

CPRE

CONSORTIUM *for* POLICY RESEARCH *in* EDUCATION
UNIVERSITY OF PENNSYLVANIA

UNIVERSITY OF
DELAWARE

Center for Research in
Education & Social Policy

READING RECOVERY: An Evaluation of the Four-Year i3 Scale-Up

A RESEARCH REPORT

Henry May

Philip Sirinides

Abigail Gray

Heather Goldsworthy

MARCH 2016



About the Consortium for Policy Research in Education (CPRE)

Founded in 1985, the Consortium for Policy Research in Education (CPRE) brings 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 research is peer-reviewed and open access.

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.

© Consortium for Policy Research in Education, University of Pennsylvania

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.

This CPRE research report was externally and internally reviewed for quality and accuracy.

Suggested Citation:

May, H., Sirinides, P., Gray, A., & Goldsworthy, H. (2016). *Reading Recovery: An Evaluation of the Four-Year i3 Scale-Up*. Philadelphia: Consortium for Policy Research in Education.

Acknowledgements

The research described in this report would not have been possible without the collaboration and support of a number of individuals. The CPRE/CRESP research team extends its gratitude to Jerry D'Agostino and Emily Rodgers at The Ohio State University, the PI and Co-PI of the i3 scale-up grant to Reading Recovery, who facilitated many of our data collection activities and helped ensure a successful project. In addition, we would like to extend special thanks to Mike Puma of Chesapeake Research Associates and Rebecca Maynard at the University of Pennsylvania for lending their expertise and support throughout this project.

This evaluation benefitted from the input of a number of contributors. These include the graduate research assistants who joined the project team over the years, particularly Michael Armijo, Helen Anderson-Clark and Andrew Schiera, who made significant contributions to the work presented here. Other research assistants involved in the project include Michael Moore, Namrata Tognatta, Rachel Skrlac Lo, Charlotte Mecozzi, and Horatio Blackman. Also, we would be remiss in failing to note the contributions of Yolanda Green, CPRE's business manager, and Jessica Gillespie and Toscha Blalock, who served as project managers. Amy Hillier and Shrobona Karkun of the University of Pennsylvania's School of Design lent us their expertise in GIS mapping. We also thank Jeff Brymer-Bashore, Director of the International Data Evaluation Center for his assistance in accessing program data for this study.

Additionally, we are grateful to Sharon Ravitch, Andy Porter, and Bob Boruch of the University of Pennsylvania's Graduate School of Education for their careful and thoughtful reviews of our reports, and to Jonathan Supovitz and the entire staffs of CPRE and CRESP for their comradery and support.

We would also like to thank Marsha Studebaker and other staff of the Reading Recovery Council of North America (RRCNA), who supported us with resources and information throughout the evaluation, and the veteran Reading Recovery teacher leaders and trainers who formed an advisory group for our work: Carol Lyons, Noel Jones, Synda Buhacevich, and Maribeth Schmitt. Last but certainly not least, we are indebted to the hundreds of practicing Reading Recovery professionals who taught us about the program and graciously participated in a multitude of data collection activities. We greatly admire your passion and commitment.

Henry May
Phil Sirinides
Abby Gray
March 2016

Table of Contents

Executive Summary	1
Study Goals and Design	1
Key Findings	2
Conclusions	4
Introduction	5
Brief Overview of Reading Recovery	6
The i3 Evaluation: Goals and Design	7
Structure and Content of the Final Report	7
Chapter 1 Taking Reading Recovery to Scale: Processes, Challenges, and Outcomes	9
Studying the Scale-Up: Research Methods	10
Findings: Recruitment Targets	10
Findings: Recruitment Processes and Strategies	12
Findings: The Evolving Challenges of Scaling Up	19
Conclusions: Learning and Adapting through Scaling Up	25
Chapter 2 The Immediate Impacts of Reading Recovery: Results of the Four-Year RCT Study	27
Research Objectives	27
Sample Selection and Assignment	28
Measures	33
Additional Data Collection	36
Analyses	37
Findings	39
Conclusions	44
Chapter 3 Sustained Impacts of Reading Recovery: A Regression Discontinuity Study from the 2011-12 Cohort	45
Methods	46
Findings	53
Conclusions	67

Chapter 4 Reading Recovery Implementation Fidelity:	
Scale-up Successes and Challenges	69
Defining Implementation Fidelity	70
Evaluating Implementation Fidelity	70
Qualitative Findings	79
Discussion.....	82
Chapter 5 Lesson-Level Implementation:	
Instructional Strength in Reading Recovery	83
Research Questions: Understanding Implementation in Depth.....	84
Research Methods: An Evolving Mixed-Methods Design.....	84
Data	85
Findings: Instructional Strength in Reading Recovery	90
Discussion.....	105
Chapter 6 School-Level Implementation of Reading Recovery:	
A Cross-Case Analysis	107
Research Methods	109
Findings	115
Discussion.....	137
Conclusions.....	145
Conclusion Learning from Scale-Up:	
Lessons from the four-year i3 evaluation of Reading Recovery	147
Appendices	150
References	167

Executive Summary

This report presents the final results of a four-year independent external evaluation of the impacts and implementation of the scale-up of Reading Recovery, a literacy intervention targeting struggling 1st-grade students. The evaluation was conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania in collaboration with the Center for Research in Education and Social Policy (CRESP) at the University of Delaware.

The evaluation was funded by an Investing in Innovation (i3) grant to The Ohio State University (OSU) from the U.S. Department of Education's Office of Innovation and Improvement. The grant, which was awarded in 2010, totaled \$45 million, with an additional \$10.1 million raised from private sources, to cover the expansion of Reading Recovery around the U.S. Of this, \$5 million was earmarked for the completion of the independent evaluation of the scale-up effort between 2011 and 2015.

Reading Recovery is an intensive intervention targeting struggling 1st-grade readers. The program was developed in the 1970s and 80s by Marie Clay, a developmental psychologist and professor at the University of Auckland whose theories about how children become literate provide the foundation for Reading Recovery's approach (Clay, 1991; 2005). Reading Recovery is delivered to students through a 12- to 20-week series of individual, daily, 30-minute lessons provided by a Reading Recovery-trained teacher. The overarching goal of the program is to intervene early, before young students' reading difficulties become lifelong obstacles.

STUDY GOALS AND DESIGN

CPRE/CRESP's evaluation of Reading Recovery 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. The primary goals of the evaluation are to:

1. Provide experimental evidence of the short- and long-term impacts of Reading Recovery on student learning in schools that are part of the i3 scale-up; and
2. Assess the implementation of Reading Recovery under the i3 grant, including fidelity to the program model and progress toward the scale-up goals.

KEY FINDINGS

The impact evaluation includes a multi-site randomized controlled trial (RCT) for estimating immediate impacts, a regression discontinuity study (RD) for estimating long-term impacts, and an implementation study for assessing fidelity of implementation and exploring program implementation in depth. The RCT includes nearly 7,000 randomized students in more than 1,200 schools over four years. The RD study measures Reading Recovery's impacts at the end of first grade and in third grade, and replicates the RCT's immediate post-treatment findings in a separate sample of students. The implementation study involves a combination of qualitative and quantitative research executed on a large scale over the same four-year timeframe.

KEY FINDINGS

The evaluation's key findings pertain to the following topics:

1. Scale-Up Processes, Challenges, and Outcomes

CPRE's evaluation of the i3 scale-up assessed Reading Recovery's progress toward the goals of training 3,675 new Reading Recovery teachers; providing one-to-one Reading Recovery lessons to an additional 67,264 students; and, delivering other instruction—generally classroom or small-group instruction—to 302,688 more students via teachers trained in Reading Recovery during the scale-up. We found that the scale-up surpassed its goals for the number of Reading Recovery teachers trained (3,747, or 102 percent of the scale-up goal), and the number of non-Reading Recovery students served by those teachers (325,458, or 108 percent of the scale-up goal). In the third area, students served with one-to-one Reading Recovery lessons provided by teachers trained with i3 funds, the project achieved 92 percent of its goal, with 61,992 students. As these results indicate, the effort to expand Reading Recovery under the i3 initiative was highly successful.

Our research into the scale-up process revealed that Reading Recovery trainers at the 19 universities that partnered with OSU to implement the scale-up nationally played a critical role in recruiting teachers and schools to participate, as did Reading Recovery teacher leaders across the country. The particular strategies trainers and teacher leaders used in recruitment were varied and context-driven, though the major obstacles—limited staff capacity, a difficult fiscal climate, and the challenges of scaling up sustainably and with fidelity to the Reading Recovery model—were consistent across most regions.

2. Immediate Impacts of Reading Recovery

The four-year, multi-site RCT examined Reading Recovery's impacts on students' scores on the Iowa Test of Basic Skills (ITBS) Reading Total assessment as well as the ITBS Reading Comprehension and Reading Words subtests, and on the Observation Survey of Early Literacy Assessment (OS). Within schools, students were matched into pairs based on pretest scores and randomly assigned, within matched pairs, to treatment and control. Students in the treatment group received 12- to 20-weeks of daily, one-to-one Reading Recovery lessons provided by a trained teacher as a supplement to regular classroom

literacy instruction. Students in the control group received regular classroom literacy instruction as well as any interventions normally provided to low-performing 1st-grade readers in their schools.

The RCT revealed medium to large impacts across all outcome measures. Effect sizes on the ITBS and its subscales at the end of 12- to 20-weeks of treatment ranged between 0.30 and 0.42 standard deviations. For the ITBS Total Reading battery, this effect size translates to a gain of +18 percentage points in the treatment group, as compared with control students. The growth rate we observed in students who participated in Reading Recovery over approximately a five-month period was 131 percent of the national average rate for 1st-grade students. Moreover, these results were similar in two subgroups of interest to the i3 program: English Language Learners and students in rural schools.

With 6,888 student participants, the RCT of short-term impacts in the scale-up schools is among the largest such studies ever conducted. Its rigorous design and large sample offer strong evidence of the effects of Reading Recovery on the short-term progress of struggling students.

3. Sustained Impacts of Reading Recovery

The evaluation's RD study focused on students who were in first grade in 2011-2012 and had reached third grade by the end of the i3 study. It assessed these students' reading achievement at the end of first grade, roughly 5 months after they completed Reading Recovery, and again in third grade. Findings at the end of first grade revealed an effect of +12 points on students' OS scores, for an intent to treat effect size of .30 standard deviations.

The RD analysis of impacts on 3rd-grade reading achievement used state test scores in reading as the outcome measure. While the impact estimate produced by this analysis was not significant, the available data were far too sparse to produce a conclusive finding. Future research is planned to better address Reading Recovery's impacts on 3rd-grade test score.

The RD study also replicated the findings of the RCT. Its baseline contrast was estimated simultaneously with the Year One RCT in a separate random sample of i3 schools, and it found similarly large impacts. Based on a sample of 3,522 students in 274 schools, the impacts estimated immediately post intervention show an effect of +29.7 points on the OS. This equates to an intent to treat effect size of .68 standard deviations.

4. Implementation Fidelity

CPRE/CRESP assessed fidelity to the Reading Recovery program model in the schools involved in the scale-up over the four years of the evaluation. Fidelity data were collected annually from Reading Recovery teachers and teacher leaders, and analyzed for consistency with the *Standards and Guidelines of Reading Recovery in the United States, 6th Edition* (Reading Recovery Council of North America, 2012). Specifically, we assessed

CONCLUSIONS

fidelity to the program model in four areas: staff background and selection; teacher leader and site capacity; Reading Recovery teacher training and professional development; and one-to-one Reading Recovery lessons.

Our analysis revealed strong fidelity to the program model in all of these areas and all years of the scale-up. This suggests that the intervention was delivered as designed to the students in the scale-up, and that teachers delivering Reading Recovery lessons were properly trained. In total, the results of the fidelity analysis support the validity of our impact findings.

5. Lesson- and School-Level Implementation

We also studied aspects of Reading Recovery's implementation that are not codified in program documents but that vary from school to school and impact program quality. These aspects are manifested both at the level of the one-to-one Reading Recovery lesson and at the level of the school as a whole. First, at the lesson level, we explore instructional strength in terms of Reading Recovery teachers' *deliberateness* and *instructional dexterity*. These related but distinct dimensions of Reading Recovery teachers' instruction are enacted through specific aspects of their practice. We also discuss school- and district-level administrative support as important facilitators of instructional strength in Reading Recovery.

Second, we present the results of 23 field-based case studies focused on schools implementing Reading Recovery under the scale-up, and identify four schemas of Reading Recovery implementation at the school level: isolation, obstruction, endorsement, and integration. Patterns of communication about Reading Recovery within the school and the principal's engagement with the program are the key differentiators of these four schemas.

CONCLUSIONS

The i3-funded scale-up of Reading Recovery was one of the most ambitious and well-documented expansions of an instructional program in U.S. history, and it was highly successful. The independent evaluation investigating its impacts and execution is one of the most comprehensive evaluations ever implemented in the field of education. It represents a contribution to the growing body of research on the conditions for and impacts of scaling up instructional programs.

Introduction

This report presents the results of a four-year independent external evaluation of the impacts and implementation of the scale-up of Reading Recovery, a literacy intervention targeting struggling 1st-grade students. The evaluation was conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania and the Center for Research in Education and Social Policy (CRESP) at the University of Delaware.

The CPRE/CRESP team's evaluation was funded by an Investing in Innovation (i3) grant to The Ohio State University (OSU) from the U.S. Department of Education's Office of Innovation and Improvement. The grant (Principal Investigator: J. D'Agostino; Co-Principal Investigator: E. Rodgers), which was awarded in 2010, totaled \$45 million to cover the expansion of Reading Recovery around the U.S. An additional \$10.1 million was raised from the private sector to support the scale-up. Of this total, \$5 million was earmarked for the completion of this independent evaluation of the scale-up effort between 2011 and 2015.

The i3 award to OSU—the U.S. seat of Reading Recovery—was made partly on the basis of prior evidence of the intervention's impacts on student achievement in literacy. Contributing further evidence of Reading Recovery's effectiveness through a large, rigorous impact study was a primary goal of the evaluation. Additional goals included documenting implementation of Reading Recovery under the i3 scale-up, and assessing OSU's success at achieving its expansion goals by training new Reading Recovery teachers and providing instruction to additional students.

In this chapter, we provide a brief introduction to Reading Recovery; describe the evaluation's overall design and purposes; and summarize the key findings of CPRE's two previous reports on this research (May et al., 2013; 2014). In addition, we discuss the structure of this final report, with a brief overview of the content and purpose of each of the chapters to come.

BRIEF OVERVIEW OF READING RECOVERY¹

Reading Recovery is an intensive intervention targeting struggling 1st-grade readers. The program was developed in the 1970s and 80s by Marie Clay, a developmental psychologist and professor at the University of Auckland whose theories about how children become literate provide the foundation for Reading Recovery's approach (Clay, 1991; 2005).

Reading Recovery is delivered to students through a 12- to 20-week series of individual, daily, 30-minute lessons provided by a Reading Recovery-trained teacher. The overarching goal of the program is to intervene early, before young students' reading difficulties become lifelong obstacles. Its instructional objective is to help students develop a set of self-regulated literacy strategies that govern the use of meaning, structure, letter-sound relationships, and visual cues in reading and writing (Clay, 1991; 2005; May et al., 2013).

Teachers who deliver the Reading Recovery intervention must complete an intensive, year-long graduate-level training course taught by a literacy coach called a teacher leader. Through this training and ongoing professional development, teachers develop expertise at analyzing students' literacy behaviors, identifying learning needs, and delivering responsive instruction. Reading Recovery theory asserts that the role of the teacher is to identify students' strengths and needs, and to facilitate his or her learning by providing appropriate opportunities to acquire and practice new skills. Other key roles in Reading Recovery implementation include teacher leaders, who train and support Reading Recovery teachers, and Reading Recovery trainers, who are faculty members at OSU and its 19 partner universities around the U.S. Each partner university operates a University Training Center (UTC) that serves as a regional hub of Reading Recovery training and activity. University-based trainers support and train teacher leaders, and oversee the program's operations within their areas.

The implementation of Reading Recovery is guided by the *Standards and Guidelines of Reading Recovery in the United States*, 6th Edition (Reading Recovery Council of North America, 2012). The *Standards and Guidelines* outline the key activities involved with the program's implementation.

While research on Reading Recovery has produced mixed findings (Iversen, Tunmer & Chapman, 2005; Rodgers, Gómez-Bellengé, Wang & Schultz, 2005; Farrall, 2006), a number of studies point to positive impacts on student learning (D'Agostino & Harmey, 2016; Ashdown & Simic, 2000; Allington, 2005; Center et al, 1995; D'Agostino & Murphy, 2004; Schwartz, 2005; Quay et al, 2001; Rodgers et al, 2004, 2005). Conducting a current, highly rigorous assessment of Reading Recovery's impacts was a key goal of the evaluation.

1.A detailed description of Reading Recovery is presented in the evaluation's Year One report (May et al., 2013).

THE I3 EVALUATION: GOALS AND DESIGN

CPRE/CRESP’s evaluation of Reading Recovery 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. The primary goals of the evaluation are:

1. to provide experimental evidence of the short- and long-term impacts of Reading Recovery on student learning in schools that are part of the i3 scale-up; and
2. to assess the implementation of Reading Recovery under the i3 grant, including fidelity to the program model and progress toward the scale-up goals.

The impact evaluation comprises a multi-site randomized controlled trial (RCT) for estimating short-term impacts and a multi-site regression discontinuity (RD) design for estimating long-term impacts. The RCT component involves over 1,250 schools over four years, and is described in detail below. The implementation study involves a combination of qualitative and quantitative research executed on a large scale over the same four-year timeframe.

Each component of the study is detailed—with discussions of methods, data, and findings—in a dedicated chapter of this report.

STRUCTURE AND CONTENT OF THE FINAL REPORT

Reading Recovery is implemented through the collaboration of a network that includes local schools, districts, UTCs, and in some cases state government entities. The independent evaluation of the i3 scale-up of Reading Recovery provides a detailed exploration of the interaction of these collaborators, and of both the strengths of the Reading Recovery program and the challenges it faces. Many of these insights emerged over time, through multiple rounds of research. This final report includes separate chapters devoted to discussion of research methods and findings in a number of key areas. The following chapters follow this introduction:

Chapter 1, *Taking Reading Recovery to Scale: Processes, Challenges, and Outcomes*, discusses OSU’s and its partner universities’ progress toward the scale-up’s expansion goals. A key finding is that scale-up goals were met and, in some cases, exceeded. Several examples highlight Reading Recovery leadership’s use of the i3 grant and evaluation to catalyze system-wide introspection and operational improvements.

Chapter 2, *The Immediate Impacts of Reading Recovery: Results of the Four-Year RCT Study*, details the findings of the rigorous, multi-year impact study. We share a key finding: That

STRUCTURE AND CONTENT OF THE FINAL REPORT

short-term impacts of Reading Recovery on the reading achievement of students who received the intervention as part of the i3-funded scale-up were medium-sized to large, with significant school-to-school variation.

In Chapter 3, *Sustained Impacts of Reading Recovery: A Regression Discontinuity Study from the 2011-12 Cohort*, we present initial findings from the estimation of long-term outcomes for Reading Recovery students as they progress into third grade and beyond. This study uses a quasi-experimental design for causal inference, and Chapter 3 presents the first round of results from this study. As data are collected for additional cohorts of students as they progress through elementary school, similar analyses will be undertaken with larger pooled samples.

Chapter 4, *Reading Recovery Implementation Fidelity: Scale-up Successes and Challenges*, focuses on the results for implementation fidelity under the scale-up, and describes the methods we used to obtain them. We report that, with very few exceptions, strong fidelity to the program model was seen in all four years of the evaluation.

The implementation fidelity analysis offers one vantage point on the intervention's implementation. Chapters 5 and 6 offer additional perspectives that emerged through in-depth qualitative research into implementation of Reading Recovery. Chapter 5, *Lesson-Level Implementation: Instructional Strength in Reading Recovery*, presents findings on the implementation of the program at the lesson level, which this evaluation revealed to be both critical and highly variable. Chapter 6, *The School-Level Implementation of Reading Recovery: A Cross-Case Analysis*, discusses the players and relationships that determine the contributing factors and extent to which Reading Recovery is integrated with other school level processes.

The Conclusion to this report, *Learning from Scale-Up: Lessons from the Four-Year i3 Evaluation of Reading Recovery*, summarizes key lessons and limitations of the evaluation, and suggests future directions.

Taking Reading Recovery to Scale: Processes, Challenges, and Outcomes

This chapter examines the i3-funded scale-up of Reading Recovery as part of an independent evaluation conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania and the Center for Research on Education and Social Policy (CRESP) at the University of Delaware. The CPRE/CRESP independent evaluation was funded by a 2010 i3 Scaling Up What Works grant to The Ohio State University (OSU) for the expansion of Reading Recovery in the U.S.

Reading Recovery was developed in New Zealand during the 1970s, and introduced in the United States in 1984. Initially serving 110 students in a single Ohio school district, Reading Recovery expanded in the U.S. to serve 152,000 students nationwide during its peak year, 2000-2001. Its rapid growth was largely due to grassroots expansion efforts by OSU and a network of partner universities operating regional Reading Recovery University Training Centers (UTCs). Beginning in 2001, expansion of Reading Recovery slowed dramatically in response to widespread fiscal pressures on schools and districts and the passage of the federal Reading First legislation in 2001. The language of Reading First led many states to discourage the use of Reading Recovery because of the perception—strongly contested by Reading Recovery advocates—that the program did not meet the legislation’s requirements for scientifically based reading instruction (May et al., 2013). The i3 scale-up award afforded an opportunity to expand Reading Recovery in the U.S. once again.

In this chapter, we first present the outcomes of the scale-up relative to this opportunity, and to the expansion goals of the i3 grant. Second, we describe patterns and trends in the nationwide scale-up. Finally, we summarize the evaluation’s findings about the scale-up process. These findings pertain to the key partners and personnel involved; the strategies OSU and its partner universities used to recruit teachers, schools, and districts to participate in the scale-up; and the successes and challenges Reading Recovery personnel encountered over the course of the expansion.

STUDYING THE SCALE-UP: RESEARCH METHODS

The goals of our research on the scale-up of Reading Recovery are as follow:

1. Determine whether OSU's recruitment targets for the scale-up were reached, and what factors supported or hindered efforts to meet the targets;
2. Investigate the processes and strategies used by OSU and its university partners to recruit districts, schools, and Reading Recovery teachers and teacher leaders;
3. Examine the challenges that resulted from the scale-up, the ways in which OSU and partner universities responded; and
4. Identify any operational adaptations that emerged over the course of the scale-up process.

The CPRE/CRESP research team used a mixed-methods approach to address these objectives. In each year of the scale-up, we obtained teacher recruitment, teacher attrition, and student-service figures from the International Data Evaluation Center (IDEC), the data management and reporting system used by OSU to track Reading Recovery's growth and implementation nationwide. These data are collected and entered into IDEC by teachers and teacher leaders working in schools and districts across the country on an ongoing basis. We received annual reports from IDEC on all key scale-up outcomes in each year of the grant period.

IDEC's count data were supplemented with extensive qualitative research. Specifically, researchers conducted semi-structured interviews with the key Reading Recovery personnel responsible for administering the scale-up in every year of the grant. These personnel included the staff of the i3 office at OSU; faculty members overseeing UTCs at OSU and its 19 partner universities across the country; and teacher leaders responsible for recruiting and training teachers in districts around the U.S. The purpose of the interviews was to explore the systems and methods leveraged by the Reading Recovery community to meet scale-up goals, and to develop a thorough understanding of the challenges they encountered during the scale-up process and the strategies they used to address them. From 2011 to 2015, our researchers discussed scale-up with 173 participants. Table 1.1 shows qualitative data collection activity by year.

FINDINGS: RECRUITMENT TARGETS

The independent evaluation of the five-year i3 scale-up assessed Reading Recovery's progress toward the following goals: (a) train 3,675 new Reading Recovery teachers; (b) provide one-to-one Reading Recovery lessons to an additional 67,264 students; (c) provide other instruction—generally classroom or small-group instruction provided by trained Reading Recovery teachers—to 302,688 students; and (d) train 15 new teacher leaders. At the beginning of the evaluation, CPRE/CRESP established a threshold of 80

Table 1.1. Qualitative Research Activities Related to Scale-Up, by year

Data Collection Method	Data Source	# Conducted 2011-12	# Conducted 2012-13	# Conducted 2013-14	# Conducted 2014-15	Total
Interviews	i3 Project Staff at OSU	4	4	—	3	11
Interviews	UTC Directors	18	18	12	—	48
Interviews	Teacher Leaders	—	24	26	—	50
Focus Groups	Teacher Leaders	34	—	—	30	64
All						173

percent of the scale-up goals as the key indicator of scale-up success (i.e. 2,940 teachers trained in Reading Recovery; 53,811 students served with one-to-one lessons; and 242,150 students taught by i3 teachers in classroom or small-group settings).

