is the posttest outcome score for student i from pair j in school k
- β_0 is the model intercept
- β_1 is the slope coefficient for the pretest covariate (i.e., from the Observation Survey)
- β_2 is the overall treatment effect
- Trt is the treatment assignment indicator, with 1=treatment and 0=control
- γ_j is the random intercept associated with matched pair j , with variance ω^2
- α_k is the random intercept associated with school k , with variance τ^2
- φ_k is the random treatment effect associated with school k , with variance ξ^2
- ρ is the correlation between random school intercepts and treatment effects
- ε_{ijk} is the student-level residual, with variance σ^2

References

- Ashdown, J., & Simic, O. (2000) Is early literacy intervention effective for English language learners? Evidence from Reading Recovery. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, 5(1), 27-42.
- Bloom, H. S., Raudenbush, S. W., & Weiss, M. (2012). Estimating variation in program impacts: Theory, practice and applications. Paper under review.
- Boushey, G., & Moser, J. (2009). *The CAFE Book: Engaging All Students in Daily Literary Assessment and Instruction*. Stenhouse Publishers.
- Briggs, C., & Young, B. (2003). Does Reading Recovery work in Kansas? A retrospective longitudinal study of sustained effects. *The Journal of Reading Recovery*, 3(1), 59-64.
- Brown, W., Denton, E., Kelly, P., & Neal, J. (1999) Reading Recovery effectiveness: A five-year success story in San Luis Coastal Unified School District. ERS Spectrum: *Journal of School Research and Information*, 17(1), 3-12.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An experimental evaluation of Reading Recovery. *Reading Research Quarterly*, 30, 240-263.
- Clay, M. (1987). Implementing Reading Recovery: Systemic Adaptations to an Education Innovation. *New Zealand Journal of Educational Studies*, 22(1), 35-57.
- Clay, M. (1991). *Becoming Literate: The Construction of Inner Control*. Portsmouth, NH: Heinemann.
- Clay, M. (2005a). An Observation Survey of Early Literacy Achievement. Portsmouth, NH: Heinemann. Clay, M. (2005b). *Literacy lessons designed for individuals: Part one*. Portsmouth, NH: Heinemann.
- Clay, M. (2005c). *Literacy lessons designed for individuals: Part two*. Portsmouth, NH: Heinemann.
- D'Agostino, J., & Murphy, J. (2004). A meta-analysis of Reading Recovery in United States' schools. *Educational Evaluation and Policy Analysis*, 26(1), 23-38.
- Denton, C.A. (1997). An evaluation of an implementation of the Reading Recovery program. Paper presented at the annual meeting of the Southwest Educational Research Association. Austin, TX.
- Denton, C.A., Ciancio, D.J., & Fletcher, J.M. (2006). Validity, reliability, and utility of the Observation Survey of Early Literacy Achievement. *Reading Research Quarterly*, 41(1), 8-34.

Elmore, R. (2004). *School Reform from the Inside Out: Policy, Practice, and Performance*. Cambridge, MA: Harvard Education Press.

Elmore, R. (2007). Local School Districts and Instructional Improvement. In *The Keys to Effective Schools: Educational Reform as Continuous Improvement*. Thousand Oaks, CA: Corwin Press.

Elmore, R. (2009). *Instructional Rounds in Education: A Network Approach to Improving Teaching and Learning*. Cambridge, MA: Harvard Education Press.

Engelhard, G. (2007). Review of the Iowa Tests of Basic Skills. In Spies, R.A., Plake, B.S., Geisinger, K.F., Carlson, J.F. (Eds.), *17 Mental Measurements Yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Fountas, I. C., & Pinnell, G. S. (2001). *Guiding Readers and Writers, Grades 3-6: Teaching Comprehension, Genre, and Content Literacy*. Westport, CT: Heinemann Education Books.

Fullan, M. (1993). *Change Forces: Probing the Depths of Educational Reform*. Levittown, PA: Taylor & Francis.

Hallinger, P. (2005, April). *Instructional leadership: How has the model evolved and what have we learned*. Paper presented at annual meeting of the American Educational Research Association, Montreal, Canada.

Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools, 4*(3), 221-239.

Harvey, Stephanie, and Anne Goudvis. *The Comprehension Toolkit*. Firsthand. Available from: Heinemann. PO Box 6926, Portsmouth, NH 03802-6926, 2005.