Table 1.2 provides an overview of the scale-up's accomplishments, as compared with the overall goals of the five-year grant. At the conclusion of the scale-up, Reading Recovery's growth had exceeded the evaluation's threshold of 80 percent of the scale-up goals in all four areas. The scale-up surpassed its goals for the number of Reading Recovery teachers and teacher leaders trained, and the number of non-Reading Recovery students served by newly trained teachers. In the third area, students served with one-to-one Reading Recovery lessons provided by teachers trained with i3 funds, the project achieved 92 percent of its goal. As these results indicate, the effort to expand Reading Recovery under the i3 initiative was highly successful.

Over the course of the scale-up, Reading Recovery personnel leveraged the i3 grant to initiate or expand Reading Recovery in over 1,300 schools. Figure 1.1 shows school recruitment over the course of the scale-up by UTC. As this figure shows, a majority of schools recruited under i3 were located in the midwest and northeastern United States. Over a third of these schools were recruited by four of the nineteen UTCs. More than 75 percent of the 1,321 schools recruited under i3 (994 schools) used the grant to expand an existing Reading Recovery implementation by training additional teachers and serving more students. In the remaining 327 i3 schools, the grant was used to implement Reading Recovery for the first time. Figure 1.2 presents counts by state of i3-funded implementations in sites that both expanded Reading Recovery and were first-time implementers.

i3 Recruitment Priorities

In addition to goals for teachers trained and students served under i3, the scale-up also included recruitment priorities established by the U.S. Department of Education. In particular, the scale-up emphasized the recruitment of schools in rural areas, schools

FINDINGS: RECRUITMENT PROCESSES AND STRATEGIES

with large English Language Learner (ELL) populations, and schools with historically low student achievement.

Figure 1.3 shows the locale of schools that received i3 funding over the course of the scale-up. Locale classifications for this analysis were obtained for each school from the National Center for Education Statistics (NCES). The largest percentage of i3 schools are located in areas classified by NCES as suburban (29.8 percent), followed closely by urban and rural areas (27.9 percent and 27.2 percent, respectively), and finally towns (13.8 percent). Figure 1.4 shows the percent of the student population classified as ELL in each i3-funded school. In 20 percent of i3-funded schools, ELL students represented more than 20 percent of the student population (represented by the non-blue dots in the figure).

FINDINGS: RECRUITMENT PROCESSES AND STRATEGIES

Previously released reports on the progress of the scale-up (May et al., 2013; 2014; 2015) provide detailed discussions of the findings of our extensive qualitative research on the process of recruitment for the Reading Recovery scale-up. Key findings from the four years of research are summarized here.

Key Players in Recruitment: UTC Directors and Teacher Leaders

As managers of the regional hubs that constitute the national Reading Recovery network, faculty and staff of the UTCs at OSU and its partner universities are responsible for recruiting teachers and schools to implement the program. This role is an established part of the program model and is specified in the *Standards and Guidelines* (RRCNA, 2012). UTC directors reported that the i3 grant was both an unprecedented opportunity to leverage and scale up pre-existing recruitment practices, and a new and unique recruitment challenge.

In some regions, UTC directors set the strategy and tightly orchestrated i3 recruitment activities. These directors often reported spending significant time traveling and devoting considerable effort to personal contact with prospective participants. These directors concluded that a deep personal involvement was key to successful recruitment. “It’s slow going because it’s so personal,” one director reported. “So I spend a lot of time on the road and in hotels.” Another director made similar comments. “You have to make a personal connection,” she explained. “Somebody’s not going to wake up some morning and decide ‘This is what I’m going to do. I’m going to do Reading Recovery.’ It’s that personal connection that you make to recruit people.”

Teacher leaders—experienced Reading Recovery teachers who have received extensive training in order to train and support Reading Recovery teachers—also played a key role in recruitment. In some regions, teacher leaders had primary responsibility for the

Table 1.2. Progress Toward Scale-Up Goals

Goal	Scale-up Target	Year 1	Year 2	Year 3	Year 4	Year 5	Total	% of Goal Met
Reading Recovery teachers trained	3,675	300	905	876	805	861	3,747	102
Teacher leaders trained	15	3	15	6	12	10	46	306
Students served with one-to-one Reading Recovery lessons	67,264	2,400	8,480	12,832	17,352	20,936	62,000	92
Other students served by Reading Recovery teachers	302,688	12,600	44,520	67,368	91,098	109,914	325,500	108

scale-up process, 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.

As the first point of Reading Recovery contact for schools and districts, many teacher leaders reported that i3 recruitment added a significant new dimension to their jobs. Some teacher leaders reported exerting considerable effort on recruitment to little avail; many struggled to recruit enough new Reading Recovery teachers to fill a training class. Others were able to draw on previously established relationships to facilitate entry into schools:

I already have relationships and the other staff that I work with have relationships with the schools, and so, often, you can sort of piggyback or follow those relationships that preexist in order to have conversations with people. And I would say most of my recruiting has been done one-to-one, myself with an administrator, talking about, “Are you interested in this? How can we make this happen?” And my Site Coordinator, who is Director of Professional Services, has a very active role in a similar way where she has all these relationships and is constantly having conversations with people.

Other teacher leaders were inundated with i3 requests. One teacher leader recalled:

FINDINGS: RECRUITMENT PROCESSES AND STRATEGIES

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.

The extent to which directors assumed personal responsibility for recruitment versus delegating to teacher leaders was a function of personal preference and 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, director involvement in recruitment was 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."

Recruitment Strategies: Varied and Context-Driven

Through the course of our evaluation, we found that UTC regions approached the tasks of scale-up differently, and met with varying levels of success. Because each UTC region operated as a largely independent system with its own norms, policies, and standardized processes, directors were free to design and carry out recruitment strategies based on their region's unique characteristics.

By and large, UTC directors and teacher leaders reported that the strategies they used to recruit schools and districts under i3 resembled outreach activities they had undertaken in the past. Some regions relied primarily on mass marketing (e.g. flyers and brochures sent to school and district decision-makers), while others emphasized personal contacts within existing networks. As a result, recruitment efforts differed vastly from region to region.

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. In several regions, university-based comprehensive literacy initiatives — for instance, Fountas & Pinnell's Literacy Collaborative and Dorn's Partnerships in Comprehensive Literacy — were important existing platforms for Reading Recovery implementation. Often well-known and highly regarded regionally, 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 complementary to Reading Recovery's approach.

While schools nationwide used federal Title I funds to support their Reading Recovery programs, a notable difference across regions was the extent to which i3 recruitment efforts leveraged state and local political and funding resources. This variability was in

Figure 1.1. Schools Participating in the i3 Scale-Up: University Training Center Affiliation (data as of May, 2015)

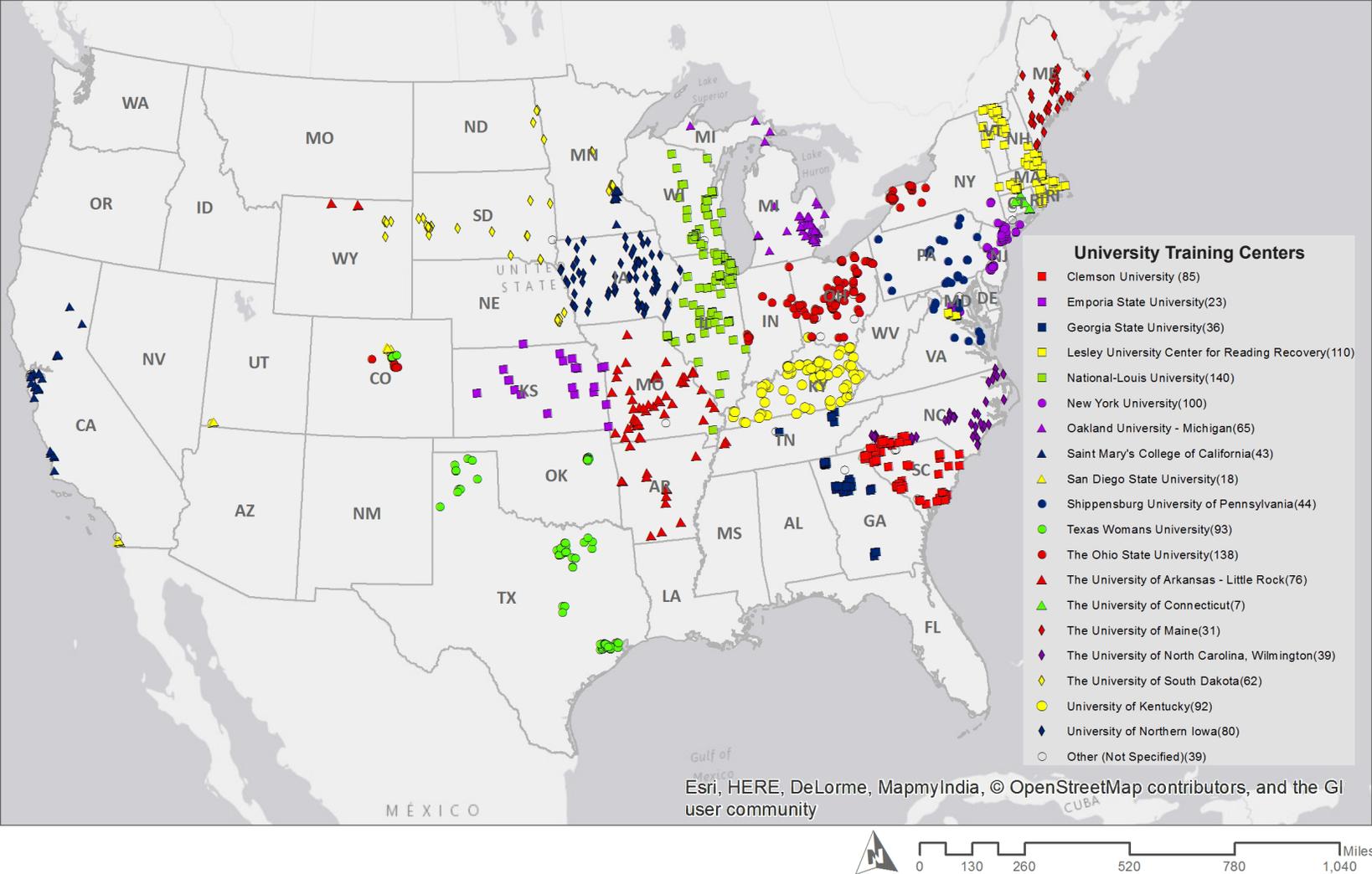


FIGURE 1.2

Figure 1.2. Schools Participating in the i3 Scale-Up: Use of i3 Funds to Support New or Expanding Implementations (data as of May, 2015)

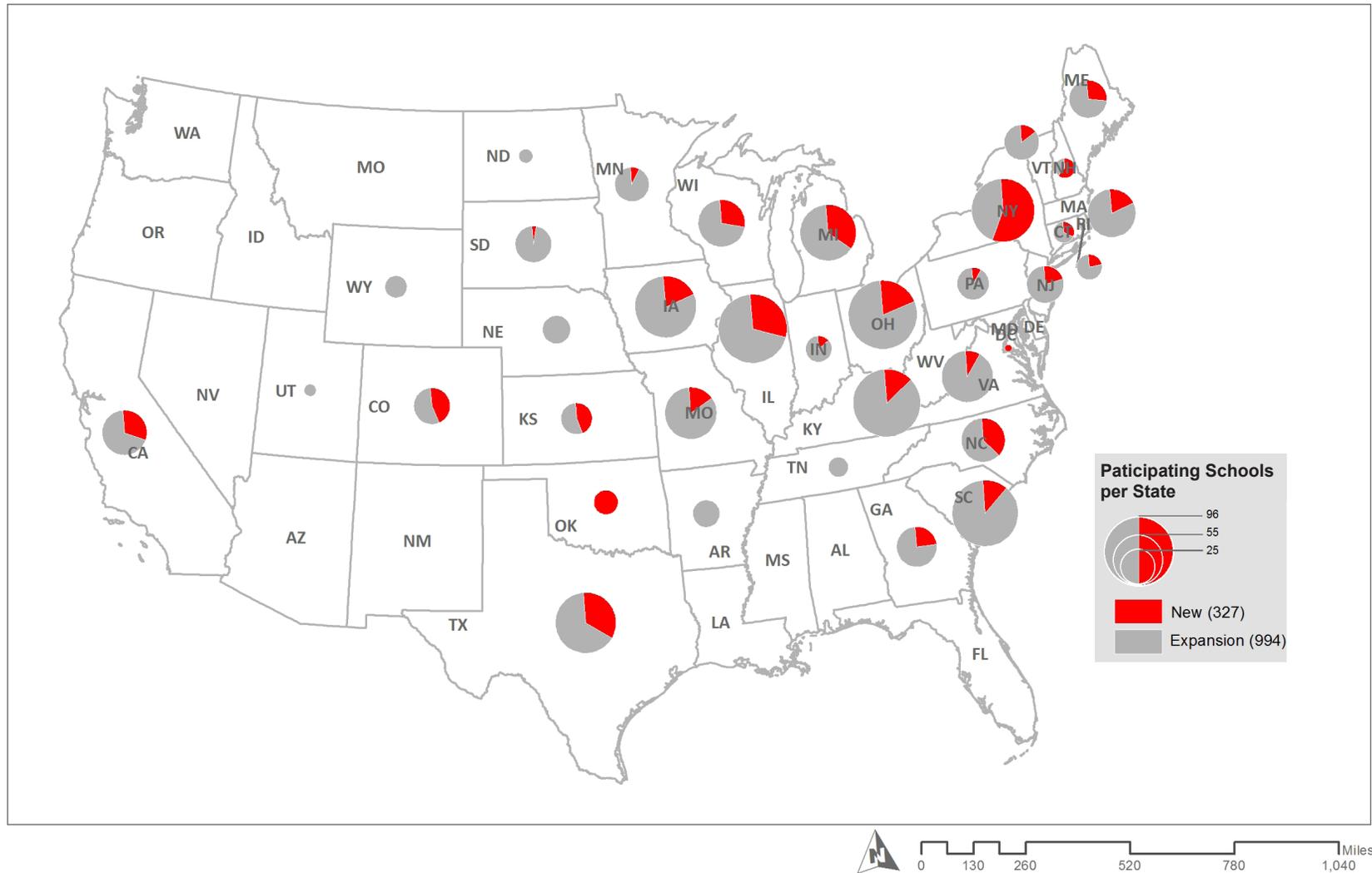


Figure 1.3. Schools Participating in the i3 Scale-Up: Locale (data as of May, 2015)

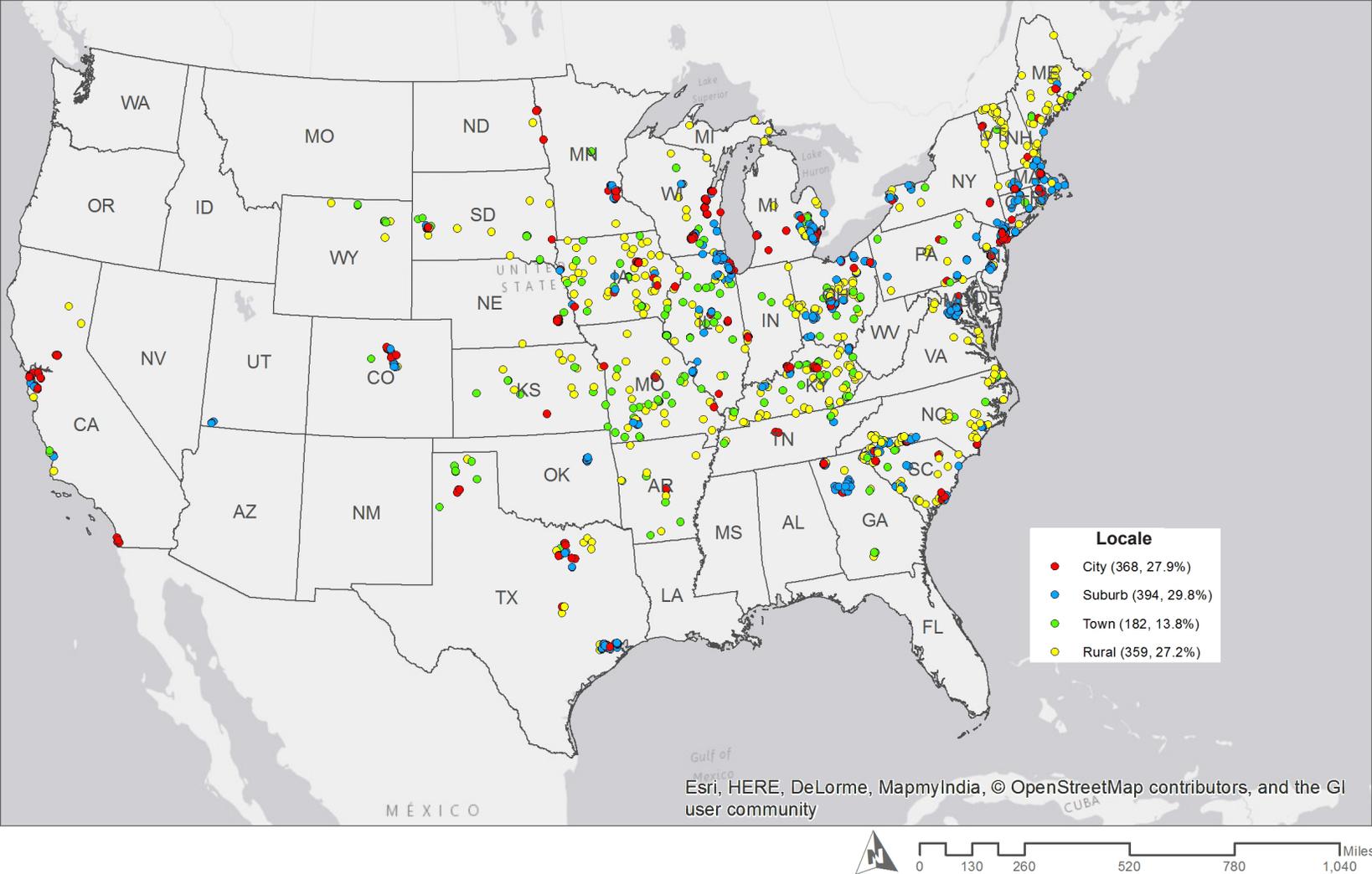
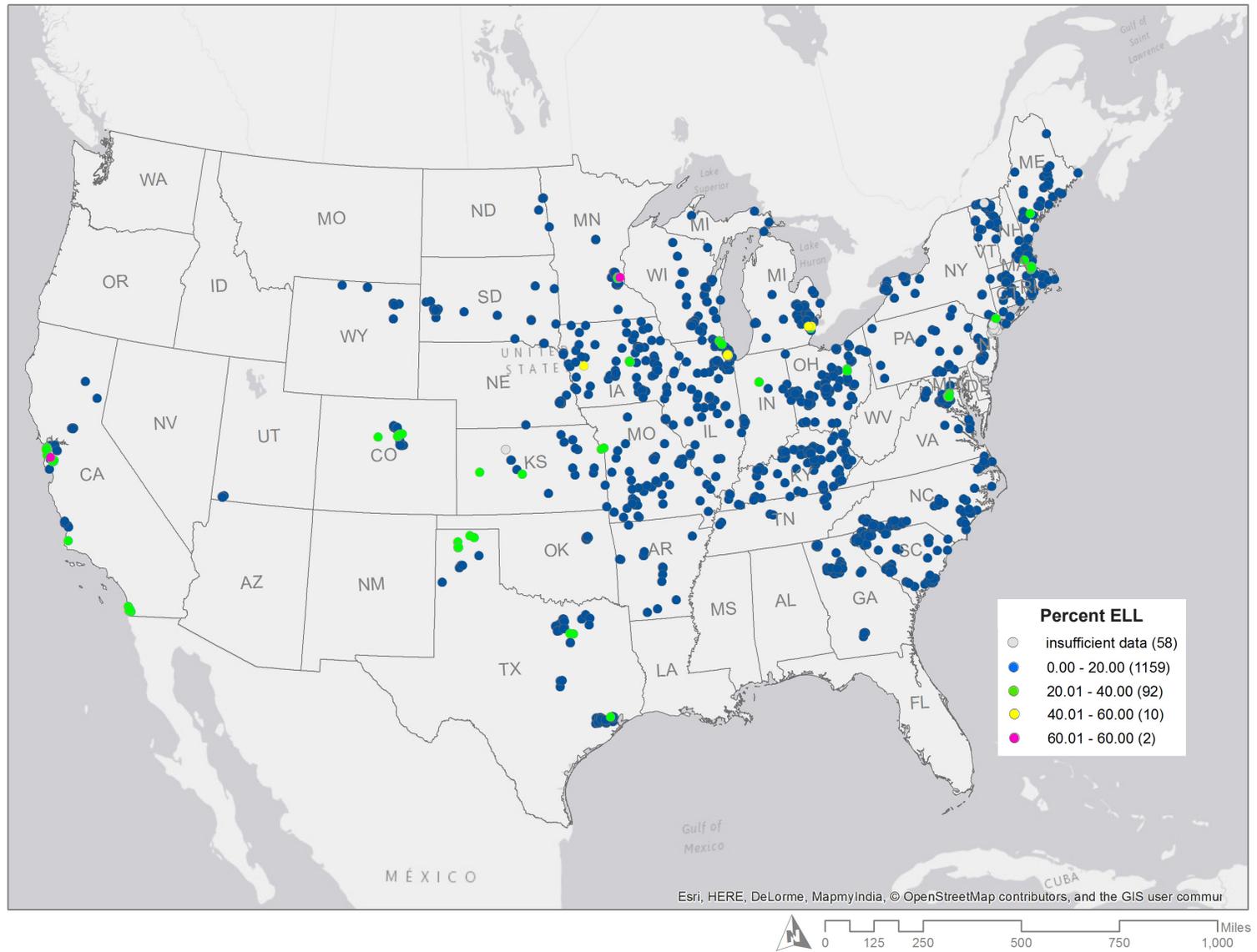


FIGURE 1.4

Figure 1.4. Schools Participating in the i3 Scale-Up: Proportion of Student Population Classified as ELL in each i3-Funded School (data as of May, 2015)



large part due to the significant differences in regional early-literacy infrastructures. We found that several regions enjoyed significant state-level political support for Reading Recovery and early-literacy programming. While much less prevalent, some regions were also afforded access to funding through state-level budget allocations. Figure 1.5 shows the breakdown of state and federal funding sources available to i3 schools for Reading Recovery.

In regions where political support existed, UTC directors reported 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. Another said:

We have been very fortunate here at [this UTC] because we have worked so closely with our state legislators, and we're just telling them the need for early [literacy] intervention and why this is so important. And so, we have been given these early literacy grants that help pay the Reading Recovery teacher's salary.

Despite the regional differences in their recruitment efforts, all UTCs emphasized the opportunity the i3 grant provided to train Reading Recovery teachers at no cost to districts and schools. 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. Others 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.”

FINDINGS: THE EVOLVING CHALLENGES OF SCALING-UP

In the early stages of the scale-up, UTC directors and teacher leaders faced challenges to recruitment that were unique to this start-up phase. In particular, their recruitment efforts were constrained by the initial timing of the i3 award, and by questions regarding the school selection criteria set forth by the U.S. Department of Education.

Confirmation of i3 grant awards in the fall of 2010 occurred two months after training for new Reading Recovery teachers typically begins. As a consequence, UTC directors were unable to use i3 funds to attract districts and schools during the peak of the recruitment season in late spring and early summer 2010. As one UTC director explained:

Year One 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, et cetera set. People are assigned to positions. It's difficult for schools to shift gears after a month of school.

As a result of the timing of the award, many UTC directors initially used i3 funds to support the existing cohort of new Reading Recovery teachers in training that year, most of whom were recruited prior to the grant award. Over time, UTCs moved past what one director called the “teething problems related to the protocols, paperwork, [and] policies” of the i3 grant, and recruitment began to accelerate and to more specifically target rural schools, those with significant ELL populations, and those with historically low student achievement.

Toward the end of the i3 grant, different challenges to school recruitment and program sustainability emerged. UTC directors and teacher leaders reported limited capacity, economic troubles, administrator turnover, and limited understanding of Reading Recovery as obstacles to overcome.

Key Challenge #1: Capacity

Beginning the second year of the scale-up, UTC directors and teacher leaders began to express concerns about the capacity limits they faced, particularly in time and workforce availability for recruitment. “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.” Many other UTC directors echoed this sentiment. They reported feeling challenged by the experience of balancing their once-primary roles as university faculty and Reading Recovery trainers with their new role as a regional administrator of the i3 grant:

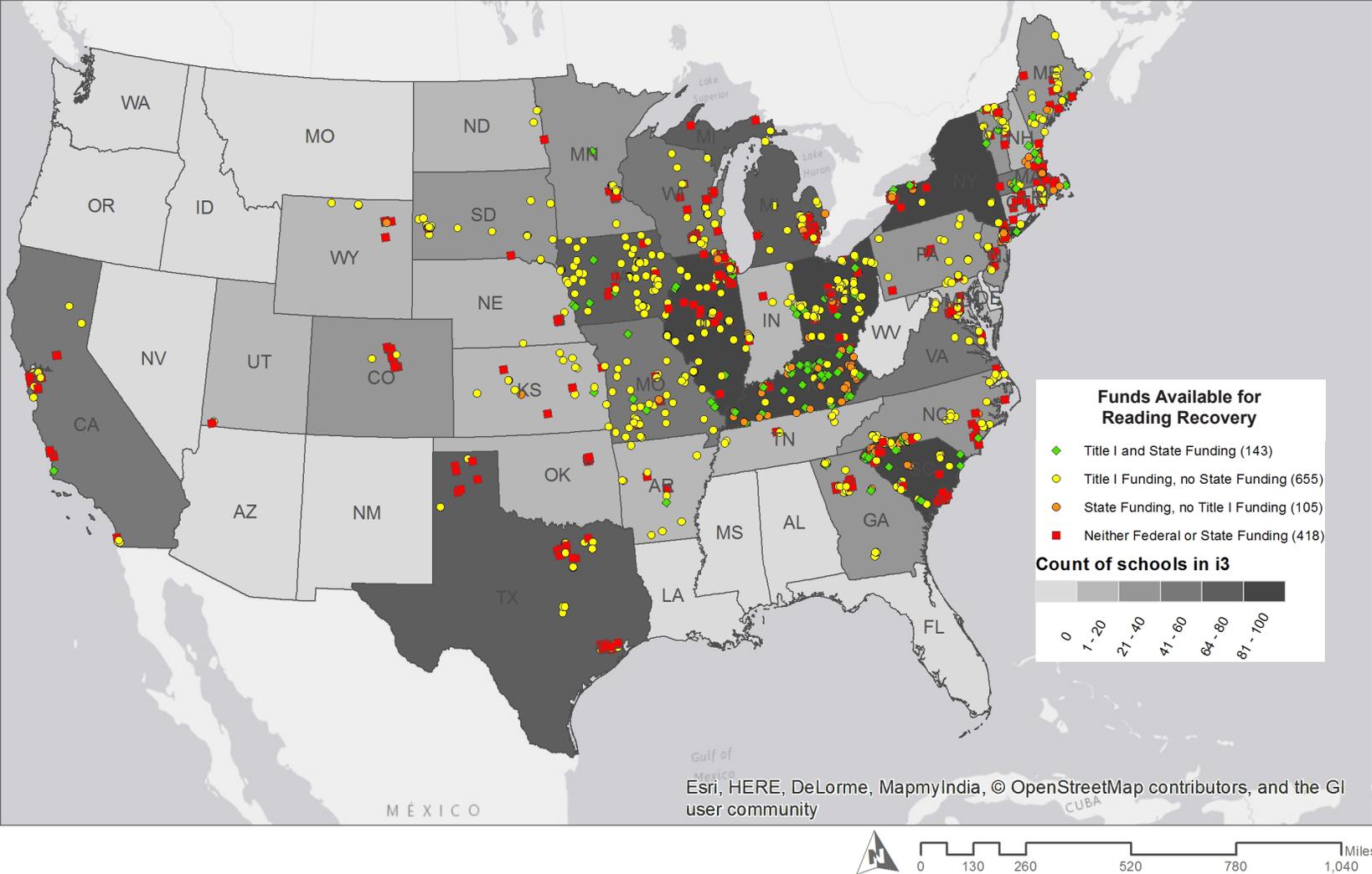
I’m doing this solo...and I’m stretched thin. I have to go over all the details with the reimbursement [of participating schools’ i3-funded expenditures]. In addition to my university faculty role, I’m also training. I’m trying to cover several large sites because there is no school district that would 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 repeated 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 oversaturated 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.

While many UTC directors reported playing central roles in generating interest for Reading Recovery among districts and schools, it was universally acknowledged that securing commitment was contingent on the availability of teacher leaders to monitor and support site-level implementations. Shortages of teacher leaders in reasonable proximity

Figure 1.5. Schools Participating in the i3 Scale-Up: Public Funds Leveraged to Support Reading Recovery



to recruitment hotspots led some UTC directors to limit the geographic areas in which they attempted to recruit or to turn away inquiring sites. 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.

Key Challenge #2: Fiscal Climate

UTC directors and teacher leaders repeatedly commented that the downturn of the U.S. economy had resounding consequences for school recruitment efforts, as well as school and teacher retention. While the i3 grant funded training and materials for participating schools' Reading Recovery programs, it did not pay the newly trained teachers' salaries, and respondents reported that many school and district administrators were daunted by that expense. "Ever since the economic crash a few years back," explained one UTC director, "schools have had to cut their budgets... So even though we're offering this training, they don't have enough in their budget to pay [a Reading Recovery teacher]." Many directors and teacher leaders commented that they heard a great deal of interest in Reading Recovery from school and district administrations, but that the cost of the implementation remained a significant barrier to adoption of the program. "Nearly every decision that's being made is being made based on funding," remarked one UTC director. 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.

In recognition of persistent financial challenges, many UTC directors and teacher leaders reported working individually with districts and schools to problem-solve the integration of an additional teacher allocation into their budget. One UTC director explained that her most successful recruitment strategy was to work on a case-by-case basis with district and school administrators to develop "creative staffing structures," emphasizing other roles Reading Recovery teachers can play in a building (e.g., small-group interventionist, literacy coach, classroom teacher), and the number of students the teacher can serve through such roles.

UTC directors and teacher leaders further reported that once schools were recruited and began implementing Reading Recovery, program cost was an ongoing threat to school and teacher retention. In interviews, teachers and school administrators frequently reported uncertainty about whether their school or district would continue to fund Reading Recovery. Because school and district budgets were frequently in flux, most principals reported that Reading Recovery was consistently in a tenuous position. “We’ll hold on to it as long as we can,” commented one principal. “But if we get to the point where we have to decide between cutting a literacy teacher and a Reading Recovery teacher, I know that the Reading Recovery teacher will have to be cut, just because of the sheer number of students [classroom teachers] serve.”

Key Challenge #3: Scaling Up with Fidelity

The complexities of balancing fidelity to a program’s model with the needs and priorities of new contexts have been discussed in the education literature for decades (Berman & McLaughlin, 1976; Dede, 2006; Dede, Honan, & Peters, 2005; Coburn, 2003; Elmore, 1996). In the second and third years of the scale-up, the challenge of scaling up with fidelity to Reading Recovery’s *Standards and Guidelines* (RRCNA, 2012) emerged as a theme in discussions with UTC directors and teacher leaders about recruitment. As Reading Recovery personnel pushed to meet scale-up goals, UTC directors sometimes found that schools that were drawn in by free training and materials were less than fully bought into the Reading Recovery model. UTC directors reported having to draw clear boundaries to ensure that program quality was not sacrificed in the scale-up process:

After that first year (of the grant), a couple of districts were kind-of weak in their implementation. And I ended up having to drop them because they weren’t going to do Reading Recovery right. They were going to do Reading Recovery after school, or every other day. And I’m like “That’s not Reading Recovery.” So I actually had to drop them.

There was a district and they flat-out refused to follow the *Standards and Guidelines* in selecting children. And the children they were selecting weren’t Reading Recovery kids, they were kids that were average or above average. And I had to just remove them from the grant. In all conscience, I couldn’t let them continue.

Some UTC directors also reported using additional measures, formal and informal, to judge a system’s potential for implementing Reading Recovery 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 ensure that 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.

Key Challenge #4: Scaling Up Sustainably

As mentioned earlier, in order to receive i3 funds, districts, schools, and teachers were asked to make a three-year commitment to implementing Reading Recovery. Many UTC directors reported that districts were wary of assuming responsibility for funding Reading Recovery for what they considered to be such a long time" to "over the long term, 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. Over time, UTC directors and teacher leaders discovered that a signed MOA was not always a guarantee that Reading Recovery would be sustained in a setting. "What has become an issue for me, from the UTC point of view," explained one UTC director, "is who signs the MOA. In a school district where I've had RR for a long time, and the central office has already agreed to support the implementation of RR, it worked very well." She continued:

But, in new districts where we've scaled-up, and the MOA did not require someone at the central office to sort of bless the implementation, that's been much more problematic. And that's where I'm losing teachers... Because when there wasn't necessarily central office buy-in, then you don't have anybody saying "But, our school district, we committed to this."... I've probably lost six to eight schools for that reason.

Several UTC directors hinted that they felt some schools and districts signed MOAs to receive free training and materials for teachers, but did not intend to maintain the program. They felt that the i3 grant, while making it possible to increase Reading Recovery presence in schools, also attracted schools to the program that were not committed to funding it themselves:

When Reading Recovery was in its heyday, teachers were very passionate about their work in Reading Recovery, and they really fought hard to have it and keep it. With the i3 grant, and I think some of this has to do with the fact that it hasn't required a financial commitment on anybody's part, because we've been giving them stuff, I don't see the commitment to doing Reading Recovery... I think if those schools had to pay for the training themselves, they wouldn't have been so nonchalant about saying "We're not going to do this anymore. It's getting too hard."

Because the training is free of charge, it's covered under the grant, many districts were eager to come on, many were eager to train. And they do sign the MOA that specifies they need to keep Reading Recovery for three years after the training year. But in some cases... the district or school pulled that teacher out after she received the training, and they're using them to meet other needs at their school.

1. i3 funds supported teacher and teacher leader training. Schools and districts are responsible for teacher and teacher leader salaries.

Many UTC directors remarked that, in an effort to combat school and district attrition from the program, they are maintaining close contact with school- and district-level implementers, working to keep Reading Recovery “on their radar,” particularly by sharing stories of Reading Recovery success and positive impacts:

I’m keeping in close contact with all of our existing sites so that Reading Recovery, even though it’s operating very well, can’t disappear from their radar. So it’s kind of in their face all the time, telling them about successes that we’re having, and how Reading Recovery is changing the academic environment and the kind of culture at the schools... We want to make sure that people in positions that are making decisions on a day-to-day or month-to-month basis know that information. And I’m keeping in touch with all of our districts with every medium that I can find, from mailing information as it comes up, to phone calls and meetings.

UTC directors reported that the potential for attrition was particularly high following district or school leadership changes. “Different administrators have different priorities,” commented one UTC director. “A lot of times school boards bring in a new administrator, and they want to make their mark... We just lost Reading Recovery in a big district because it was not [the new superintendent’s] thing. And I’ve seen it over and over and over again... It’s a constant fight.”

CONCLUSIONS: LEARNING AND ADAPTING THROUGH SCALING UP

Our regular communication with UTC directors and teacher leaders over the course of the four-year evaluation afforded visibility into Reading Recovery’s scale-up process, including the strategies and challenges underlying the scale-up. In addition, this regular communication has provided a window into the Reading Recovery system’s evolution in response to the scale-up opportunity. In summary, by engaging in open and consistent communication about both the pitfalls and successes of the scale-up process, OSU and its partner universities leveraged the scale-up experience to build a stronger and more responsive national Reading Recovery network.

A prime example of this growth surrounds student selection. As discussed in Chapter 4: *Implementation Fidelity*, early findings pointed to significant variation in the processes schools and districts used to select students for Reading Recovery intervention. In some cases, this variation resulted in deviations from the *Standards and Guidelines*’ instructions about which students should be served (See Year Two report for an in-depth discussion of this issue. May et al., 2014). As a result, student selection was one of the few areas in which inadequate adherence to the *Standards and Guidelines* surfaced early on (May et al., 2014). Qualitative research revealed that this variation in student-selection procedures stemmed both from schools’ own priorities and preferences and from inconsistent messaging from Reading Recovery personnel about which students should be identified for one-to-one lessons. In response to the evaluation’s findings on student selection in the early years of the scale-up, OSU worked in collaboration with the partner universities to clarify and communicate messaging about student selection in the second half of the

CONCLUSIONS: LEARNING AND ADAPTING THROUGH SCALING UP

scale-up. As detailed in Chapter 4: *Implementation Fidelity*, by the third and fourth years of the scale-up, high fidelity to the student selection standards was observed.

Data collection during the last two years of the scale-up also indicated that the structural impacts of the expansion may extend beyond the Reading Recovery system itself, as some UTC directors reported shifts in regional policy or infrastructure where early literacy was concerned. One director said:

The i3 grant has really improved the ways that [the state] thinks about literacy instruction over the past couple of years. It's been exciting to see that happen. [The state's] advocacy efforts have pulled together... a whole list of about 30 organizations... to pool their efforts in ways that they hadn't before. So, just for the future of education in the state, it's been a very promising development.

I have districts training their second, third, and in a couple cases fourth teacher leaders for very large districts because they captured this whole idea of a comprehensive [early literacy] plan, and I think Reading Recovery really is going to be the tipping point for that... The i3 grant is just the seed. The whole tree and fruit is the long-term view.

Despite this progress, sustainability remained a concern as grant funds were no longer available. UTC directors and teacher leaders expressed concern about both continuing to grow, and maintaining existing implementations as budgets change:

I really am concerned about what's going to happen to Reading Recovery as the grant ends. It has been a wonderful infusion of funds to do exactly what the grant was intended to do... I don't know where we would be now without the grant. But I worry about five years from now, without it. How are we going to keep this momentum that the grant provided us going?

I worry about what happens when we can't go into schools, and say "We have this grant we can help you with." Because, money speaks. When you tell people "We've got money to help you," they spring immediately. But, when you go in and talk about expanding... one principal said to me "We are literally doing canned food drives [to fund Reading Recovery]"... I think about that. I think what we do in terms of tight economics to secure and expand the implementation... What do we do there?

One UTC director explained that in the final year of the evaluation her goal was to "keep as many people that were trained through the grant doing Reading Recovery as possible." She continued:

I think sustaining [Reading Recovery] beyond the committed time is a big challenge. So [a priority is] creating a culture within the school setting to have these people not want to keep doing it because they get to keep all their stuff, but they want to keep doing it because they see what an impact they're having. So I'm trying to make them aware of that, motivating them to see that they might have to stand up and fight for Reading Recovery, and how to go about doing that. That is a high priority for me.

The Immediate Impacts of Reading Recovery: Results of the Four-Year RCT Study

This chapter presents the findings of a four-year, multi-site randomized controlled trial (RCT) to investigate the immediate impacts of Reading Recovery. The RCT was part of CPRE/CRESP'S independent evaluation conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania and the Center for Research on Education and Social Policy (CRESP) at the University of Delaware. As *Chapter 1: Taking Reading Recovery to Scale* details, the independent evaluation was funded by a 2010 i3 Scaling Up What Works grant to The Ohio State University (OSU) for the expansion of Reading Recovery in the U.S. The RCT analysis was paired with a quasi-experimental study examining Reading Recovery's long-term impacts. The results of the long-term impacts study are discussed in *Chapter 3: The Sustained Impacts of Reading Recovery*.

The RCT was conducted from the 2011-2012 school year through 2014-2015. To implement the RCT, CPRE and CRESP worked in close collaboration with OSU and the International Data Evaluation Center (IDEC), which collects and monitors operational data on Reading Recovery in the U.S. on an ongoing basis.

RESEARCH OBJECTIVES

The RCT component of the evaluation of the i3 scale-up of Reading Recovery was designed to answer the following question: What is the immediate impact of Reading Recovery on the reading achievement of struggling 1st-grade readers, as compared with business-as-usual literacy instruction? In investigating this question, the RCT addressed the following research objectives:

1. Estimate the impact of Reading Recovery at the end of the 12- to 20-week intervention on students' reading achievement, as measured by the Iowa Tests of Basic Skills (ITBS).
2. Estimate the impacts of Reading Recovery on the Reading Comprehension and

SAMPLE SELECTION AND ASSIGNMENT

Reading Words subscales of the ITBS, as well as the Observation Survey, an aligned instrument used for Reading Recovery screening and monitoring.

3. Estimate the impacts of Reading Recovery on two subgroups of particular interest under the i3 scale-up: English language learner (ELL) students and students enrolled in rural schools.

Our findings on the overall research question and the three objectives are detailed below. First, we describe our process for randomly selecting schools and students for inclusion in the RCT, and assigning students to experimental condition.

SAMPLE SELECTION AND ASSIGNMENT

Three distinct research studies were implemented in each year of the scale-up: The RCT discussed in this chapter; the regression discontinuity study that was also part of the independent evaluation of the i3 scale-up (see Chapter 3); and Reading Recovery's internal study, which is implemented by OSU on an ongoing basis. Each school was randomly selected to contribute students to one of the three studies in the first year following its recruitment into the i3 scale-up. In subsequent years, schools rotated through the other two studies (See Appendix Table A1). This strategy limited each school's participation to one study per year and also ensured that the sample of schools contributing students to each of the three studies in any year was representative of the population of i3 schools. Schools that randomized students for the RCT during the first year were not asked to participate in the fourth year so that no school was selected to implement the RCT in more than one year.

All schools that participated in the i3 Scale-Up of Reading Recovery were considered eligible to contribute students to the RCT study. A total of 1,490 i3 schools were randomly selected to randomize students for the RCT, of which 1,254 (84 percent) did. Two hundred and thirty-six schools were selected to contribute students for the RCT but either failed to implement random assignment of students or dropped out of the i3 project altogether.

Identification of Students for the RCT

Students in the 1,254 schools that participated in the RCT were identified to be part of the experiment via the following screening process: Reading Recovery teachers first screened a pool of candidates for Reading Recovery intervention using the Observation Survey of Early Literacy Achievement (OS) (Clay, 2013). Candidates included 1st-grade students who were identified by school staff—including kindergarten, 1st-grade, and intervention teachers—as struggling readers. The eight students with the lowest OS scores in a given school were then selected to participate in the RCT. This screening and identification process, which was used at the beginning of each of the study's four school years, is consistent with Reading Recovery's normal procedures.¹ A total of 9,784 students were identified for the RCT through this screening process.

1. CPRE's Year One report (May, 2013) includes a detailed discussion of the process by which these students were identified.

Random Assignment of Students

Once schools had identified eight 1st-grade students with the lowest OS scores, Reading Recovery teachers entered the names of the selected students into an online random assignment tool, noting their English language learner (ELL) status and their baseline OS Text Reading Level (TRL) subtest scores (measures are described later in this chapter). The tool then matched them into pairs by first matching any students with ELL designations, then matching the student with the lowest TRL subtest score with the next-lowest student, and so on. Once the students were matched, a randomizing algorithm then randomly assigned one student in each pair to the treatment group and the other to the control group. The result was recorded in IDEC, and the tool was locked so that randomization in that school could not be redone.

The study was designed as a delayed-treatment RCT, control students began receiving their Reading Recovery lessons after the posttests were administered to both treatment and control students in each matched pair. The “blocking” of students in matched pairs was a central part of the study design that produced, essentially, four mini-experiments per school.

Analytic Sample

A total of 4,892 students were randomized to treatment, and 4,892 to control. Both pretest and posttest data were available for 7,855 of these students (4,136 treatment; 3,719 control). Pairs in which either student was missing assessment data were dropped from the RCT, leaving a total of 6,888 students who were able to be matched into pairs with complete data (3,444 matched pairs in 1,122 schools). These students comprise the analytic sample. The analytic sample therefore represents 70 percent of the students who were randomized to treatment and control. Because the entire pair was dropped in the event that one student in a pair was missing outcome data, there is no differential attrition overall. Table 2.1 illustrates the pooled analytic sample for the four years of the study.

Figure A1 in the Appendix illustrates sample attrition from assignment through analysis. A breakdown of the analytic sample by study year is included as Appendix Table A2. Appendix Table A3 provides detail on the analytic sub-samples for ELL students and students in rural schools.

Representativeness of the Randomized Sample

In order to be able to generalize our RCT findings to the population of students in i3 schools, ideally we would be able to compare students in the randomized sample to students who attended schools that participated in the scale-up but who were never randomized. While this was not possible, we did perform sensitivity checks to determine whether the schools attended by randomized students were significantly different from those attended by non-randomized students. We observed no differences between the schools attended by randomized vs. non-randomized students based on factors such as rural classification ($p=0.60$), Title I status ($p=0.63$), and the average percentage of students

Table 2.1. Sample Size and Retention Rate

	Treatment	Control	Total
Students Randomly Assigned	4,892	4,892	9,784
Students Dropped from the Analysis	1,448	1,448	2,896
Due to Own Missing Assessment Test Data	756	1,173	1,929
Due to Missing Data on Matched Student	692	275	967
Students Included in the Analytic Sample	3,444	3,444	6,888
Sample Retention Rate	70%	70%	70%

Table 2.2. Characteristics of Study Sample

Variable	Randomized Sample	Analytic Sample	Attrited Sample	Difference	p-value for Difference
Gender (n=9,425)					
Male	5,687	60%	60%	0	0.87
Female	3,738	40%	40%	0	
ELL Status (n=9,399)					
ELL	1,759	19%	18%	1%	0.21
Non-ELL	7,640	81%	82%	1%	
Race (n=9,340)					
Black	1,211	13%	14%	1%	<0.001
Hispanic	1,809	19%	19%	0	
White	3,833	43%	35%	8%	
Other	2,487	25%	31%	6%	
Text Reading Level (n=9,375)					
0	4,492	48%	49%	1%	0.54
1	1,832	20%	20%	0	
2	1,807	20%	19%	1%	
3+	1,244	13%	13%	0	

Notes. *p*-values based on χ^2 test of independence. Source: IDEC student demographic and test score data. Some data were unavailable for some students.

in a school that are identified as in need of Reading Recovery ($p=0.13$). We found no evidence to suggest that students who were not randomized attended schools that were any different from those attended by randomized students. In addition, we surveyed Reading Recovery teachers at the 236 schools that were selected to randomize students but did not do so about why their schools did not randomize. In a majority of cases, schools' failure to randomize was a result of misunderstandings about when and how to use the IDEC randomizing tool. In five instances, the Reading Recovery program was temporarily suspended due to extenuating circumstances such as Reading Recovery teacher maternity or medical leave. The results of both of these efforts suggest that the impact estimates yielded by the RCT are generalizable to the population of students in i3 schools.

Table 2.3. Baseline Balance Tests for Student Demographics

Pretreatment Variable	Treatment Group	Control Group	p-value for Difference
Gender (n=6,867)			
Male	60%	61%	0.18
Female	40%	39%	
ELL Status (n=6,851)			
ELL	19%	19%	0.47
Non-ELL	81%	81%	
Race (n=6,820)			
Black	12%	13%	0.06
Hispanic	20%	19%	
White	42%	44%	
Other	26%	24%	
Text Reading Level (TRL) Score (n=6,888)			
0	47%	48%	0.63
1	20%	19%	
2	19%	20%	
3+	14%	13%	

Notes. *p*-values based on χ^2 test of independence. Source: IDEC student demographic and test score data. Some data were unavailable for some students.

Analysis of Attrition

We performed statistical tests of differences in student demographics for students included in the analytic sample ($n=6,888$) and those dropped due to their own or their matched-pair partners' incomplete data ($n=2,896$). Analyses of differences in student characteristics for those students who were included and excluded from the analytic sample suggest no significant differences in pretest OS text reading levels ($p=0.54$), sex ($p=0.80$), or ELL status ($p=0.21$). Students that were dropped from the analysis were disproportionately non-White (43 percent vs 35 percent) and the difference was significant ($p<0.001$)². The characteristics of the sample are detailed in Table 2.2.