Heck, R. H. (1993). School context, principal leadership, and achievement: The case of secondary schools in Singapore. *The Urban Review, 25*(2), 151-166.

Herman, R. (2009). Professional development in Reading Recovery. Retrieved from http://www.readingrecovery.org/pdf/professional_development/PD_Overview-09.pdf

Hoover, H.D., Dunbar, S.B., Frisbie, D.A., Oberley, K.R., Ordman, V.L., Naylor, R.J., Bray, G.B., Lewis, J.C., Qualls, A.L., Mengeling, M.A., & Shannon, G.P. (2003). *The Iowa tests: Guide to research and development*. Itasca, Illinois: Riverside Publishing.

Hoover, H.D., Dunbar, S.B., Frisbie, D.A., Oberley, K.R., Bray, G.B., Naylor, R.J., Lewis, J.C., Ordman, V.L., & Qualls, A.L. (2006). *The Iowa tests: 2005 norms and score conversions*. Iowa City: University of Iowa.

Iversen, S., & Tunmer, W. (1993) Phonological processing skills and the Reading Recovery Program. *Journal of Educational Psychology, 80*(4), 437-447.

- Johnston, P.A., Allington, R.L., & Afflerback, P. (1985). The congruence of classroom and remedial reading instruction. *The Elementary School Journal*, 85, 465-478.
- Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., Schulte, A., & Olson, R. (2001). Rethinking learning disabilities. In Finn, C. E., Rotherham, A. J., and Hokanson, C. R. (Eds.), *Rethinking Special Education For a New Century*. Washington, D.C.: Progressive Policy Institute and Thomas B. Fordham Foundation.
- Morrow, L.M. (1993). *Literacy Development in the early years: Helping Children read and write*. Needham Heights, MA: Allyn & Bacon.
- National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development, National Institutes of Health.
- Neal, J., & Kelly, P. (1999). The success of Reading Recovery for English Language Learners and Descubriendo la Lectura for bilingual students in California. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, 4(2), 81-10
- Pikulski, J. (1994). Preventing reading failure: A review of five effective programs. *The Reading Teacher*, 48, 30-39.
- Pinnell, G. (1989). Reading Recovery: Helping at-risk children learn to read. *The Elementary School Journal*, 90, 161-183.
- Pinnell, G., Lyons, C., DeFord, D., Bryk, A., & Seltzer, N. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 29, 8-39.
- Quay, L., Steele, D., Johnson, C., & Hortman, W. (2001). Children's achievement and personal and social development in a first-year Reading Recovery program with teachers in training. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, 5(2), 7-25.
- Reading Recovery Council of North America. (2009). *Standards and Guidelines of Reading Recovery in the United States* (5th ed.). Columbus, OH: Author. Available: http://www.readingrecovery.org/pdf/implementation/Standards_Guidelines-09_Full_Version.pdf
- Rodgers, E., Gómez-Bellengé, F., Wang, C., & Schultz, M. (2005). *Predicting the literacy achievement of struggling readers: Does intervening early make a difference?* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Rodgers, E., Wang, C., & Gómez-Bellengé, F. (2004). *Closing the literacy achievement gap with early intervention*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Rowe, K. (1995) Factors affecting students' progress in reading: Key findings from a longitudinal study. *Literacy Teaching and Learning: An International Journal of Early Literacy*, 1(2), 57-110.

Ruhe, V., & Moore, P. (2005) The Impact of Reading Recovery on Later Achievement in Reading and Writing. *ERS spectrum* 23(1), 20-30.

Schmitt, M.C., & Gregory, A. E. (2005). The impact of an early literacy intervention: Where are the children now?. *Literacy Teaching and Learning: An International Journal of Early Literacy*, 10, 1-20.

Schwartz, R. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, 97, 257-267.

Smith-Burke, M.T., Pinnell, G.S., Jackson, M., Wey, S., Askew, B.J., & Hambright-Brown, E. Reading. *A principal's guide to Reading Recovery*. Recovery Council of North America. (2002) Columbus, OH.

Strickland, D.S. (2002). The Importance of Effective Early Intervention. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about reading instruction* (Third Edition) (pp. 69-86). Newark, DE: International Reading Association.

Supovitz, J. A., Sirinides, P., & May, H. (2010) How principals and peers influence teaching and learning. *Journal of Education Administration Quarterly*, 46(1), 31-56.



A COLLABORATIVE PUBLICATION BETWEEN



Center for Research in
Education & Social Policy