Baseline Balance

Baseline balance was assessed for the analytic sample in order to confirm that the treatment and control groups were equivalent on observed characteristics. Table 2.3 presents results of baseline balance tests for student demographics and baseline TRL of the pooled analytic sample (6,888 students in 1,122 schools).

2. Following estimation of main impacts (see Findings), sensitivity checks showed that student race (represented as a binary indicator for White/non-White) was non-statistically significant as a moderator of the treatment. As such, using available data from the analytic sample, there is no evident bias on the average treatment effect that is associated with higher attrition among non-White students in the RCT.

No significant differences were found between the groups on sex, ELL status, or race, suggesting that random assignment produced treatment and control groups that were well balanced immediately prior to the start of treatment students' lessons. Baseline balance on prior reading performance, as measured by the TRL subtest, was also assessed. Again, no significant differences were found between the treatment and control groups at baseline. It is only possible to test for differences on measured characteristics; therefore, the possibility remains that there are systematic differences between the groups on characteristics we did not measure. However, this assessment provides strong evidence that the baseline characteristics and reading achievement of treatment and control groups were effectively identical immediately prior to the start of the experiment.

Treatment Condition

Students who were randomly assigned to the treatment group began receiving Reading Recovery lessons as soon as possible after assignment—typically in September or October of first grade. These students were expected to receive a typical Reading Recovery intervention, which consists of one-to-one lessons for 30 minutes per day as a supplement to regular classroom literacy instruction.

As is standard in Reading Recovery, each student's intervention was expected to last 12-20 weeks, depending on the student's pace of progress toward the goal of reaching an average text reading level for first graders. Typically, the target text reading level is 16. Students might at any time be found, via OS assessment, to have completed the program successfully, at which point their lessons would be concluded. Those who progressed more slowly might receive the full 20 weeks of lessons. Reading Recovery's program model requires that lessons are terminated after 20 weeks, whether or not a student reaches the target text level.

In an analysis of dosage, we calculated the percentage of treatment that was received for each student based on the student's recorded number of Reading Recovery lessons and final text reading level. For instance, a student who reached Level 16 or completed 60 or more lessons was considered for this i3 evaluation to have received 100 percent of the treatment; a student who completed 30 lessons without reaching Level 16 was considered to have received 50 percent of the treatment. We found that 91.5 percent of students in the intervention received 100 percent of the treatment, and that the average treatment dosage for students in the treatment group was 98.1 percent. Although our dosage calculation is an approximation, findings indicate that the estimated effects in this evaluation represent the effect of near-complete delivery of the intended treatment on students' reading achievement.

Control Condition

In this intent to treat study, students in the control condition received regular classroom literacy instruction and were also able to receive supplemental supports. Students in the control group had access to any literacy supports that were normally provided to low-achieving 1st-grade readers by their schools, other than Reading Recovery.

To confirm the experience of control group students, we surveyed their 1st-grade teachers in the second and third years of the study, and asked them to report, individually for each control group student in their classroom, what supplemental instructional services the students received during the experiment (i.e., when the treatment students were participating in Reading Recovery). Of the 3,579 1st-grade teachers of control students in the study, we received surveys from 1,898 (54 percent response rate). From these responses, we obtained information on 1,245 (57 percent) of all students assigned to the control group in Years Two and Three of the RCT. The 1st-grade teachers reported that 39 percent of these control students received no supplemental instructional supports; 37 percent participated in some individual or small-group intervention (other than Reading Recovery) provided by a Reading Recovery-trained teacher³; 23 percent participated in a literacy intervention that was not delivered by a Reading Recovery teacher; and 8 percent received ELL or special education supports. Seven percent of control students received a combination of more than one of the supplemental instructional services listed above. These findings indicate that the majority of control group students (61 percent) did experience some form of supplemental literacy support in addition to regular classroom instruction. Therefore, in this study, we are comparing the effectiveness of Reading Recovery to that of classroom instruction plus a range of other support services that schools provide to struggling readers.

MEASURES

The study's pre-specified primary research objective is the estimation of Reading Recovery's impact on students' ITBS Total Reading scores (confirmatory). Additional objectives include the estimation of impacts on ITBS sub-scales (Reading Words and Reading Comprehension) and OS scores, as well as impacts on all outcomes for ELL students and those attending rural schools (exploratory).

Assessment and Data-Collection Procedures

All students in the RCT were assessed immediately pre-intervention via the OS, which was generally administered by a Reading Recovery teacher working in the student's school. The OS was administered again immediately post-intervention, along with the ITBS (measures are discussed in detail below). Reading Recovery teachers did not administer posttest measures to their own students; rather, posttests were administered by other Reading Recovery teachers or by teacher leaders. This is standard practice in Reading Recovery and helps ensure the validity of student scores.

The precise timing of posttest administration varied, as posttests were administered to both the treatment and the control student in a given matched pair immediately after the treatment student completed his or her 12- to 20-weeks of Reading Recovery lessons. This

3. Reading Recovery teachers generally provide one-to-one lessons for half the school day. During the rest of the day, many of them work as interventionists who provide support to students in literacy or other content areas in individual or small-group settings. For more on Reading Recovery teachers' roles, see the Introduction to this report or prior publications from this evaluation (May et al., 2013, 2014, 2015).

ensured that the two students experienced their assigned conditions for an identical time period. Typically, posttests were administered roughly halfway through the school year. As the study was designed as a delayed-treatment RCT, control students began receiving their Reading Recovery lessons after the posttests were administered to both treatment and control students in each matched pair.⁴

CPRE and CRESP collaborated with IDEC at OSU to collect pre- and posttest data for the RCT, as well as student- and school-level demographic data. IDEC provides data collection, management, and analysis support to Reading Recovery programs throughout the U.S., and its existing infrastructure facilitated large-scale data collection for the i3 evaluation at low cost. Teachers entered pre- and posttest data for each student into IDEC's database, teacher leaders reviewed the data for accuracy, and IDEC staff then securely shared the data with the evaluation team via files that included student, teacher, and building characteristics.

Pretest and Exploratory Outcome Measure: OS

The OS (Clay, 2013) 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 (TRL). 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 subscales (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 information yielded through administration of the OS has been provided by several studies that 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 (Gómez-Bellengé, Rodgers, Wang, & Schultz, 2005); (b) measurement of early reading constructs (Tang & Gómez-Bellengé, 2007; Gómez-Bellengé, Gibson, Tang, Doyle, & Kelly, 2007); and (c) prediction of the attainment of performance benchmarks (Denton et al., 2006).

The full OS measure, including all six subtests, was administered to all students at pretest, and again at the end of the treatment period. Total OS scores were used as both baseline and outcome measures in the exploratory analysis of the intervention's impacts on the OS.

Pretest and Matching Variable: TRL subtest

The TRL subscale of the OS provides a single, consolidated measure of students' reading

4. In some instances, the control student was found to have made enough progress during the experiment that his or her Reading Recovery slot was given to another student. However, in most cases the control student began Reading Recovery lessons after the administration of the posttests.

abilities. The TRL subtest task involves the administration of a running record to assess reading speed and accuracy, and provides an overall assessment of a students' reading skill. TRL was used to match students during the random assignment process. In addition, it was used as the baseline measure and pretest covariate in the statistical models of impacts on the ITBS (see the impact models, below).

Confirmatory Outcome Measure: ITBS Reading Total

The ITBS Reading Total served as the confirmatory outcome measure for the RCT. 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 & Gómez-Bellengé, 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 subtests that measure achievement for students in kindergarten through the eighth grade. 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 (Hoover, Dunbar, & Frisbie, 2011). ITBS raw scores can be converted to several other types of scores, including developmental scores (grade equivalents), developmental standard scores, and status scores. 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 et al., 2011).

Information regarding the technical characteristics of the ITBS was obtained through the *Guide to Research and Development* (GRD Manual), 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 decoding 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.

Exploratory Outcome Measures: ITBS Reading Comprehension and Reading Words Subtests

In addition to the use of the ITBS Reading Total score as the confirmatory outcome measure, two subscales of the ITBS were used as outcome measures for exploratory analyses. The Reading Words subtest was used to assess impacts on decoding ability, and the Reading Comprehension subtest was used to assess impacts on comprehension.

ADDITIONAL DATA COLLECTION

Control Group Contact with Reading Recovery Teachers

CPRE/CRESP collected survey data from 1st-grade teachers in the schools participating in the scale-up in order to understand the services provided to control students while the treatment students received Reading Recovery. In addition, we surveyed Reading Recovery teachers in order to assess the extent of their instructional contact with control students during the experiment. This is especially important in this study because Reading Recovery teachers spend half of their time engaging in instructional activities outside of their Reading Recovery lessons (e.g., small-group work, co-teaching, push-in support). Control group students were allowed to receive any available supports or intervention services other than Reading Recovery, therefore we wanted to understand the extent to which control students were instructed by Reading Recovery teachers during the experiment.

Survey data were collected from each Reading Recovery teacher in an RCT school on the precise types of interaction (whole-class, small-group, individualized), frequency of interaction (from never to daily), and the timing of interaction they had with the control group students (during or after the experiment). We were able to collect information on 1,177 (54 percent) of all 2,171 students *assigned* to the control group in Years Two and Three of the RCT. We found that 8 percent of the control group students had some exposure to their Reading Recovery teacher in a whole-class setting; 34 percent had exposure in a small-group setting; and 10 percent of the control group students received individualized support from the Reading Recovery teacher during the experiment. Individualized instructional interaction with a Reading Recovery teacher has the highest potential for imitating the intervention. It does not represent contamination since it does not involve one-to-one Reading Recovery lessons. Although the rate of one-to-one instructional contact is low, overall, the fact that nearly 40 percent of the control students were taught by a Reading Recovery teacher during the experiment may have caused some attenuation of the estimated impacts.

Reasons for Missing Assessment Data

In the spring of Year Three, we administered a survey to a sample of 145 Reading Recovery teachers in the RCT schools with missing assessment data. The purpose of this data collection was to understand the reasons for missing student data and to explore any systematic differences between pairs that were retained throughout the study and those that were not. Sixty-six Reading Recovery teachers responded (46 percent response rate) with detailed information about their students for whom we did not have ITBS data. In 48 percent of instances (37 out of 77 students) no assessment data was entered into IDEC because the child was unavailable (i.e. student moved or did not complete their full series of Reading Recovery lessons.) Most of the reported instances of students

moving or leaving the school occurred in the control group. The remaining 52 percent of students ($n=40$) were reported to have missing ITBS data for logistical reasons related to the assessment—for instance, because the teacher was not aware that she was required to administer and record the assessment, or because she never received the testing materials. These survey results suggest that a large percentage (possibly half) of the missing test data was a result of natural attrition through student mobility.

Non-Compliance with Random Assignment and Control Contamination

We analyzed IDEC intervention records—which include pre- and posttesting dates, intervention start and exit dates, and the total number of lessons provided to each student—to assess control contamination. We found only a small number of matched pairs—31 pairs out of 3,444 (<1 percent)—in which the control student was exposed to Reading Recovery before the posttest measures were administered. In these cases, the period of overlap (when both treatment and control students in a pair were receiving the intervention) ranged from 10 days to 177 days. In 14 of the 31 cases, the control student’s intervention start date preceded that of the treatment student, indicating non-compliance with random assignment (0.4 percent non-compliance). In the 17 remaining cases the treatment student began first but overlapped with control (0.5 percent control contamination). Given the very limited amount of non-compliance and control contamination in our analytic sample (less than 1 percent combined), the impact estimates we present here would remain practically unchanged after adjustment for non-compliance. As such, we present only the “intent-to-treat” estimates.

ANALYSES

The RCT addressed the following research objectives:

1. Estimate Reading Recovery’s impact on the early literacy competence of 1st-grade students, as measured by the ITBS;
2. Estimate the impacts of Reading Recovery on subscales of the ITBS as well as the OS, an aligned instrument used for Reading Recovery screening and monitoring; and
3. Estimate the impacts of Reading Recovery on two subgroups of particular interest under the i3 scale-up: ELL students and students enrolled in rural schools.

Impacts on student reading performance were estimated in all analyses by comparing immediate post-intervention reading achievement of students randomly assigned to participate in Reading Recovery at the beginning of first grade to students randomly assigned to the control condition. We used a three-level hierarchical linear model (HLM) (Raudenbush & Bryk, 2002) with students nested within matched pairs, and matched pairs nested within schools. 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 estimated causal impact is less prone to problems associated with attrition through student and school non-participation.

Differences in the posttest performance of the treatment and control students were estimated after controlling for pretest performance. For the confirmatory analysis of Reading Recovery's impacts on the ITBS Reading Total scores for the full sample as well as exploratory analyses of ELL and rural subgroup impacts and impacts on ITBS subscales, this HLM included each TRL score as a covariate. For exploratory analyses of impacts on the OS, the OS Total score was used as the covariate. All models also included a binary indicator of treatment condition, a four-level fixed effect for year, an interaction effect for treatment by year, a random effect for matched pair, a random effect for overall school performance (i.e., school intercepts), and a random effect for the impact of Reading Recovery (i.e., school treatment effects). A completely general (unstructured) covariance matrix was used, which included a correlation between random effects for school-level intercept and slope. Additionally, a grouped residual variance was included to account for differences in dispersion of outcome scores within the treatment versus control groups. 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.

Model-adjusted average student scores on the outcome measures were estimated for treatment and control groups pooled across all years, schools, and pairs, along with the group contrasts and associated standard errors. Least square means estimates provide estimates of the group means and group mean differences that are based on the HLM model parameters. The results are raw differences that can be converted to standardized effect size units and benchmarked to effect sizes typically found in evaluations of reading interventions. Appendices A3-A8 present raw means and standard deviations for the full sample and the subgroups on all measures.

Parameter estimates from the HLM model were used to calculate the overall mean differences in outcomes between treatment and control groups after controlling for baseline scores. The impact estimate was then standardized using the standard deviation of the outcome for the control group to produce Glass' Δ . We also calculated a population-based Cohen's d standardized effect size, which is calculated by dividing the estimate of treatment impact by the standard deviation of outcome measure for the national norming sample. This allowed 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.

Imputation methods for missing outcome data were not used because they do not guarantee unbiased causal estimates if the data are not missing at random. Pairs with missing assessment data were instead dropped from the experiment. Although this resulted in a reduced sample size, the impact estimates are still valid indicators of causal impacts.

Table 2.4. Descriptive Statistics for ITBS Scores and OS Total Scores for Treatment and Control Groups

Post-Intervention Outcomes	Treatment Group (n=3444)	Control Group (n=3444)
ITBS Total Scale Scores		
Mean (Standard Deviation)	138.8 (7.5)	135.4 (7.2)
Mean Percentile Rank ^a	36	18
ITBS Comprehension Scale Scores		
Mean (Standard Deviation)	140.0 (9.5)	136.0 (9.0)
Mean Percentile Rank ^a	39	23
ITBS Reading Words Scale Scores		
Mean (Standard Deviation)	140.7 (9.0)	137.1 (8.2)
Mean Percentile Rank ^a	43	27
OS Total Scores ^b		
Mean (Standard Deviation)	496.5 (44.2)	451.4 (49.0)
Mean Percentile Rank	33	7

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

^b For OS scores, treatment $n=3371$; control $n=3322$.

FINDINGS

Table 2.4 shows simple descriptive statistics for the treatment and control groups on scale scores from the posttest administration of the ITBS and OS measures. For each set of scores, the means are one-third to one-half of a standard deviation larger in the treatment group. Differences in percentile ranks are +18 for ITBS Total Scores, +16 for ITBS Reading Words, +16 for ITBS Reading Comprehension, and +26 for OS Total Scores. Appendix Table A4 shows means and standard deviations for both groups on the ITBS.

Model Estimates for Main Impact Analyses

Table 2.5 shows the model-based average scores for ITBS Total Reading scale scores pooled across all years of the RCT. The full set of parameter estimates and standard errors from the HLM model are included in Appendix Table A10. The impact estimate for the difference between treatment and control students' ITBS Total Reading scores was 3.41 points ($p<.0001$; 95 percent CI = 3.09, 3.72). This mean impact estimate translates into an increase of .48 standard deviations relative to control group mean and a .37 standard deviation increase relative to the national population of first graders.

Impacts on Target Subgroups

In keeping with the goals of the i3 scale-up funding, the evaluation of Reading Recovery includes a specific focus on rural schools and students who are ELL. Exploratory analyses of treatment effects on ITBS Total Reading Scores for these two subgroups are presented

Table 2.5. Impact Estimates on ITBS Total Scale Scores

Mid-Year Outcomes	Treatment Group (n=3444)	Control Group (n=3444)	Difference	Glass's Δ^b	Cohen's d^c
ITBS Total Reading Scores					
Mean	138.71	135.30	+3.41		
(Standard Error)	(0.16)	(0.15)	(0.16)	+0.48	+0.37
Mean Percentile Rank ^a	36	18	+18		

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

^b Control SD: ITBS-T SD=7.16

^c Population SD, ITBS Level 6, Fall: ITBS-T SD=9.1

below. Appendix Tables A6 through A9 show simple descriptive statistics for ITBS Total Reading Scores of students in rural schools and for ELL by treatment and control groups. Once again, the means are over one-half of a standard deviation larger in the treatment group.

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 impact estimate for the difference between rural treatment and control students' expected Total Reading scores on the ITBS was 3.00 ($p < .0001$; 95 percent CI = 2.49, 3.51). Dividing that impact estimate by the standard deviation of the control group yields a Glass' Δ effect size of 0.43 standard deviations. Additional analyses using the full analytic sample that include an additional interaction between the Rural indicator and Treatment assignment indicator showed no significant difference in impacts on ITBS Total Reading Score for rural and non-rural schools. Estimates of the school-level variability in treatment effects were statistically significant at the 99 percent confidence level, and the variance was smaller in rural schools than the effect variance in the overall analysis. This suggests that while the majority of rural schools' Reading Recovery programs have similar program effects on student performance than the full RCT sample, there is less variability in school-level impact estimates. The school intercept/impact correlation was not statistically significant, suggesting no relationship between the average ITBS scores and the magnitude of Reading Recovery impact estimates in rural schools.

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 impact estimate for the difference between ELL treatment and control students' expected Total Reading Scores on the ITBS was 4.08 with a p -value significant at greater than 99 percent confidence. Dividing that impact estimate by the standard deviation of the control group yields a Glass' Δ effect size of 0.57 standard deviations. Additional analyses using the full analytic sample that include an additional cross-level interaction between the ELL indicator and Treatment assignment indicator showed no significant difference in impacts on ITBS Total Reading Score for ELL and

non-ELL students. Estimates of the school-level variability in treatment effects were statistically significant ($p < 0.0001$, 95 percent CI = 3.35, 4.81), and the variance was even larger in the ELL student population than the effect variance in the overall analysis. This suggests that while the majority of Reading Recovery programs have positive impacts on ELL student performance, they vary greatly in their ability to produce sizeable impact estimates. The school intercept/impact correlation was not statistically significant, suggesting no relationship between the average ITBS scores and the magnitude of Reading Recovery impact estimates among ELL students.

Variation in Impact Estimates

The significant variance components for random effects in the HLM models of impacts on ITBS scores suggest that the magnitude of the Reading Recovery impact estimates varies substantially across schools. Findings show an overall treatment effect of 3.41 points, with a random effect variance estimate of 10.01 points for the school-level impacts (see Appendix Table A10). Taking the square root of this variance estimate yields a standard deviation of 3.2 points. Some but not all of this variation is the result of noisy school-level impact estimates due to small n s at the school level. Furthermore, the school intercept/impact correlation was negative and statistically significant at the 99 percent confidence level ($\rho = -.24$, $p < .0001$). This suggests that Reading Recovery impact estimates tended to be larger in schools where students have lower average ITBS scores overall.

The Impact of Reading Recovery on ITBS Subscales and OS Total Scores

Table 2.6 shows the model-based average scores for ITBS subscales and OS Total Scores pooled across all years of the RCT. The full set of parameter estimates and standard errors from the HLM model are included in Appendix Table A10.

The impact estimate for the difference between treatment and control students' expected OS Total scores was 43.5 with a p -value significant at greater than 99 percent confidence. Dividing that impact estimate by the standard deviation of the control group yields a Glass' Δ effect size of 0.89 standard deviations. Alternatively, dividing the impact estimate by the standard deviation from the U.S. norms for OS Mid-Year (D'Agostino, et. al. 2012; $s = 10.2$) yields a Cohen's d effect size of 0.99 standard deviations.

The impact estimate for the difference between treatment and control students' expected Reading Words scores on the ITBS was 3.57 points ($p < .0001$). Dividing this impact estimate by the standard deviation of the control group yields a Glass' Δ effect size of 0.43 standard deviations. Alternatively, dividing the impact 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 0.35 standard deviations.

Analyses of impacts on the ITBS Reading Comprehension subscale showed similar results. The impact estimate for the difference between treatment and control students' expected

Table 2.6. Impact Estimates on ITBS Subscales and OS Total Scores

Mid-Year Outcomes	Treatment Group (<i>n</i> =3444)	Control Group (<i>n</i> =3444)	Difference	Glass's Δ^d	Cohen's d^e
ITBS Reading Words Scale Scores					
Adjusted Mean	140.55	136.98	+3.57		
(Standard Error)	(0.19)	(0.17)	(0.20)	+0.43	+0.35
Mean Percentile Rank ^a	43	27	+16		
ITBS Comprehension Scale Scores					
Adjusted Mean	139.82	135.92	+3.90		
(Standard Error)	(0.21)	(0.18)	(0.21)	+0.43	+0.38
Mean Percentile Rank ^a	39	23	+16		
OS Total Raw Scores^{bc}					
Adjusted Mean	495.37	451.88	+43.49		
(Standard Error)	(0.76)	(0.79)	(0.95)	+0.89	+0.99
Mean Percentile Rank	31	7	+24		

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

^b Percentile ranks based on U.S. Norms for OS Mid-Year (D'Agostino, et. al. 2012).

^c treatment *n*=3371; control *n*=3322.

^d Control SD: ITBS-C SD=8.98; ITBS-RW SD=8.23; OS-T SD=49.43

^e Population SD, Level 6, Fall: ITBS-C SD=10.2; ITBS-RW SD=10.2; OS-T SD=43.96

Reading Comprehension scores on the ITBS was 3.90 with a *p*-value significant at greater than 99 percent confidence. Dividing that impact estimate by the standard deviation of the control group yields a Glass' Δ effect size of 0.43 standard deviations. Alternatively, dividing the impact estimate by the standard deviation from the ITBS 2005 national norming sample of 1st-grade readers (i.e., *s* = 10.2) yields a Cohen's *d* effect size of 0.38 standard deviations.

Benchmarking the Effects of Reading Recovery

Our impact analysis revealed effect sizes on the ITBS and its subscales that range between 0.30 and 0.42 standard deviations (see Appendix Table A10) in each year of the evaluation. These are large relative to typical effect sizes found in educational evaluations. In their paper on the interpretation of effect sizes, Lipsey et al. (2012) offer a number of useful benchmarks for understanding the magnitude of these effects. For randomized studies that use "broad scope" standardized tests as the outcome measure for interventions at the elementary level, the authors report average effects of 0.08 standard deviations (Lipsey et al., 2012, p. 34). This benchmark suggests that the total standardized effect sizes (using Cohen's *d*) for Reading Recovery of 0.37, was 4.6 times greater than average for studies that use comparable outcome measures.

Table 2.7. Reading Recovery Treatment Effects as Compared with National Benchmarks for First Graders

	ITBS Total	ITBS RW	ITBS Comp
Treatment Effect (growth in ITBS scores)	3.41	3.57	3.90
Cohen's <i>d</i>	0.37	0.35	0.38
Treatment students' additional months of learning, over national growth average for first graders	1.55	1.62	1.77
Treatment students' growth rate, as a percentage of national average for first graders	131%	132%	135%

Note. From the start of first grade through the fifth month (i.e., the period during which the treatment students received Reading Recovery instruction), ITBS Reading Total scale scores are expected to increase from 133 to 144 for the average student in the U.S. (Hoover et al., 2003). The treatment effect represents additional gains experienced by students who received Reading Recovery.

Based on their analysis of 181 different samples, Lipsey et al. (2012) also present mean effect sizes for different types of educational interventions. They report a mean standardized effect size of 0.13 for “curricula or broad instructional programs.” The authors specifically include Reading Recovery in this group. This indicates that Reading Recovery’s effects were 2.8 times greater than the reading outcomes of other instructional interventions. Similarly, the impacts of Reading Recovery were 3.5 times larger than the average effects of Title I programs reviewed by Borman and D’Agostino (1996).

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 permits the interpretation of impacts as an increase in growth rate during the study period. Table 2.7 shows the expected gains on the ITBS benchmarked against the national sample, the gains in terms of additional months of learning, and the growth rate for Reading Recovery students compared to the national average for beginning first graders.

From the start of first grade through the fifth month of the school year (the period during which the treatment students received Reading Recovery instruction), ITBS Reading Total scale scores for the average student in the U.S. are expected to increase from 133 to 144 (Hoover et al., 2003). This increase of 11 points over a five-month period suggests that the additional gains of 3.41 points experienced by Reading Recovery students of our evaluation is roughly equivalent to an additional 1.6 months of learning, and translates to a growth rate that is 31 percent greater than the national average growth rate for beginning first graders. Table 2.7 also includes data for the Reading Comprehension and Reading Words Subscales.

CONCLUSIONS

This four-year evaluation revealed significant positive impacts of Reading Recovery on students' reading achievement. Treatment students who participated in Reading Recovery outperformed students in the control group on the Total Reading battery of the ITBS, Reading Comprehension and Reading Words subscales of the ITBS, and the OS. The average ITBS Total Reading score for the Reading Recovery (treatment) group was equivalent to the 36th percentile for students nationally, while the average score for the control group was equivalent to the 28th percentile for students nationally—a difference of +18 percentile points. A similar pattern of large gains in test scores for the Reading Recovery students relative to their control group counterparts was observed using subtests of the ITBS and the OS. Moreover, these findings were generally similar for students attending schools in rural and their counterparts in non-rural areas and for ELL students and their non-ELL counterparts.

Our findings also revealed substantial variation in effect estimates across schools. The vast majority of schools experienced positive impacts, and some schools produced effect estimates that were many times larger than those of typical reading interventions. Schools with lower-than-average ITBS scores tended to have larger treatment effects. Our efforts to understand this variation via statistical analysis of available data—including school- and student-level demographic information and, as Chapter 4 details, fidelity to the program model—did not yield clear answers. However, the extensive implementation study component of this evaluation produced several hypotheses as to possible sources of school-to-school variation in impacts. These hypotheses are discussed in Chapter 5 and Chapter 6 of this report.

Finally, the goal of the RCT was to assess the immediate impacts of Reading Recovery on students' achievement in literacy. The findings presented in this chapter therefore pertain only to the impacts that were evident immediately following the intervention. In the following chapter, we present our research on the longer-term impacts of the program.

Sustained Impacts of Reading Recovery: A Regression Discontinuity Study from the 2011-12 Cohort

This chapter examines the sustained impacts of Reading Recovery as part of an independent evaluation conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania and the Center for Research on Education and Social Policy (CRESP) at the University of Delaware. The independent evaluation was funded by a 2010 i3 Scaling up What Works grant to The Ohio State University (OSU) for the expansion of Reading Recovery in the U.S.

Although the multi-site randomized controlled trial (RCT) conducted under CPRE/CRESP's external evaluation of the i3 Scale-Up of Reading Recovery produced rigorous evidence of large impacts, the RCT provides no information about whether these impacts are sustained beyond the first half of first grade. This is because most students assigned to the control condition in the RCT ended up participating in the Reading Recovery intervention during the second half of first grade, after the posttest outcome data were collected for the RCT. Thus, while the RCT provides excellent information about the immediate impacts of Reading Recovery for those students, a different design is needed to evaluate the impacts at the end of first grade and the persistence of these impacts through later grades.

To estimate these sustained effects, CPRE/CRESP implemented an alternative research design in a separate, randomly selected sample of schools during each year of the independent evaluation (i.e., 2011-2015). A regression discontinuity (RD) design used cutoff-based assignment using pre-intervention test scores. Students with the lowest scores on the Observation Survey of Early Literacy (OS; Clay, 2005), relative to other students in their school, were assigned to Reading Recovery. Those who scored just above the cutscore never received the intervention and became the comparison group for the RD study.

This design aligns with the selection process advocated in the *Reading Recovery Standards and Guidelines of Reading Recovery in the United States* (6th ed.) (RRCNA, 2012), which states that students selected for Reading Recovery should be those with the lowest OS

scores in a school. The result is that no student who would have otherwise received Reading Recovery services was denied those services as a result of the study. At the same time, this design allows us to obtain rigorous evidence of long-term causal effects by tracking the outcomes for Reading Recovery and RD comparison students into later grades. The details of this RD study of sustained effects are provided in the sections that follow.

METHODS

The RD study of sustained impacts was conducted each school year from 2011-12 through 2014-15. The sections that follow describe the research questions; logic of the research design; and the measures, sample, procedures, and statistical analyses employed in this study.

Research Questions

This RD Study and analyses were guided by several research questions as follows:

1. What was the short-term impact of Reading Recovery during the first half of the school year under the RD design and how did this compare to the impact estimate produced by the more rigorous RCT?
2. What was the short-term impact of Reading Recovery during the second half of the school year under the RD design, and how did this compare to the impact during the first half of the school year?
3. What was the long-term impact of participation in Reading Recovery in first grade on students' reading and/or English Language Arts (ELA) performance in third grade?
 - a. What was the “intent-to-treat” effect, regardless of whether a student completed Reading Recovery, or was referred for additional intervention?
 - b. What was the effect for those students who successfully completed Reading Recovery¹ and were not referred for additional intervention?

Research Design

This study employs a multi-site RD design. Each participating school carried out the research procedures described below, including administration of a pre-intervention

1. Successful completion of Reading Recovery, or “discontinuing” the intervention, is determined by the Reading Recovery teacher in consultation with the Teacher Leader, and typically coincides with the student’s reading performance reaching a level near the average for his or her class. Thus, this subsample excludes all students who failed to complete at least 12 weeks of Reading Recovery lessons (typically because they moved or the school year ended) and all students who failed to make sufficient progress and were referred for additional intervention (e.g., special education services).

assessment, assignment of students to the intervention based on pre-assessment scores, and collection of outcomes data at the middle and end of first grade and again at the end of third grade.

The *What Works Clearinghouse* (WWC) has published draft standards for RD studies (WWC, 2014; Schochet, Cook, Deke, Imbens, Lockwood, Porter, & Smith, 2010) that include four elements: (a) integrity of the forcing variable, (b) attrition, (c) continuity of the outcome-forcing variable relationship, and (d) functional form and bandwidth for statistical models estimating impacts. The results reported in the sections that follow address each of these standards.

Measures

The instrument used to assess students' reading performance at the beginning, middle, and end of first grade was the OS (Clay, 2005). The OS is the primary screening, diagnostic, and monitoring instrument for Reading Recovery. It is a one-to-one, teacher-administered, standardized assessment that includes six subscales: Letter Identification, Concepts about Print, Ohio Word Test, Writing Vocabulary, Hearing and Recording Sounds in Words, and Text Reading Level.

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, 2005); measures of the inter-assessor reliability of the Text Reading and Writing Vocabulary tasks yielded coefficients of .92 and .87 (Denton, Ciancio, & Fetcher, 2006). In addition, evidence of the validity of information yielded through administration of the OS has been provided by several studies that assess the construct and criterion validity of the OS tasks using the sub-tests of various norm-referenced tests, including the Iowa Tests of Basic Skills (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).

Long-term outcomes were measured by collecting scale scores on state achievement tests in reading (in Georgia, Iowa, Illinois, Kentucky, Ohio, and Texas) or ELA (in Massachusetts and South Carolina) administered at the end of third grade. The published reliabilities (or median reliability based on standard error of measurement) of each of these instruments is .80 or higher. Scores on each state test were standardized to a common *z*-score scale by subtracting the statewide mean and dividing by the statewide standard deviation separately for each state (see May, Perez-Johnson, Haimson, Sattar, & Gleason, 2009).

The results reported here are based on our analysis of a subset of the 2011-12 cohort of first graders—the first cohort of participants in the i3-funded scale-up and RCT study. Because most state assessments are first administered in third grade, this is the only

cohort for which state test data were available prior to the August, 2015, conclusion of the i3 project. Third-grade state assessment data for the 2012-13 cohort of students will be available in Fall, 2015, and for the 2013-14 and 2014-15 cohorts in Fall, 2016 and 2017, respectively. A follow-up study is planned that will include collection of long-term outcomes data for all four cohorts from the i3 scale-up.

Selection of Schools

During the summer before the start of each school year (i.e., 2011-2014), each new school joining the Reading Recovery i3 Scale-Up was randomly selected to participate in one of three research designs during their first year. The three research designs included (a) the RCT, (b) the RD study, and (c) an internal study used by Reading Recovery to monitor and update national norms for the OS assessment. Each school was then rotated through these three research designs based on a predetermined order (e.g., RD in 2011-12, RCT in 2012-13, OS norming in 2013-14, etc.). Prior to the 2011-12 school year, 370 schools were randomly selected from the population of i3 schools to implement the RD study. Of these, 331 schools actually participated in the RD study; of the remaining schools, 21 dropped out of the i3 Scale-Up before the school year began and 18 failed to administer the OS as directed (e.g., no students were tested who were not assigned to Reading Recovery). These schools are excluded from the RD study sample. Over subsequent years, more than 1,500 schools participated in the RD study. As noted, however, because long-term outcomes data (i.e., state assessment data) are not yet available for Cohorts 2-4, only the schools and students participating in the RD study during the 2011-12 school year are included in the analyses presented here.

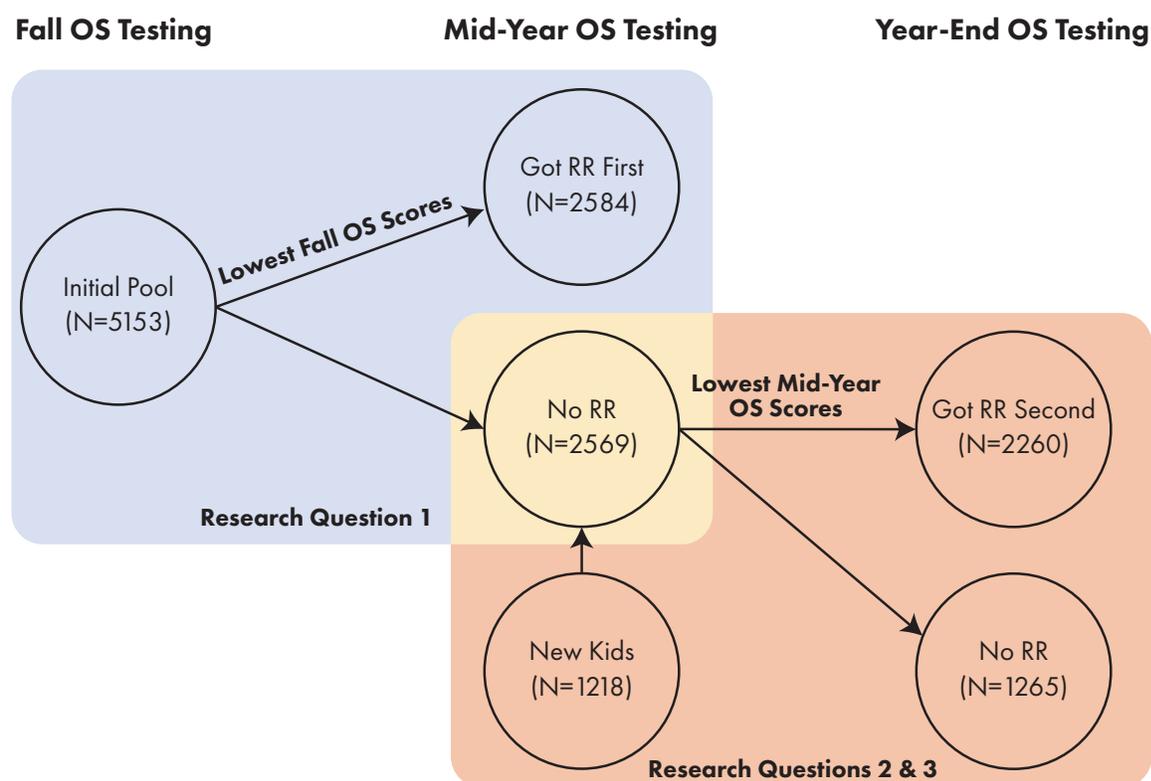
Because collection of long-term outcomes (i.e., state test scores in third grade) required extensive effort to locate each student in district databases and record their 3rd-grade scores in IDEC, collection of long-term outcomes data was carried out only in eight states with the greatest number of schools participating in the i3 Scale-Up: Georgia, Iowa, Kentucky, Massachusetts, Ohio, South Carolina, and Texas. In these states, there were 85 schools who participated in the 2011-12 RD study (26 percent of the 331 that participated 2011-12).

Assignment of Students

In each school participating in the RD study, Reading Recovery teachers were instructed to administer the OS at the beginning of the school year to all students who were selected for Reading Recovery screening based on low reading performance.² This is consistent with the normal screening process for Reading Recovery. Furthermore, for the purposes of the RD study, teachers were instructed to identify at least 4 more students for screening than their schools' Reading Recovery program could accommodate, even if these

2. A detailed description of schools' screening processes is presented in CPRE's Year One report (May et al., 2012). Briefly, most schools identify a pool of low-achieving students for screening based on a combination of teacher reports and assessments administered at the end of kindergarten and/or the beginning of first grade. The students from this pool with the lowest OS scores are assigned to Reading Recovery.

Figure 3.1. Flow of Students through the RD Study



additional students were not as low-performing as others being screened. Typically, this resulted in pre-intervention testing of 8 to 25 students with the OS at the beginning of first grade in each school (median $n=14$; total $N=5,153$). This expanded screening pool, created the sample for the RD study.

Teachers were instructed to then rank order the students by their OS scores and assign those students with the lowest scores to Reading Recovery, working from lower to higher scores until all open Reading Recovery slots were filled. This typically resulted in 4 to 12 students assigned to Reading Recovery during the first half of the school year (median $n=7$; total $N=2,584$) and an additional 4 to 14 students in each school assigned to a waitlist (median $n=7$; total $N=2,569$).

As the first wave of Reading Recovery students completed their intervention about midway through first grade, Reading Recovery teachers were instructed to test all previously assessed students again using the OS.³ In addition, any students who had been selected for Reading Recovery screening midway through first grade were also assessed with the OS (total $N=1,218$). Those students who did not receive Reading Recovery lessons in the first group were then rank-ordered based on their mid-year OS scores, and those students with the lowest scores were assigned to Reading Recovery in order of lower

3. Of the 2,569 students not assigned to Reading Recovery during Wave 1 of the RD, 262 students were not tested with the OS at mid-year and therefore did not continue into Wave 2 of the RD.

METHODS

to higher scores until all open Reading Recovery slots were filled. This typically resulted in 4 to 12 students assigned to Reading Recovery (median $n=6$; total $N=2,260$) and 1 to 7 students (median $n=4$; total $N=1,265$) assigned to a waitlist in each school during the second half of the school year. Figure 3.1 shows the flow of the RD assignment process through the two waves.

Because the RD in this study involved two distinct waves with separate assignment processes, each wave is treated as a separate RD. More specifically, the first wave allows us to use the OS scores from the beginning of first grade as the forcing variable in an RD analysis with mid-year OS scores as the outcome measure. This Wave 1 RD is identical in timing of assessments to the RCT conducted in a separate random sample of schools at the same time. Thus, the first wave of the RD study serves as a replication of the RCT study of short-term impacts.

Analyses of the second-wave RD data are restricted to include only those students who had not previously participated in Reading Recovery and who were tested with the OS midway through first grade. We then use these mid-year OS scores as the forcing variable in an RD analysis with year-end OS scores and 3rd-grade test scores as the outcome measures. Note that those students who were not assigned to Reading Recovery during the second half of first grade never participated in the intervention—the Reading Recovery *Standards and Guidelines* (2012) strictly prohibit the provision of Reading Recovery to second graders or to students who are repeating first grade. Thus, the second wave RD allows long-term tracking of student outcomes through third grade and beyond.

Although Reading Recovery requires that students with the lowest OS scores be selected for the intervention, there is no predetermined formula for creating a composite OS score based on the six subscales. As such, Reading Recovery teachers are expected to use the information across the subscales to rank the students according to the following process.

Select the lowest-achieving students first. Generally speaking, these are the children with the most stanine 1s, 2s, and 3s on the six tasks of the Observation Survey. But also take into account the students' responses and raw scores on the individual tasks. Where children's profiles are similar, select the student with the least evidence of problem-solving activity (monitoring, self-correcting, initiating solving). Also, ask the classroom teacher who is gaining the least from classroom instruction (North American Trainers Group, 2015).

This suggests two things: (a) the relative weight of each subscale for determining assignment to Reading Recovery is likely to vary from one school to another given differences in level and variation of OS scores in different schools, and (b) there is the potential for inclusion of information beyond the OS subscales in determining assignment when students have similar OS score profiles.

To deal with the first issue, we used a statistical model to estimate OS subscale score weights separately for each school (see next section). To deal with the second issue, we analyzed both a “fuzzy” RD (Trochim, 1984), which estimates impacts given

noncompliance with pure OS-based assignment (Bloom, 2009), and a “sharp” RD including only schools where the OS scores completely accounted for assignment to Reading Recovery. The next section describes how each school’s compliance with the RD design was determined and how their subscale weights, OS composite scores, and cutscore were derived from the data.

Determining Cutscores and Composite OS Scores

To derive each school’s OS composite scores and ranking of students for selection into Reading Recovery, we estimated a separate logistic regression model for each school in which we predicted assignment to Reading Recovery based on the smallest subset of OS subscale scores that would yield perfect prediction. More specifically, we estimated all possible logistic regression models using all possible combinations of one to six subscales, and then identified the model with the fewest predictors that produced a concordance index of 1.00 (i.e., perfect prediction) and quasi-complete or complete separation in the data (Allison, 2012). Schools for which perfect prediction was not possible using all six OS subscales were deemed to have implemented an imperfect RD and were included only in the fuzzy RD analyses. Separate logistic regression models were run and composite scores calculated for assignment to Reading Recovery at the beginning of first grade and at the middle of the year (i.e., separately for each wave of the RD study).

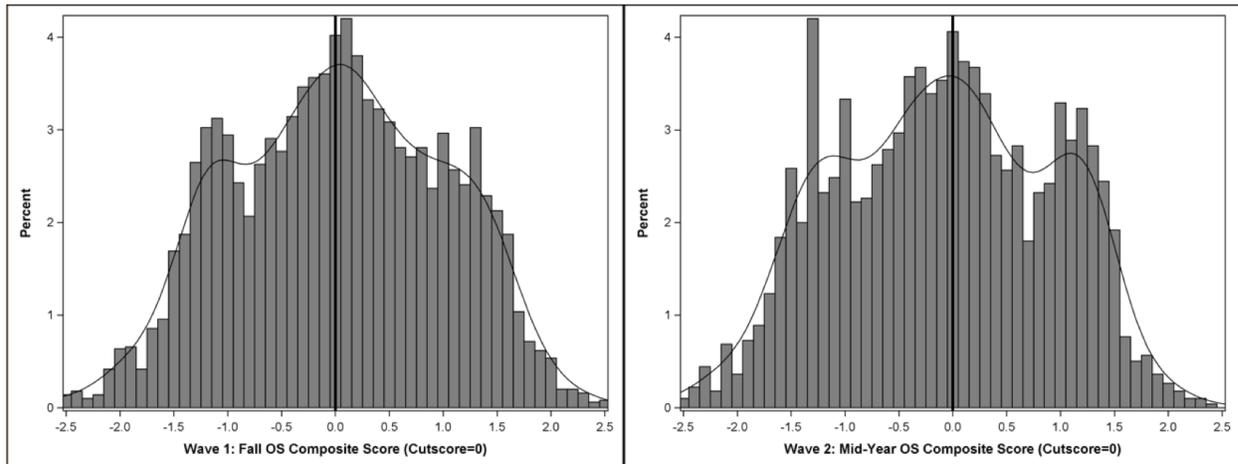
The final model for each school was used to calculate a predicted value for each student (p_{ik}) representing a composite score where each contributing subscale was weighted by its regression coefficient for school k . These scores were then normalized using Blom’s (1958) method via PROC RANK in SAS and rescaled to create final pretest OS composite scores ($Pretest_{ik}$) on a standard normal (i.e., z -score) scale with lower scores associated with assignment to Reading Recovery.

The cutscore for each school ($Cutscore_k$) was then set as a uniform random value between the $Pretest_{ik}$ score for the student with the highest score in school k who was assigned to Reading Recovery and the student with the lowest $Pretest_{ik}$ score in school k who was not assigned to Reading Recovery. Each student’s composite OS score was then centered around the cutscore for that student’s school (i.e., $Pretest_{ik} - Cutscore_k$).

Integrity of the Forcing Variable and Cutscores

OS testing was completed and scores recorded before the cutscore was determined for each school. This precludes teachers from manipulating students “true” OS scores as a means for influencing treatment assignment (WWC, 2014; Schochet et al., 2010). One could argue that manipulation could occur as a result of differentially weighting OS subscales in order to influence assignment to Reading Recovery. Unfortunately, there is no way to formally test for such manipulation. Anecdotal reports from interview data collected during the i3 evaluation (see May, et al., 2012; May, et al., 2014) suggest that assignments are intended to rely exclusively on OS scores, that teachers feel a strong obligation to prioritize students with the lowest OS scores, and that three OS subscales (i.e., Text Reading Level, Letter Identification, and Hearing and Recording Sounds in

Figure 3.2. Histograms of Forcing Variables for Waves 1 & 2 of the 2011-12 RD



Words) were often the dominant elements when ranking students.

We were able to assess the continuity of the forcing variable used in both waves through histograms depicting the distribution of OS scores for each forcing variable. Both histograms suggest excellent continuity in the forcing variables.

Statistical Models of Impacts

Program impacts under the RD design were estimated by comparing performance of students below and above derived cutoff scores for Reading Recovery eligibility. Because there is variability in schools' cutoff values (i.e., the lowest eight students in one school have OS scores that are higher or lower than the lowest eight students in another school), the generalizability of results beyond students near a single cutoff score is enhanced. Using multilevel statistical models, the performance of students above and below the cutoff score was compared, with students nested within each participating school. This multilevel design included the centered pretest assignment variable as a covariate at the student level, a parameter for the discontinuity associated with assignment to Reading Recovery, a random effect for overall school performance (i.e., a random school intercept), a random effect for the pretest slope (i.e., a random school slope), and a random effect for the impact of Reading Recovery (i.e., a random treatment effect across schools).⁴ Models were estimated using PROC MIXED in SAS 9.3 via Restricted Maximum Likelihood (REML) with degrees of freedom based on within- and between-cluster sample sizes.

The mathematical form of the primary impact model for the RD study is included as Equation B1 in the Appendices.

Given the small school and student sample sizes for some analyses (e.g., long-term

4. WWC standards for RD designs (WWC, 2014; Schochet et al., 2010) require that multi-site studies estimate separate treatment effects for each site. We accomplish this by including random intercepts, slopes, and treatment effects for each school in our multilevel models of impacts.

analyses involving only 75 schools from eight states), the correlation parameters ($\rho_{\alpha,\lambda}$, $\rho_{\alpha,\varphi}$, or $\rho_{\lambda,\varphi}$) were dropped from the model due because they were inestimable. In some cases, the variance parameters (α_{κ} , λ_{κ} , or φ_{κ}) were dropped from the model due because their point estimates were zero or inestimable.

In accordance with WWC standards for RD studies (WWC, 2014; Schochet et al., 2010), model fit and potential misspecification were assessed graphically via scatterplots and spline curves and also by testing for an interaction between pretest scores and the treatment assignment variable. Assumptions of linearity in the RD analyses were further assessed by testing polynomial parameters and by imposing various restrictions on the bandwidth around the cutoff score. More specifically, analyses were restricted to include only students whose pretest scores fell within ± 1.0 or ± 0.5 standard deviations of the cutoff score).

FINDINGS

Attrition

Of the 331 schools that implemented the RD design, 274 schools (83 percent) were deemed to have implemented a sharp RD cutoff for assignment of students to Reading Recovery at the beginning of first grade (i.e., concordance index = 1.0). The majority of these schools (62 percent) exhibited perfect prediction of assignment to Reading Recovery using just one or two OS subscales. The 56 schools exhibiting fuzzy RD implementation had concordance indices that ranged from .70 to .99 (median = .89) using all six subscales.

After restricting analyses to only those students for whom mid-year OS scores were available, the sample of students included in analyses for the first wave of the fuzzy RD was reduced from 5,153 to 4,716 (8 percent student-level attrition; 0 percent school-level attrition). When further restricting analyses to only the 274 schools implementing a sharp RD cutoff (17 percent school-level attrition), the sample of students included in analyses for the first wave of the RD study was further reduced to 3,522 (32 percent student-level attrition).

For the mid-year assignment to Reading Recovery (i.e., the second wave of the RD study), 325 schools successfully implemented a second wave of the RD study (i.e., they conducted mid-year OS testing of both Reading Recovery and non-Reading Recovery students); however, only 263 of these schools collected year-end OS scores for both Reading Recovery and non-Reading Recovery students (19 percent school-level attrition). Of these schools, 231 (88 percent) were deemed to have implemented a sharp RD cutoff for assignment of students to Reading Recovery during the second half of first grade. The majority of these schools (64 percent) exhibited perfect prediction of assignment to Reading Recovery using just one or two OS subscales. The 32 schools exhibiting implementation of a fuzzy RD had concordance indices that ranged from .78 to .99 (median = .94) using all six subscales.

After restricting analyses to the 263 schools collected year-end OS scores for both Reading

TABLE 3.1

Table 3.1. School and Student Attrition Rates by RD Study Wave, Design, and Outcome Measure

Wave	Design	Outcome	School N	Student N		School Attrition	Student Attrition	
				Treatment Group	Untreated group		Treatment Group	Untreated Group
Wave 1 Short-Term	Initial Sample		331	2,584	2,569			
	Fuzzy RD	Mid-Year OS Score	331	2,292	2,424	0%	11%	6%
	Sharp RD	Mid-Year OS Score	274	1,685	1,837	34%	35%	28%
Wave 2 Short-Term	Initial Sample		325	2,260	1,265			
	Fuzzy RD	Year-End OS Score	263	1,759	1,221	19%	22%	3%
	Sharp RD	Year-End OS Score	231	1,499	1,045	29%	34%	17%
Wave 2 Long-Term	Initial Sample		85	638	431			
	Fuzzy RD	3 rd -Grade State Test	75	390	275	12%	39%	36%
	Sharp RD	3 rd -Grade State Test	73	366	264	12%	43%	39%

Recovery and non-Reading Recovery students, and to only those students for whom year-end OS scores were available, the sample of students included in analyses for the second wave of the fuzzy RD was reduced from 3,525 to 2,980 (15 percent student-level attrition). When further restricting analyses to only those schools implementing a sharp RD cutoff (29 percent school-level attrition), the sample of students included in analyses for the second wave of the RD study was further reduced to 2,544 (28 percent student-level attrition).

Within the 85 schools in the eight states where long-term 3rd-grade test scores were collected, there were 1,069 students who were tested with the OS in the middle of first grade and assigned to Reading Recovery or not. Ten schools were dropped from the analysis for failure to test any students not assigned to Reading Recovery (12 percent school-level attrition). After restricting analyses to the remaining 75 schools and those students for whom 3rd-grade state test scores were available, the sample of students included in analyses for the second wave of the RD study was reduced to 665 students (38 percent attrition). When further restricting analyses to only those schools implementing a sharp RD cutoff ($N = 73$), the sample of students included in analyses for the second wave

of the RD study was further reduced to 630 (41 percent attrition). Table 3.1 shows attrition rates for RD analyses by study wave, design (fuzzy vs. sharp), and outcome measure.

Replicating the Short-Term Impacts from the RCT (Research Question 1)

Given that Wave 1 of the RD study was conducted at exactly the same time as the 2011-12 RCT but in a separate random sample of i3 schools, it is possible to examine how closely the design replicates the results from the i3 RCT in terms of impacts on mid-year OS scores (i.e., $ES = .88$ standard deviations on the OS in the RCT). After restricting RD analyses to only those students for whom mid-year OS scores were available, the sample of students included in analyses for the first wave of the fuzzy RD was reduced from 5,153 to 4,716 (8 percent student-level attrition).

Figure 3.3. Fuzzy RD Piecewise Spline Regression of Assignment to Treatment in Wave 1 from 2011-12

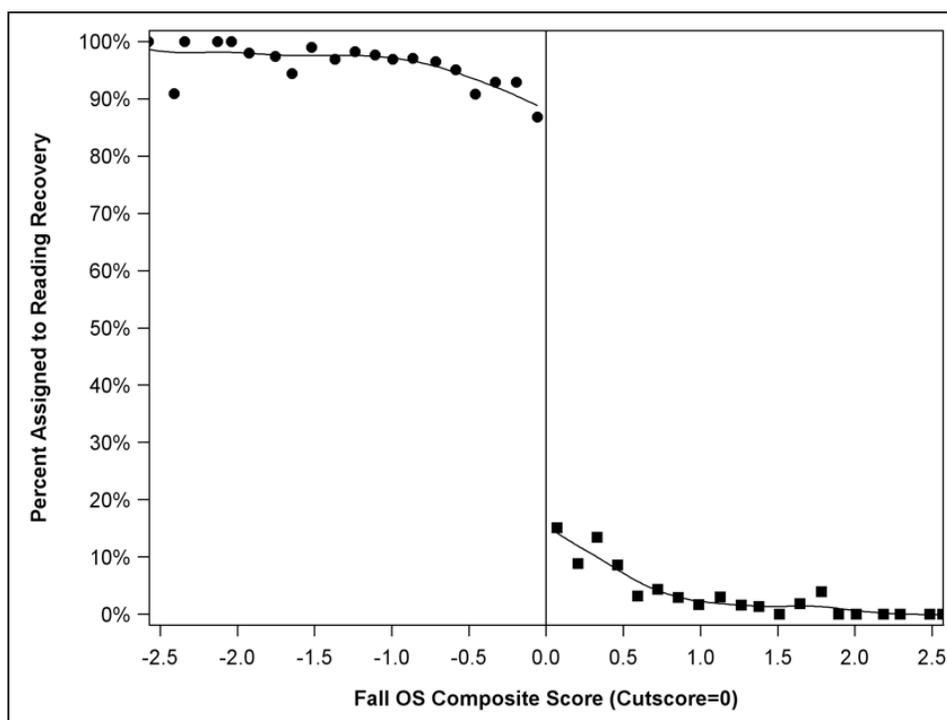


Figure 3.3 shows a piecewise spline regression confirming the discontinuity in treatment assignment at the cutscore for the 2011-12 Wave 1 RD study. Compliance with intended treatment assignment was good, with 89.6 percent of students with Fall OS composite scores between -0.25 and 0.0 assigned to the treatment, and 88.8 percent of students with Fall OS composite scores between 0.0 and $+0.25$ not assigned to the treatment.⁵

For those students below the RD cutscore, the median number of Reading Recovery lessons

5. These compliance rates are used to calculate complier average causal effect at the cut-point ($CACE_c$) using equation 12 from Bloom (2009).

Table 3.2. Fuzzy RD Parameter Estimates for ITT Effects of Reading Recovery on Mid-Year 1st-Grade OS Reading Scores in Wave 1 from 2011-12

Dependent Variable: Mid-Year OS Reading Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	474.71	1.72	<.0001
Pretest (β_1)	8.44	1.09	<.0001
Treatment Effect (β_2)	22.41	1.97	<.0001
Random Effects			
School Intercept Variance (τ^2)	610.27	76.84	<.0001
School Pretest Slope Variance (ψ^2)	94.22	27.77	.0003
School Treatment Impact Variance (ξ^2)	238.07	101.55	.0095
School Intercept/Slope Correlation ($\rho_{\alpha,\lambda}$)	-0.35	0.12	.0025
School Intercept/Impact Correlation ($\rho_{\alpha,\varphi}$)	-0.53	0.11	<.0001
School Slope/Impact Correlation ($\rho_{\lambda,\varphi}$)	0.32	0.21	.1227
Student-Level Residual Variance (σ^2)	1092.96	25.14	<.0001

Note. Sample includes 331 schools and 4,716 students.

received was 71 (e.g., about 14 weeks), with 80.1 percent of students receiving at least the minimum 12 weeks of lessons (i.e., 60 lessons). Considering 60 or more lessons as having received 100 percent of the intervention and less than 60 lessons as a percentage of the intended 60 (e.g., 30 lessons is 50 percent of the treatment), the average treatment dosage for students below the Wave 1 RD cutscore was 96.2 percent, suggesting almost complete delivery of the intended treatment.

Results from the Fuzzy RD model of short-term impacts on OS scores at the middle of first grade are shown in Table 3.2.

HLM analyses for the Wave 1 Fuzzy RD revealed an effect of +22 points on mid-year OS scores. Dividing by the mid-year 1st-grade population standard deviation of OS scores of 44 points (see D'Agostino, et. al. 2012) yields an intent-to-treat (ITT) effect size of .51 standard deviations.

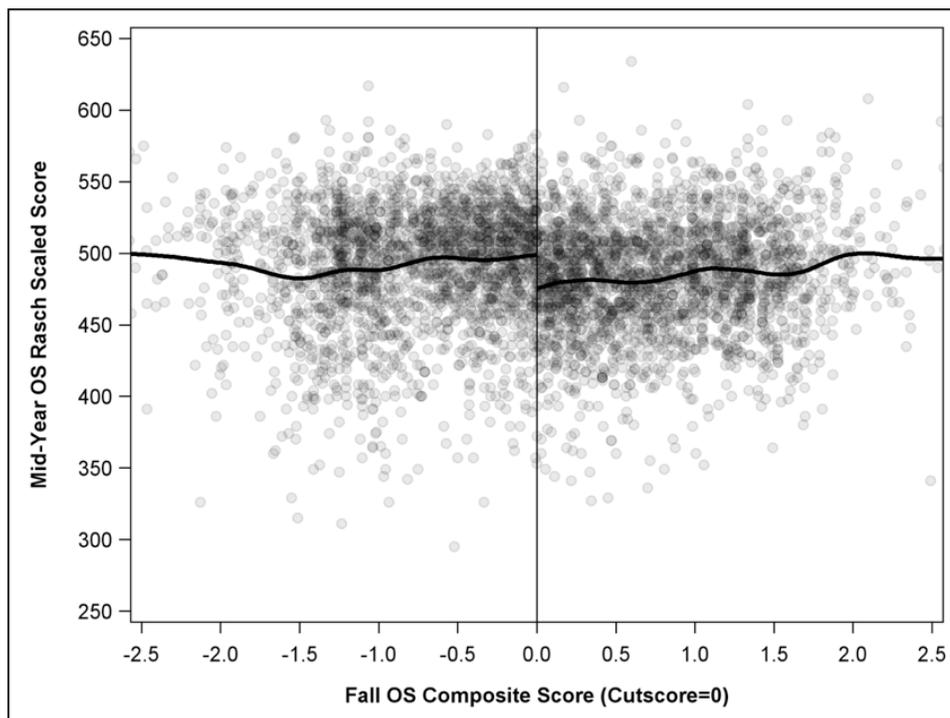
To produce the complier average causal effect at the cut-point ($CACE_c$), we followed the formula from Bloom (2009, p. 12, equation 12) and divided the raw effect estimate by the difference in treatment participation rates (i.e., 89.6 percent - 11.2 percent) for those students whose forcing variable score was within .25 standard deviations of the cutscore (i.e., just below and above the cut-point). This yielded a $CACE_c$ impact estimate of +.65 standard deviations on mid-year OS scores, after adjusting for non-compliance near the cutscore. This RD impact estimate is somewhat similar to the RCT estimate of .88 standard deviations from 2011-12.

Results in Table 3.2 also include several significant random effects parameters. Perhaps

the most interesting of these is the significant School Treatment Impact Variance of 238.07, which corresponds to a standard deviation of 15.4 points on the mid-year OS scale. A 95 percent plausible value interval for site specific impact estimates ranges from -7.8 points to +52.7 points (-.18 to +1.2 standard deviations). These results are very similar to those from the RCT suggesting that the majority of schools exhibit medium to large effects, with some schools showing very large effects in excess of a full standard deviation.

Figure 3.4 shows a piecewise spline regression confirming the discontinuity in the outcome variable for Wave 1 of the 2011-12 RD study. The piecewise spline regression is very linear, with no evidence of an interaction between the forcing variable and the treatment. Formal statistical tests found no significant polynomial terms (up to the third degree), no interaction, and no appreciable change in effect estimates when restricting bandwidth to within ± 1.0 or ± 0.5 standard deviations of the cutscore (raw effect estimates ranged from +19.9 to +22.4 points).

Figure 3.4. Fuzzy RD Piecewise Spline Regression of Mid-Year OS Reading Scores in Wave 1 from 2011-12 (331 schools, 4,716 students)



After restricting analyses to only those students from the 274 schools exhibiting a sharp RD design in Wave 1, the sample of students included was reduced from 5,153 to 3,522 (31.7 percent student-level attrition). Figure 3.5 shows a piecewise spline regression confirming the perfect discontinuity in treatment assignment at the cutscore for this restricted sample. All of the students below the cutscore were assigned to Reading Recovery, while all of those students above the cutscore were not assigned to Reading Recovery during the first half of first grade.

Results from the HLM model of short-term impacts on OS scores at the middle of first grade based on the sample of 274 schools and 3,522 students in the sharp RD in Wave 1 of 2011-12 are shown in Table 3.3. Results show an effect of +29.7 points on mid-year OS scores. Dividing by the

Table 3.3. Sharp RD Parameter Estimates for ITT Effects of Reading Recovery on Mid-Year 1st-Grade OS Reading Scores in Wave 1 from 2011-12

Dependent Variable: Mid-Year OS Reading Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	469.60	1.89	<.0001
Pretest (β_1)	11.86	1.23	<.0001
Treatment Effect (β_2)	29.71	2.14	<.0001
Random Effects			
School Intercept Variance (τ^2)	593.01	83.41	<.0001
School Pretest Slope Variance (ψ^2)	82.04	33.15	.0067
School Treatment Impact Variance (ξ^2)	145.04	110.92	.0955
School Intercept/Slope Correlation ($\rho_{\alpha,\lambda}$)	-0.16	0.16	.3053
School Intercept/Impact Correlation ($\rho_{\alpha,\varphi}$)	-0.44	0.16	.0056
School Slope/Impact Correlation ($\rho_{\lambda,\varphi}$)	-0.18	0.53	.7295
Student-Level Residual Variance (σ^2)	1025.54	27.69	<.0001

Note. Sample includes 274 schools and 3,522 students.

mid-year 1st-grade population standard deviation of OS scores of 44 points yields an ITT effect size of .68 standard deviations, which is nearly identical to the CACE_C estimate from the Fuzzy RD presented earlier.

Figure 3.6 shows the discontinuity in the outcome variable for Wave 1 of the 2011-12 Sharp RD sample of 274 schools and 3,522 students. Once again, the piecewise spline regression is quite linear in all but the tails of the data, and it exhibits no evidence of curvilinearity or an interaction between the forcing variable and the treatment. Formal statistical tests found no interaction, no significant polynomial terms (up to the third degree), and no appreciable change in effect estimates when restricting bandwidth to within ± 1.0 or ± 0.5 standard deviations of the cutscore (raw effect estimates ranged from +28.9 to +30.6 points).

We also estimated one additional sensitivity check in which we tested for discontinuities in outcomes at a point in the forcing variable where there should be no impact. Specifically, we re-estimated the impact model after shifting the cutscore by one-half standard deviation above and then below the original cutscore. Results confirmed the absence of a discontinuity with non-significant point estimates ($p > .05$) near zero ($< .08$ standard deviations).

Short-Term Impacts during the Second Half of First Grade (Research Question 2)

Wave 2 of the RD allows us to estimate short-term effects of Reading Recovery on year-

Figure 3.5. Sharp RD Piecewise Spline Regression of Assignment to Treatment in Wave 1 from 2011-12

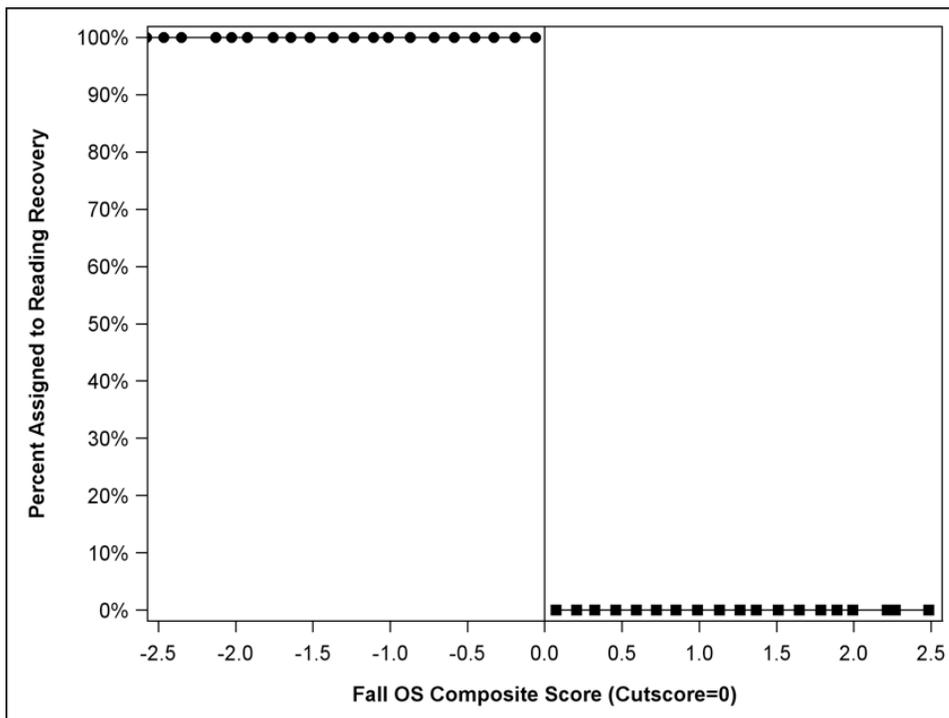


Figure 3.6. Sharp RD Piecewise Spline Regression of Mid-Year OS Reading Scores in Wave 1 from 2011-12 (274 Schools, 3,522 students)

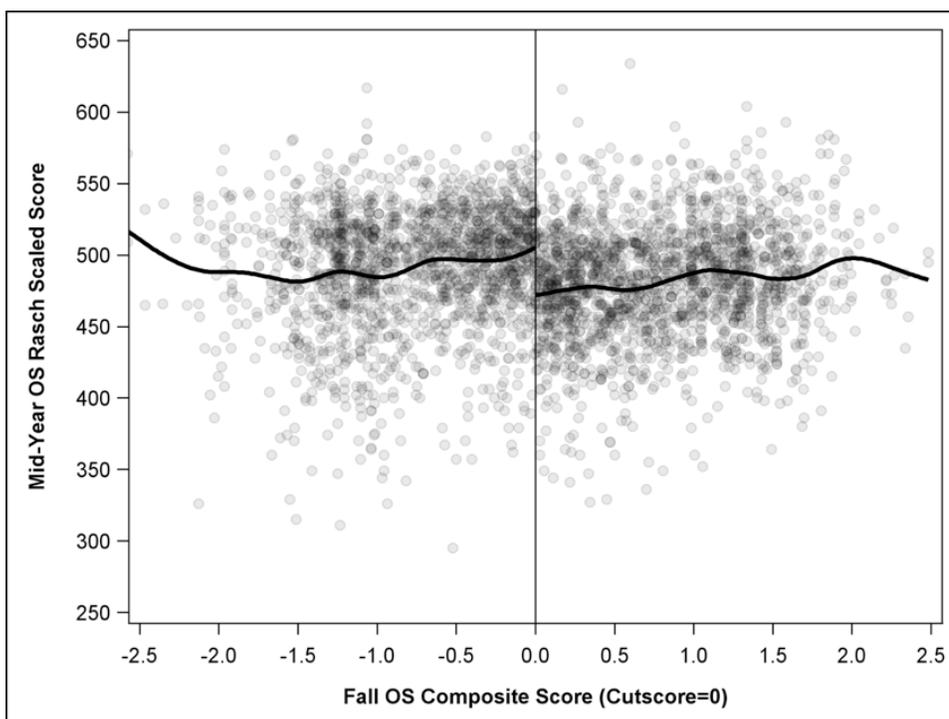


Table 3.4. Fuzzy RD Parameter Estimates for ITT Effects of Reading Recovery on Year-End 1st-Grade OS Reading Scores in Wave 2 from 2011-12

Dependent Variable: Mid-Year OS Reading Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	524.09	1.99	<.0001
Pretest (β_1)	10.67	1.16	<.0001
Treatment Effect (β_2)	12.04	1.94	<.0001
Random Effects			
School Intercept Variance (τ^2)	745.36	94.49	<.0001
School Pretest Slope Variance (ψ^2)	87.31	29.27	.0014
School Treatment Impact Variance (ξ^2)	158.48	85.83	.0324
School Intercept/Slope Correlation ($\rho_{\alpha,\lambda}$)	-0.12	0.15	.4553
School Intercept/Impact Correlation ($\rho_{\alpha,\varphi}$)	-0.38	0.15	.0114
School Slope/Impact Correlation ($\rho_{\lambda,\varphi}$)	0.06	0.34	.8639
Student-Level Residual Variance (σ^2)	710.01	21.21	<.0001

Note. Sample includes 263 schools and 2,980 students.

end OS reading scores for the students who participated in Reading Recovery during the second half of first grade. After restricting RD analyses to only those students for whom year-end OS scores were available, the sample of schools included in analyses for the second wave of the fuzzy RD was reduced from 325 to 263 (19.1 percent school-level attrition), and the sample of students was reduced from 3,525 to 2,980 (15.4 percent student-level attrition).

Figure 3.7 shows a piecewise spline regression confirming the discontinuity in treatment assignment at the cutscore for the 2011-12 Wave 2 RD. Compliance with intended treatment assignment was very good, with 94.1 percent of students with mid-year OS composite scores between -.25 and 0.0 assigned to the treatment, and 94.0 percent of students with mid-year OS composite scores between 0.0 and +0.25 not assigned to the treatment.

For those students below the RD cutscore, the median number of Reading Recovery lessons received was 48 (e.g., about 9.6 weeks), with only 18.1 percent of students receiving at least the minimum 12 weeks of lessons (i.e., 60 lessons). Considering 60 or more lessons as having received 100 percent of the intervention and less than 60 lessons as a percentage of the intended 60 (e.g., 30 lessons is 50 percent of the treatment), the average treatment dosage for students below the Wave 1 RD cutscore was 77.2 percent, suggesting incomplete delivery of the intended treatment during the second half of first grade in 2011-12. Qualitative results from May et al. (2013) suggest that this was largely the result of a late start in providing Reading Recovery, and that most students participating during the second half of first grade did not reach the required minimum of 12 weeks before the school year ended.

Figure 3.7. Fuzzy RD Piecewise Spline Regression of Assignment to Treatment in Wave 2 from 2011-12

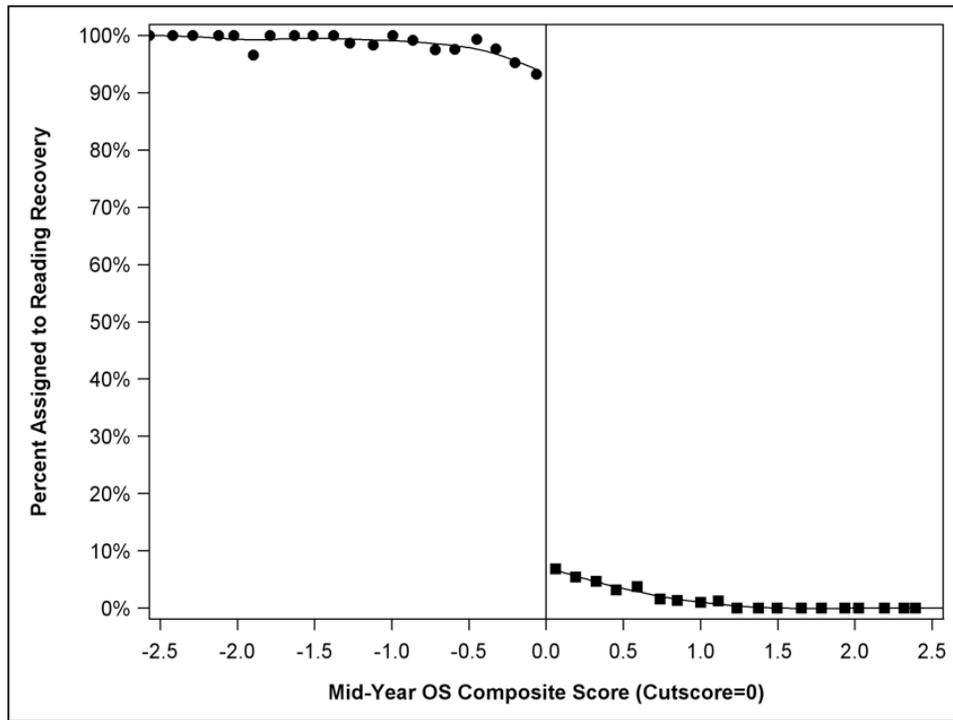
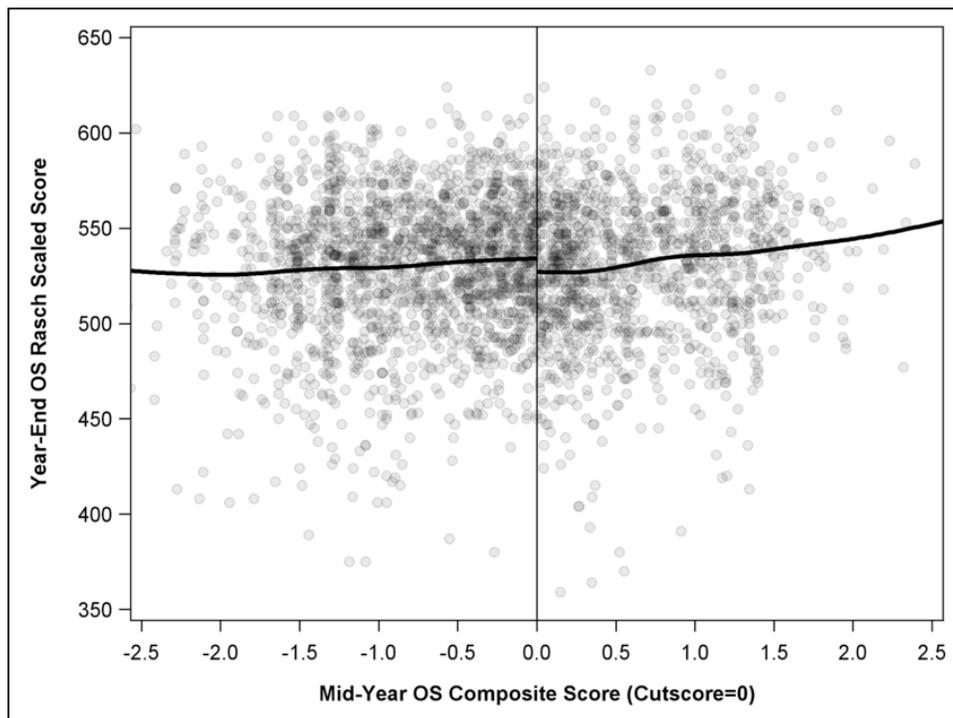


Figure 3.8. Fuzzy RD Piecewise Spline Regression of Mid-Year OS Reading Scores in Wave 2 from 2011-12 (263 schools, 2,980 students)



FINDINGS

Results from the HLM model of short-term impacts on OS scores at the end of first grade are shown in Table 3.4.

HLM analyses revealed an effect of +12 points on year-end OS scores. Dividing by the year-end 1st-grade population standard deviation of OS scores of 40.4 points yields an ITT effect size of .30 standard deviations. After adjusting for non-compliance near the cutscore, a CACE_c impact estimate on year-end OS scores was +.34 standard deviations. Although these impact estimates are smaller than those from RD Wave 1 (see pages 13-14), they are still substantial given that the vast majority of Wave 2 Reading Recovery students (81 percent) received only a partial intervention.

Figure 3.8 shows the discontinuity in the year-end OS Reading scores after Wave 2 of the 2011-12 RD study. The piecewise spline regression is linear, with no evidence of an interaction between the forcing variable and the treatment. Formal statistical tests found no interaction, no significant polynomial terms (up to the third degree), and no appreciable change in effect estimates when restricting bandwidth to within ± 1.0 or ± 0.5 standard deviations of the cutscore (raw effect estimates ranged from +10.9 to +13.4 points).

Results were similar after restricting analyses to the 231 schools and 2,544 students from the Sharp RD at Wave 2 (see Appendix for detailed results). The ITT impact estimate was .37 standard deviations, and sensitivity checks revealed no problems with model specification.

Estimating Sustained Impacts in Third Grade (Research Question 3)

Given that students in the comparison group during Wave 2 of the RD never received Reading Recovery, we can estimate long-term effects on state reading test scores in third grade and beyond. However, because collection of long-term outcomes data required extensive effort to locate each student in district databases and record 3rd-grade scores in IDEC, collection of long-term outcomes data under the original i3 Scale-Up grant was carried out only in eight states with the greatest number of schools participating in the i3 Scale-Up: Georgia, Iowa, Kentucky, Massachusetts, Ohio, South Carolina, and Texas. In these states, there were 85 schools who participated in the 2011-12 RD study (26 percent of the 331 that participated 2011-12).

Within these 85 schools, there were 1,069 students who were tested with the OS in the middle of first grade and assigned to Reading Recovery (n=638) or not (n=431). Ten schools were dropped from the analysis for failure to test any students not assigned to Reading Recovery. After restricting analyses to the remaining 75 schools and those students for whom 3rd-grade state test scores were available, the sample of students included in analyses for the second wave of the RD was reduced to 665 students, including 390 treatment students and 275 untreated students.

Table 3.5. Fuzzy RD Parameter Estimates for Long-Term ITT Effects of Reading Recovery on 3rd-Grade Reading/ELA State Test Scores

Dependent Variable: Mid-Year OS Reading Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	-0.613	0.091	<.0001
Pretest (β_1)	0.220	0.055	<.0001
Treatment Effect (β_2)	0.013	0.101	.8991
Random Effects			
School Intercept Variance (τ^2)	0.355	0.082	<.0001
School Pretest Slope Variance (ψ^2)	n/e		
School Treatment Impact Variance (ξ^2)	-0.024	0.037	.5031
School Intercept/Slope Correlation ($\rho_{\alpha,\lambda}$)	n/e		
School Intercept/Impact Correlation ($\rho_{\alpha,\varphi}$)	n/e		
School Slope/Impact Correlation ($\rho_{\lambda,\varphi}$)	n/e		
Student-Level Residual Variance (σ^2)	0.528	0.031	<.0001

Note. Sample includes 75 schools and 665 students. “n/e” is not estimable.

There were two sources of attrition at issue: one related to attrition of schools, which was considerable (i.e., down to 75 schools from the complete Wave 2 sample of 325 schools, or 77 percent attrition), and the other related to within-school attrition among the final sample of 85 and then 75 schools. If student-level attrition is calculated based on the complete Wave 2 sample of 325 schools, 2,260 treatment students, and 1,265 untreated students, then student-level attrition is 83 percent for the treatment group, and 78 percent for the untreated group. If student-level attrition is calculated by comparing the initial and final samples of students from only the sample of 85 and then 75 schools, then attrition at the student level is 38.9 percent for the treatment group, and 36.2 percent for the untreated group, and 37.7 percent overall (see Table 3.1).

Given the small sample size, statistical power for this analysis is somewhat low. Power calculations performed using PowerUp! version 01/22/2015 (Dong & Maynard, 2013) show a minimum detectable effect (MDE) of .33 standard deviations for this RD with 75 schools and 665 students (about 8 students per school), 59 percent of students assigned to the treatment group, an ICC of .25 and 8 percent of level-1 variance explained (estimated from the data collected), and value of .20 for treatment effect heterogeneity. This suggests that nearly all of the short term impact of Reading Recovery would need to be maintained on average across all schools in order to be detected in this analysis.

Not surprisingly, the HLM analyses of long-term outcomes under the intent-to-treat analyses (Research Question 3a) failed to yield a significant overall impact on 3rd-grade

Table 3.6. Fuzzy RD Parameter Estimates for Long-Term Effects of Successful Completion of Reading Recovery on 3rd-Grade Reading/ELA State Test Scores

Dependent Variable: Mid-Year OS Reading Scores	Estimate	Standard Error	p-value
Fixed Effects			
Intercept (β_0)	-0.606	0.091	<.0001
Pretest (β_1)	0.208	0.061	.0007
Treatment Effect (β_2)	0.187	0.111	.0924
Random Effects			
School Intercept Variance (τ^2)	0.326	0.074	<.0001
School Pretest Slope Variance (ψ^2)	n/e		
School Treatment Impact Variance (ξ^2)	n/e		
School Intercept/Slope Correlation ($\rho_{\alpha,\lambda}$)	-0.082	0.033	.0124
School Intercept/Impact Correlation ($\rho_{\alpha,\varphi}$)	n/e		
School Slope/Impact Correlation ($\rho_{\lambda,\varphi}$)	n/e		
Student-Level Residual Variance (σ^2)	0.537	0.035	<.0001

Note. Sample includes 75 schools and 555 students. “n/e” is not estimable.

state test scores ($ES=.01$; $p>.05$).⁶ Furthermore, the small sample size forced several parameters to be dropped from the model because they were inestimable (i.e., the random slopes, and all correlations between random effects). Detailed results from the HLM model for the Fuzzy RD of long-term impacts on 3rd-grade scores on state reading/ELA tests are shown in Table 3.5.

Figure 3.9 shows the piecewise spline regression of 3rd-grade reading/ELA state test scores (rescaled to a z-score scale). The piecewise spline regression is fairly linear except for the far left tail of the distribution, with no evidence of an interaction between the forcing variable and the treatment. Formal statistical tests found no interaction, no significant polynomial terms (up to the third degree), and no appreciable change in effect estimates when restricting bandwidth to within ± 1.0 or ± 0.5 standard deviations of the cutscore (raw effect estimates ranged from -0.17 to +0.04 points).

6. In an effort to maximize the size of the analytic sample, we also estimated the long-term effects model after combining both Wave 1 and Wave 2 samples as follows. Assignment to Reading Recovery is based on low Fall OS scores or low Mid-Year OS scores, given that a student has not yet participated in Reading Recovery. Using Rubin’s (1986) potential outcomes framework, the mid-year OS scores for Wave 1 Reading Recovery students had they not received the intervention are unobserved (only the post-intervention scores are observed). Therefore, we set the Mid-Year OS scores for Wave 1 Reading Recovery students to missing and used multiple imputation prior to estimating the composite OS scores predicting assignment to Reading Recovery at any point during the school year. The result is a Reading Recovery RD treatment group that includes all Wave 1 and Wave 2 students with available long-term test data ($n=752$), and an RD untreated group that includes all other students who were tested with the OS, and had available long-term test data, but never received Reading Recovery ($n=295$). This analysis also failed to yield a significant overall impact on 3rd-grade state test scores ($ES=-.10$; $p>.10$).

Figure 3.9. Fuzzy RD Piecewise Spline Regression of 3rd-Grade Reading/ELA State Test Scores (75 schools, 665 students)

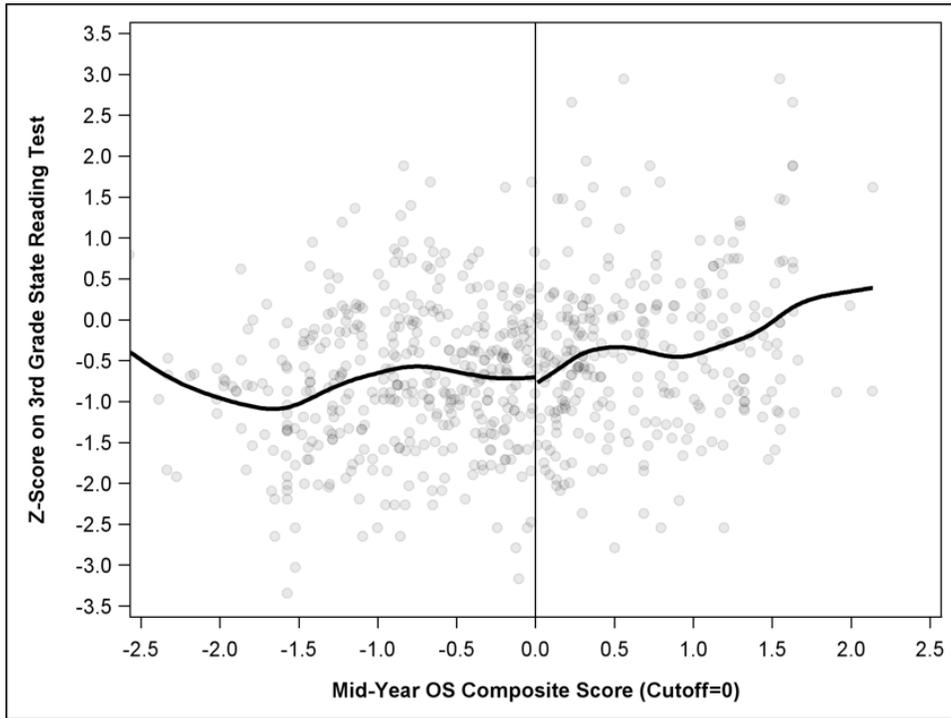
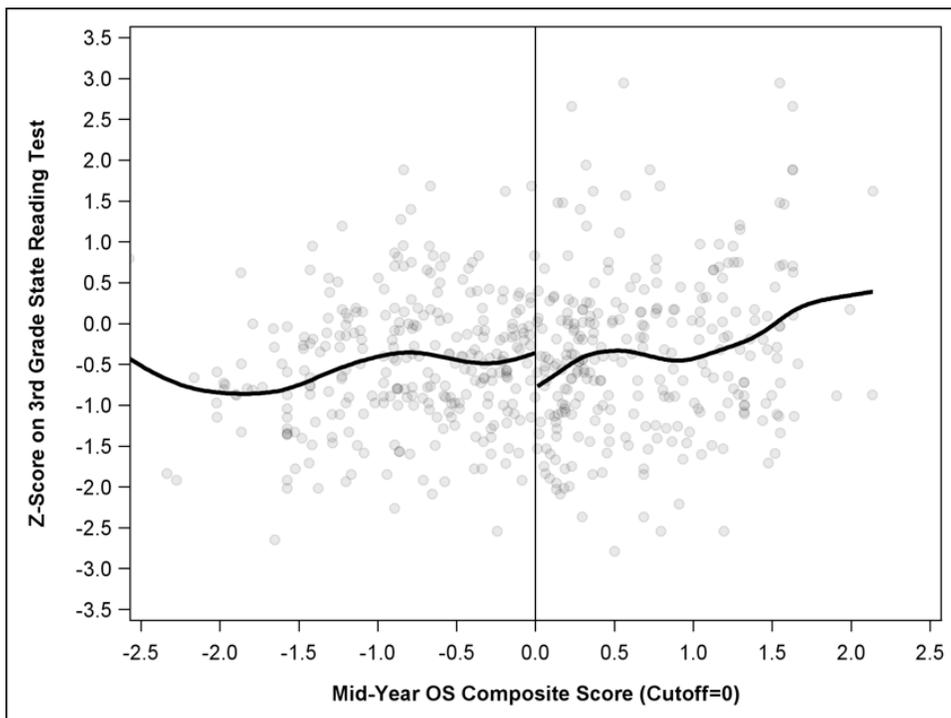


Figure 3.10. Fuzzy RD Piecewise Spline Regression of 3rd-Grade Reading/ELA State Test Scores for Successful Completion of Reading Recovery (75 schools, 555 students)



FINDINGS

Results were similar after restricting analyses to the 73 schools and 630 students included in the long-term Sharp RD from Wave 2 (see Appendix for detailed results). The ITT impact estimate was .04 standard deviations, and sensitivity checks revealed no problems with model specification.

After restricting the analyses to those students who successfully completed Reading Recovery and were not referred for additional intervention (Research Question 3b), the sample included 281 Reading Recovery students and 274 comparison students. HLM analyses of long-term outcomes for this subsample revealed a marginally significant overall impact on 3rd-grade state test scores ($ES=.19$; $p>.09$) for successfully “discontinued” students. Given that the state test scores are scales as z -scores, the impact estimate from the HLM model is already in standardized form. In other words, students who successfully completed Reading Recovery and were not referred for additional intervention had reading/ELA scores in third grade that were .19 standard deviations higher than their predicted scores under the RD design had they not participated in Reading Recovery.

Detailed results from the HLM model for the Fuzzy RD of long-term impacts on 3rd-grade scores on state reading/ELA tests are shown in Table 3.6.

Figure 3.10 shows the piecewise spline regression of 3rd-grade reading/ELA state test scores (rescaled to a z -score scale). The piecewise spline regression is fairly linear, with possible evidence of an interaction between the forcing variable and the treatment. A formal statistical test found a marginally significant interaction ($p=.07$), suggesting that the predictive relationship between OS scores in first grade and third grade reading/ELA scores was stronger for students not assigned to Reading Recovery. Further tests found no significant polynomial terms (up to the third degree), and no appreciable change in effect estimates when restricting bandwidth to within ± 1.0 or ± 0.5 standard deviations of the cutscore (raw effect estimates ranged from +0.02 to +0.33 points).

Because this analysis involves a restricted sample of Reading Recovery participants (i.e., only those who successfully discontinued and were not referred for additional intervention such as special education services), there is the risk of bias in the impact estimate due to sample selection. This is especially true given that the decision to refer a student to additional intervention is based on lack of progress (e.g., small gains on the OS) while in Reading Recovery.⁷ In this analysis, the sample of Reading Recovery students was reduced by 28 percent, from 390 to 281. To place a lower bound on the impact estimate and determine the worst-case-scenario of the amount of selection bias, we re-estimated the model after deleting 28 percent of the comparison group with the smallest gains

7. Re-estimation of the Wave 1 short-term impacts confirms likely selection bias for the subsample of successful Reading Recovery completers in that the estimated impact is 34.87 points, which is 56% larger than the 22.41 point impact estimate for the full sample. However, analysis of gains on the OS during first-grade suggest that successful completion of Reading Recovery (i.e., discontinuing) is related to but not perfectly predicted by gains on the OS—40% of the excluded treatment children were in the bottom 28%, while 21% of the included treatment children were in the bottom 28%. This corresponds to an odds ratio of 2.5, suggesting that students with small OS gains were 2.5 times less likely to be discontinued.

on the OS during first grade. This approximates the effect of “creaming” the sample by retaining only the students with the largest gains during first grade. The results of this sensitivity analysis produced an impact estimate that while not statistically significant, was still positive and more than half the magnitude of the original effect ($ES=.10$; $p>.40$). This suggests that the observed effect of .19 standard deviations is unlikely to be entirely attributable to selection bias.

CONCLUSIONS

The amount of data currently available for RD analyses of long-term effects is quite limited in comparison to that available for the short-term effects RD analysis (see Table 3.1) or the RCT (see Chapter 2). As a result, the findings of this single-year RD study are inconclusive with regard to third-grade effects. However, the RD study was able to replicate reasonably well the short-term effect estimate observed in the large-scale RCT conducted in a separate sample of i3 schools. This suggests that the RD study as implemented in CPRE/CRESP’s evaluation has strong potential for valid causal inference.

While the intent-to-treat analyses for Research Question 3a showed no significant effects, it would be incorrect to conclude that Reading Recovery has no long-term effects. At this time, there simply are not enough students with 3rd-grade state test scores recorded to produce a precise impact estimate or perform subgroup analyses, and the true sustained effect may very well be quite substantial. Additionally, attenuation in the impact is likely given that the outcomes being measured are 3rd-grade state reading test scores, which are distal relative to the Observation Survey measure (the RCT found consistently larger impacts on the OS than on the ITBS test). It would also be incorrect to conclude that the positive results presented here from the subgroup analyses for Research Question 3b prove that Reading Recovery *does* produce long-term impacts, since the strength of causal inference for that result is weakened by potential selection bias.

The long-term outcomes data available for the analysis discussed here are a small fraction of what might be collected for the cohorts of students who reached third grade after data collection activities for the i3 project ended. An efficacy follow-up study is planned in which we will collect long-term follow-up data on state test scores for the full sample of students participating in the RD study from 2011-2015, including more than 8,000 students who received Reading Recovery during RD Wave 2, and over 5,000 students in the long-term comparison group. This long-term efficacy follow-up study will provide important information about the overall long-term impacts of Reading Recovery under i3 and the malleable factors that may explain school-level variation in the persistence of effects.

Reading Recovery

Implementation Fidelity:

Scale-Up Successes and

Challenges

This chapter discusses the implementation of Reading Recovery during the i3 scale-up. Specifically, it examines the degree to which participating schools implemented the intervention with fidelity to the Reading Recovery model. The research described here was part of the independent evaluation conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania and the Center for Research on Education and Social Policy (CRESP) at the University of Delaware. The independent evaluation was funded by a 2010 i3 Scaling Up What Works grant to The Ohio State University (OSU) for the expansion of Reading Recovery in the U.S.

In each year of CPRE/CRESP's independent evaluation of the i3 scale-up of Reading Recovery, a key research objective was to measure fidelity of implementation to the program's model. Thorough assessment of implementation fidelity is increasingly regarded as an integral component of program evaluation (Abry, Hulleman, Rimm-Kaufman, 2014; Gerstner & Finney, 2013; Pas & Bradshaw, 2012; Nelson et al., 2012; Hulleman & Cordray, 2009), as information about program effects is of limited use without an understanding of how the effects were achieved. Evidence of adherence to or departure from a program's model also provides important context when program effects are absent (Dhillon, Darrow & Meyers, 2015; Dusenbury et al., 2003).

CPRE and CRESP observed high fidelity to Reading Recovery's program model in all years of the evaluation. This finding suggests that the impacts we observed were, indeed, the result of faithful implementation of the intervention, and that the i3 scale-up successfully replicated Reading Recovery in the schools involved in the expansion. This chapter presents our key findings in this area.¹

1. Additional detail on fidelity, and on the Reading Recovery program model, is reported in CPRE's Year One report (May et al., 2013) and a recent article that appeared in the *American Educational Research Journal* (May et al., 2015).

DEFINING IMPLEMENTATION FIDELITY

Despite broad consensus about the importance of assessing implementation fidelity, recent studies have documented a lack of coherence in the education literature about how fidelity is best defined and measured (Dhillon, Darrow & Meyers, 2015; Century, Cassata, Rudnick, & Freeman, 2012; Century, Rudnick, & Freeman, 2010; Mowbray et al., 2003). As a result, it often falls to the evaluator to grapple with questions about how fidelity should be understood and assessed in the context of a given intervention (Goodson, Price, & Darrow, 2015; Century, Rudnick, & Freeman, 2010; Summerfelt, 2003).

Our understanding of implementation fidelity in the context of Reading Recovery is informed by the *Standards and Guidelines of Reading Recovery in the United States, 6th Ed.* (Standards and Guidelines) (Reading Recovery Council of North America, 2012). The *Standards and Guidelines* codify Clay's understandings about the activities and practices that constitute adherent implementation of the Reading Recovery program (Clay, 2005). The *Standards and Guidelines* are structured according to the key implementer roles in Reading Recovery: teacher leader, teacher, site coordinator, and university trainer. For each of these roles, the *Standards and Guidelines* delineate specific required activities related to the delivery of lessons, the training and support of teachers, and/or the establishment and maintenance of Reading Recovery sites.

While the *Standards and Guidelines* provide the framework for our research on implementation fidelity, we observed that not all of the activities it delineates are central to the delivery of the program—some pertain primarily to program growth and sustainability. As a result, we focused our fidelity study on the 51 program standards that are related to the core activities of Reading Recovery: the training of teachers and the delivery of one-to-one lessons. These 51 program standards—which are the focus of our fidelity study—met the following criteria:

1. Clearly stated as program standards in the *Standards and Guidelines*;
2. Essential to the core functions of Reading Recovery: the training of teachers and the provision of one-to-one lessons; and
3. Performed by core Reading Recovery personnel: university trainers, teacher leaders, and Reading Recovery teachers.

The definition of implementation fidelity as the adherent completion of these 51 program standards provided the framework for the collection and analysis of fidelity data.

APPROACH FOR EVALUATING FIDELITY OF IMPLEMENTATION

The approach to evaluating fidelity of implementation in the i3 scale-up of Reading Recovery involved a five-step process:

1. Construct a logic model for Reading Recovery i3 implementation, which groups the 51 standards that are related to the core activities of Reading Recovery into several Key Components.
2. Operationalize each of the 51 program standards as measurable Program Indicators.
3. Define a minimum threshold by which to assess adequacy of implementation for each of the Program Indicators; define a minimum threshold by which to assess fidelity of implementation for each of the Key Components.
4. Collect data on each Program Indicator directly from implementers.
5. Calculate adherence to the Program Indicators and determine adequacy of implementation for each; calculate number of Program Indicators with adequate implementation within each Key Component and determine fidelity of implementation for each.

Steps 4 and 5 were repeated in each year of the evaluation. This approach was informed by the NEi3 framework for high-quality implementation studies, and it is consistent with recent literature on implementation fidelity, both within and outside education research (Dhillon, Darrow & Meyers, 2015; Goodson, Price, & Darrow, 2015; Laurentson, Oh, & LaBlanca, 2015).

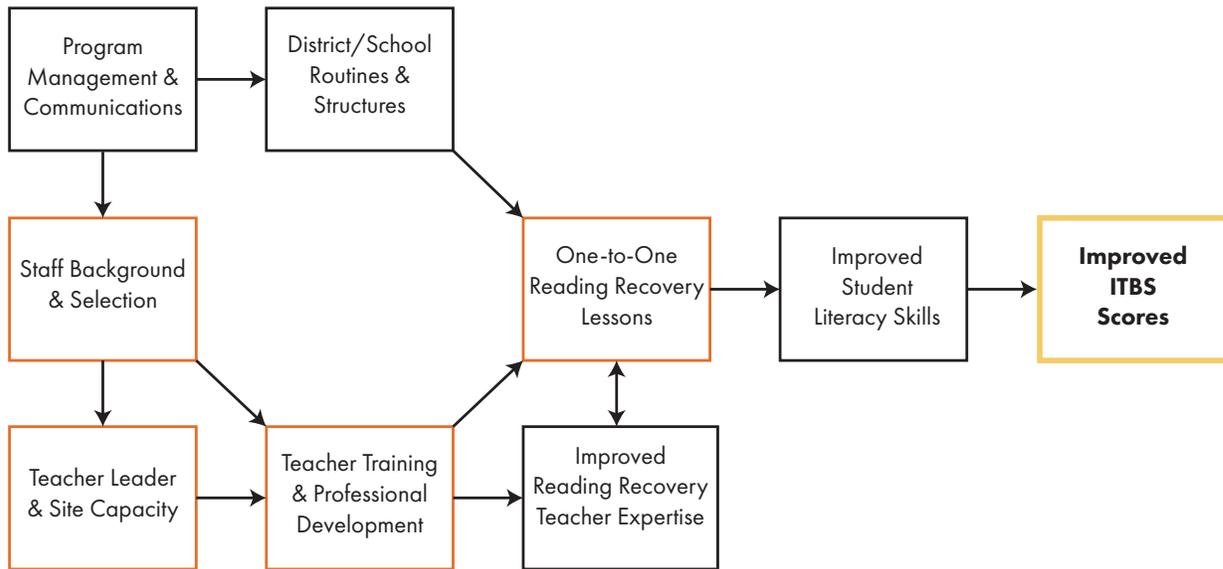
Logic model for Reading Recovery i3 implementation

In Years One and Two of the evaluation, we constructed a logic model, grouping the 51 program standards into four Key Components:

- » **Staff Background & Selection:** This component includes standards that specify the selection criteria for teachers trained in Reading Recovery and teacher leaders.
- » **Teacher Leader & Site Capacity:** This component includes standards that specify the training experience of teacher leaders as well as the standards that characterize the training environment.
- » **Reading Recovery Teacher Training & Ongoing Professional Development:** This component includes standards that specify the training and continuous professional development experience of trained and in-training Reading Recovery teachers.
- » **One-to-One Reading Recovery Lessons:** This component includes standards that specify the selection, assessment, and instruction of individual Reading Recovery students.

These Key Components enabled us to represent, as concisely as possible, the complex set of required activities reflected in the *Standards and Guidelines*. Figure 4.1 shows the Logic Model for Reading Recovery i3 Implementation that CPRE constructed to specify

Figure 4.1: Logic Model for Reading Recovery i3 Implementation



the relationships between these four Key Components of fidelity (outlined in red) and the primary program outcome—student achievement—via two mediators: improved Reading Recovery teacher expertise and improvements in student literacy.

As Figure 4.1 illustrates, staff background and selection activities are expected to facilitate both teacher leader & site capacity and Reading Recovery teacher training and professional development. In practice, this would mean that qualified individuals are selected for key roles related to the training of teachers and for the provision of one-to-one instruction to students. Training and ongoing professional development are expected to support both a teacher’s capacity to conduct the lesson and the development of that teacher’s expertise in early literacy and formative assessment. As such, the one-to-one Reading Recovery lessons are depicted as both a direct outcome of the training and professional development, and an indirect outcome of training via the effect on improved teacher expertise. Improved teacher expertise is also supported through ongoing implementation of the Reading Recovery lessons; this is specified by a reciprocal arrow in Figure 4.1. Improved teacher expertise is expected to mediate the relationship between training and provision of instruction. Participation in the one-to-one lessons is expected to produce improvements in students’ literacy, which in turn produces improvements in ITBS scores. Thus, the Logic Model for Reading Recovery i3 Implementation shown above represents our understanding of how these broad categories of activities—each encompassing a number of inputs and activities—progress toward and directly support a single, prioritized outcome: gains in student achievement as measured by the ITBS. A complete list of the program standards encompassed in the Logic Model for Reading Recovery i3 Implementation is included in Appendix C.

Indicators of Program Implementation

The 51 *Standards and Guidelines* related to the core activities of Reading Recovery were each operationalized as clear and measurable Program Indicators. All Program Indicators were assessed using at least one fixed-response survey item administered to Reading Recovery teachers and teacher leaders involved with the i3 scale-up. Several precautions were taken to avoid problems related to self-reported data; for instance, when measuring program activities, respondents were asked to report objective facts about Reading Recovery implementation in their context, not to make evaluative judgments about themselves or their own contributions. In addition, questions were carefully worded so as not to suggest that one response was correct or more appropriate than the others.

Defining Implementation Adequacy and Fidelity

Adequate implementation of individual Program Indicators was defined in advance as 80 percent adherence to the full definition of the indicator. This criteria is consistent with recent literature on the measurement of fidelity in education research (Balu & Quint, 2015). As noted, Reading Recovery teachers and teacher leaders are bound by different standards during training years. To account for these differences, CPRE calculated the adequacy index for each of the Program Indicators (i.e. percent of implementers fully meeting the standard) using only relevant respondent data based on his/her specific position and stage in the training process. To calculate fidelity for selection and training of teachers, we used data collected from teachers who were in their training year². A majority of the new i3 teacher leaders were trained in the first year of the scale-up. To account for the small numbers of i3 teacher leaders in later years of the scale-up, adequacy of Program Indicators for selection and training was calculated as a cumulative measure across all i3 teacher leaders.

A final adjustment was made to account for waivers. The North American Trainers' Group (NATG), the governing body of Reading Recovery in the U.S., has developed a waiver process by which university trainers can exempt sites from meeting particular standards with proper justification. The waiver process requires that a teacher leader make a formal, written request to deviate from a particular standard. For example, to meet Standard 2.01, a teacher leader teaching a training class of fewer than 8 or more than 12 Reading Recovery teachers must first request and obtain a waiver from her or his affiliated university training center. A university trainer determines whether the deviation is permissible and notifies the schools and regional training center of the allocation or refusal of a waiver. If granted, the waiver remains in effect for a single year. CPRE surveys asked respondents to specifically indicate when a waiver had been granted for a given indicator, and non-compliance with standards was not considered a deviation from fidelity when waivers had been approved.

2. Our Year One and Year Two reports (May et al., 2013; 2015) did include some retrospective self-report questions about training experiences that were administered to all survey participants. However, for the final analysis reported here, we restrict the data to include only teachers and leaders for whom each standard *currently* applies. This results in fewer respondents being included and some differences with previous reports for some indicators.

Fidelity to Key Components was defined in advance as a minimum number of Program Indicators that were found to have adequate implementation in each year. The minimum threshold for Key Component fidelity was defined as having not fewer than 80 percent of the number of the constituent Program Indicators with adequate implementation. Table 4.2 presents the total number of Program Indicators in each component and the number of indicators with adequate implementation needed to meet the standard of fidelity. Failure to demonstrate adequate implementation of the minimum number of Program Indicators would reflect a lack of fidelity for a given Key Component in a given year.

Collecting implementation data

The surveys were administered via Qualtrics™, a secure online survey platform. Web links were emailed to individual Reading Recovery teachers and teacher leaders. Survey instruments were individualized to ensure that the appropriate Program Indicators were administered (i.e. the correct standards were assessed) for each respondent. For example, Reading Recovery teachers and teacher leaders who are in training are expected to adhere to different standards than those who are already trained (e.g., Reading Recovery teachers in their training year are required to attend weekly classes, whereas trained teachers meet only occasionally). Teachers in training therefore received items pertaining to the relevant standards for them, and those who were already trained received different items.

READING RECOVERY TEACHER SURVEY

In May of each year, all i3-supported Reading Recovery teachers were asked to complete an online survey about their experiences during the current school year.³ The survey included a combination of fixed-response and open-ended items, and was designed to explore multiple components of the i3 scale-up and Reading Recovery implementation. These included: the training Reading Recovery teachers received; implementation of Reading Recovery in one-to-one lessons with students; fidelity to the standards and guidelines; use of curricular materials; and the nature and frequency of communication with Reading Recovery teacher leaders, school and district administrators, students' parents, other teachers at the school, and other Reading Recovery teachers.

READING RECOVERY TEACHER ACTIVITY LOGS

In the first three years of the evaluation, the full population of i3-supported Reading Recovery teachers was asked to complete daily activity logs on randomly selected days from January through April.⁴ The daily log included the question “Did you conduct any one-to-one Reading Recovery lessons today?” Only those Reading Recovery teachers who

3. In Year Four, Reading Recovery teachers were surveyed only in schools that were randomly selected from all i3 schools to participate in the RCT.

4. In the first year, 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. In the second and third year, all teachers received a total of five activity log invitations to be completed on five assigned days.

responded that they had conducted a Reading Recovery lesson with at least one student were asked to complete questions about the activities that took place during each lesson.⁵ These questions asked the teacher to report whether or not each lesson included key structural components of Reading Recovery lessons (i.e., Roaming Around the Known, Reading Familiar Books, Administering a Running Record, Word and/or Letter Work, Story Writing, Assembly of a Cut-up Story, Introducing a New Book, Reading a New Book, Creating/Updating a Lesson Record, Creating/Updating a Prediction of Progress, Creating/Updating a Record of Reading Vocabulary, Creating/Updating a Record of Writing Vocabulary, or Creating/Updating a Record of Text Level). These data were used to calculate the frequency with which Reading Recovery lessons were conducted following the expected format of lessons.

READING RECOVERY TEACHER LEADER SURVEY

In May of each year, all teacher leaders were asked to complete an online survey about their experiences during the current school year. The teacher leader survey focused on systems and patterns related to the support and oversight of the i3 scale-up. In particular, fixed-response and open-ended items on the survey examined respondents' experiences with key scale-up activities, including: the training and ongoing professional development provided to Reading Recovery teachers; implementation of Reading Recovery in one-to-one lessons with students; fidelity to the standards and guidelines; and the exchange of resources and data with UTCs, Reading Recovery teachers, and other teacher leaders. Teacher leaders were also asked to provide information on attendance at training sessions for each of the Reading Recovery teachers they trained, and on the content and pedagogy of the training sessions. These data were used to measure fidelity to the standards for Reading Recovery teacher training.

SURVEY SAMPLE SIZES AND RESPONSE RATES

Sample sizes and response rates for each of the four years of the evaluation are shown in Table 4.1. In all but Year Four, the samples for each survey included all Reading Recovery teachers and teacher leaders involved with the implementation of the program in schools in the scale-up. In Year Four, surveys were again distributed to all teacher leaders, but because of increased numbers of Reading Recovery teachers in the scale-up overall, the teacher survey was distributed only to teachers in RCT schools, which is a random sample of i3 schools (see *Chapter 2: Immediate Impacts of Reading Recovery*). Samples for the surveys were drawn from population data provided to CPRE by IDEC.

The response rates for activity logs in Table 4.1 reflect the percent of Reading Recovery Teachers completing at least one log. Of these respondents in each of the three years of administration, most respondents completed the majority of the logs. The percent that

5. In Year Two, teachers were asked to report on activities during one randomly sampled Reading Recovery lesson for that day, while in Years Three and Four, teachers were asked to report on activities during all of their Reading Recovery lessons for that day.

completed only one log ranged from 13 percent to 16 percent, the percent that completed two logs ranged from 9 percent to 17 percent, the percent that completed three logs, ranged from 12 percent to 23 percent, the percent that completed four logs, ranged from 17 percent to 44 percent, and the percent completed five logs ranged from 43 percent to 46 percent (note that only four logs were administered in 2011-12).

Determining Adequacy and Fidelity of Implementation

The final step in evaluating fidelity of implementation in the i3 scale-up of Reading Recovery is to (a) calculate adherence to the Program Indicators and determine adequacy of implementation for each; and (b) calculate the number of Program Indicators with adequate implementation within each Key Component and determine fidelity of implementation for each.

After response data on implementation fidelity were collected, we examined the extent to which Program Indicators had *adequate* or *inadequate* implementation by calculating the percentage of respondents for whom the indicator was met (i.e. full adherence to applicable standard)⁶. As a final step, we assessed the fidelity of implementation for each of the four Key Components, separately in each year. To make a determination about fidelity, we evaluated the number of implementation Fidelity Indicators within a given Key Component that met the benchmark of adequacy (i.e. 80 percent adherence among respondents for whom the indicator represented an applicable standard). If, in a given year, a Key Component had the minimum number of indicators with adequate implementation (i.e. 80 percent of Fidelity Indicators within a Key Component), then the Key Component as a whole was found to be implemented with fidelity. All indicators were weighted equally in determining whether or not a Key Component was implemented with fidelity.

Table 4.2 presents the number of Program Indicators that were adequately implemented by Key Components. In Year One, five of the 51 indicators were not assessed reliably, so the total and minimum of adequately implemented indicators were adjusted to maintain an 80 percent minimum threshold. As Table 4.2 reveals, Reading Recovery was implemented in the i3 scale-up with high fidelity to the *Standards and Guidelines*. Overall, 90 percent of the Program Indicators used to assess implementation had adequate implementation. Furthermore, all four Key Components represented in the Logic Model for Reading Recovery i3 Implementation (Figure 4.1) were found to be implemented with fidelity. In some areas, we consistently observed very high fidelity rates of 100 percent or nearly 100 percent. Appendix Tables C1 and C2 details fidelity findings by individual program standard, and includes information about the survey instruments used.

6. Five of the 51 Fidelity Indicators were not reliably assessed in Year One due to inadvertently vague wording of the survey questions.

Table 4.1. Survey Response Counts and Rates

	Y1		Y2		Y3		Y4		All	
	n	%	n	%	n	%	n	%	n	%
Teacher Leader	169	89%	221	85%	215	74%	160	81%	765	82%
Reading Recovery Teacher Survey	813	79%	1511	76%	1567	74%	696	70%	4587	75%
Reading Recovery Teacher Log	882	78%	1373	75%	2005	69%	n/a	n/a	4260	73%

Table 4.2. Indicators with Adequate Implementation Required for Fidelity

Key Component	Indicators per Component	Indicators required for fidelity	Y1	Y2	Y3	Y4	Met fidelity benchmark all 4 years
Staff Background & Selection	7	6	7	6	6	6	Yes
Teacher Leader & Site Capacity	21	17	21	20	21	21	Yes
Reading Recovery Teacher Training & Ongoing PD	15 (12)	12 (10)	11	14	14	14	Yes
One-to-One Reading Recovery Lessons	8 (6)	7 (5)	6	7	7	7	Yes
All	51 (46)	N/A	45	46	48	48	N/A

Note. Parenthetical numbers are for Year One in which fidelity for five indicators was not measured.

Deviations from implementation fidelity

Deviations from the program model were observed in a few specific areas, as follow:

- » *Commitment to Implementation.* A critical factor that can influence school-level implementation is the school system's commitment to Reading Recovery. In the Teacher Leader and Site Capacity Key Component, Standard 3.01 applies to Reading Recovery teachers: Be employed in a school system that has a *commitment to implementation* (RRCNA, 2012, p. 9). CPRE/CRESP operationalized this standard as a school's commitment to training enough Reading Recovery teachers to reach "full implementation," Reading Recovery's term for the capacity to provide Reading Recovery to all students who need it. To measure fidelity to this standard, we asked teachers whether their schools (a) was currently at full implementation (i.e. every 1st-grade student who needed Reading Recovery received the intervention), or (b) had a plan to train more Reading Recovery teachers. When teachers responded positively to either question, the standard was considered met. This standard was not met in Years Two, Three, and Four. Our qualitative research suggests that many schools in the scale-up fell short of full implementation because of the large numbers of students who need intervention.
- » *Trainer support of teacher leaders.* Standard 4.66 states that trained teacher leaders should *receive a minimum of two site visits from a trainer during the teacher leader's first year(s) in the field* (RRCNA, 2012, p. 21). Our qualitative research found great variation in the ways university trainers support and supervise teacher leaders, and varying degrees of contact between teacher leaders and their trainers. This finding is consistent with the relatively lower fidelity rates on this standard. The percent of teacher leaders reporting deviations from fidelity to this standard fell below the 80 percent benchmark only in Year Two, when it was 78 percent.
- » *Ongoing Professional Development.* Also in the Key Component of Teacher Training and Professional Development, Standard 3.44 applies to trained Reading Recovery teachers: *Participate in a minimum of six professional development sessions each year, including a minimum of four behind-the-glass sessions with two lessons each session* (RRCNA, 2012, p. 12). Reading Recovery teachers reported deviation from fidelity on this standard in every year of the study; the percentage of teachers reporting adherence ranged from 66 percent in Year Four to 76 percent in Years One and Two.
- » *Student selection.* An important issue in school-level implementation of Reading Recovery is how students are selected to receive the treatment. In the One-to-One Reading Recovery Lessons Key Component, Standard 2.05 applies to sites: *Select the lowest-achieving children for service first (based upon Observation Survey [OS] tasks) in all decisions.* Based on survey data, Reading Recovery teachers were found to have inadequate implementation of this standard in Year Two; 78 percent of teachers reported that children with the lowest scores on the OS were

selected for service first. As reported in our Year Two report, CPRE has observed that student selection is a variable process across schools and districts—one that sometimes results in students begin selected for or excluded from Reading Recovery for reasons other than their OS scores. For instance, as reported in the Year Two report, some schools exclude students from receiving Reading Recovery based on special education status or poor attendance records. However, we observed better than 80 percent fidelity to this standard in Years Three and Four of the evaluation (84 percent and 88 percent, respectively). We attribute this increase in fidelity in part to a concerted effort on the part of Reading Recovery leadership—in response to the finding of non-adherence to this standard in our Year Two report—to ensure that guidance to schools and Reading Recovery personnel about student selection is clear and consistent. In 2015, program leadership adopted and documented formal guidance for schools and Reading Recovery personnel on how students should be selected for Reading Recovery, and the circumstances under which a student should be excluded. We regard this as an example of Reading Recovery’s ability to learn from and improve in response to the scale-up process and the findings of the evaluation.

- » *Timing of the start of lessons.* Standard 2.06, in the One-to-One Lessons Key Component, states that Reading Recovery teachers should *administer the Observation Survey and begin service to children within 2 weeks of school opening at the beginning of the year* (RRCNA, 2012, p. 7). We observed deviation from this standard in Years Three and Four of the evaluation. Our qualitative research reveals that Reading Recovery teachers often have additional responsibilities related to assessing students and supporting teachers at the beginning of the school year. These responsibilities can make it difficult for teachers to complete the administration of the OS and begin lessons within two weeks of the start of school. Roughly 25 percent of Reading Recovery teachers surveyed reported that they were unable to start providing lessons within the first two weeks of school. Analysis of IDEC student records indicated that most students (75 percent) began within the first month of school.

A full listing of the program standards included in the fidelity analysis, by Key Component, is included as Appendix Table C1. Table C2 presents corresponding adequacy findings for each year of the evaluation.

QUALITATIVE FINDINGS

The analysis of survey data was our primary method for assessing implementation fidelity in the Reading Recovery scale-up. However, as the discussion of deviations from fidelity above reveals, the evaluation also included a significant qualitative component that produced additional insights on fidelity. This qualitative effort included 334 interviews with implementers, more than 70 observations of Reading Recovery lessons, and review of 50 Reading Recovery lesson records over the course of the evaluation.

The findings from our mixed-methods research on implementation are detailed at length

in the Year One and Year Two reports (May et al., 2013, 2014). In addition, Chapters 5 and 6 of this report focus on particularly important findings from the qualitative research. However, we briefly summarize our major qualitative findings as they pertain to implementation fidelity below.

Training

The training of new Reading Recovery teachers and the support of trained teachers are core activities in the Reading Recovery program model, and training activities are specified in detail in the *Standards and Guidelines*. Reading Recovery teachers reported that their training year was intense and rigorous. They consistently reported that training prepared them well for implementation and working with Reading Recovery students. Many teachers described Reading Recovery training as transformative for their own understanding of literacy; one teacher described it as “a complete shift in thinking.” Teachers spoke particularly positively about the behind-the-glass teaching component of training, citing it as a powerful exercise that allowed them both to receive feedback on their own teaching and to learn from watching others.

Most teachers—both trained and training—expressed appreciation for the support they received from their teacher leaders whom they regarded as important and accessible sources of information about literacy instruction and as supports for their own professional growth. Expressing a common sentiment, one teacher described her teacher leader as “extremely supportive. She’s always available by phone or by email. I just feel that anything that I need help with there’s always someone to answer and assist me.”

Many Reading Recovery teachers reported traveling long distances to their weekly classes. In addition, teacher leaders reported having to travel considerable distances to adhere to training standards. “I single-handedly support a site of 50 teachers in a region of 3,000 square miles,” explained one teacher leader. The scale-up process exacerbated this geographic challenge for teacher leaders in some regions, as new teachers recruited under i3 were sometimes located at considerable distance from regional training centers.

Communication

The *Standards and Guidelines* require that Reading Recovery 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” (RRCNA, 2012, p. 11). Qualitatively, we observed that teachers and teacher leaders generally work to ensure that this communication takes place, with mixed results. Reading Recovery teachers reported frequent contact with 1st-grade teachers, mostly about Reading Recovery students’ progress. However, many wished for more time or more formal opportunities to talk with their students’ classroom teachers. A minority of Reading Recovery teachers reported that they had explicit leadership roles in their schools, or that they regularly shared information about literacy instruction with other teachers. Most Reading Recovery teachers we interviewed reported having a role in their schools’ Response to Intervention

or referral processes. The extent to which Reading Recovery data were utilized in these processes varied.

Teacher leaders expressed a range of experiences in communication with principals and/or district officials about Reading Recovery, and described varying levels of influence over decisions affecting Reading Recovery implementation at the school level. A minority of teacher leaders reported providing professional development and support to all staff, not just Reading Recovery teachers, and/or having input in choice of curricula or other school-wide instructional decisions. Others reported having very little influence or contact with non-Reading Recovery staff. Some of the observed variation in student selection practices and the timing of the beginning of one-to-one lessons may be in part a function of these differences. Teacher leaders with limited access to principals often reported that it was difficult to ensure that student selection was implemented correctly, or that lesson schedules were closely followed. (For more on this issue, see Chapters 5 and 6.)

Lesson Delivery

The *Standards and Guidelines* include specific guidance on the allocation of Reading Recovery teachers' time. Adherent implementation requires that teachers provide no more and no fewer than four lessons per day. Our qualitative findings about Reading Recovery teachers' job structures were consistent with the results of our teacher log analysis. A majority of the Reading Recovery teachers interviewed reported that they provided one-to-one Reading Recovery lessons for half the day and fulfilled other roles in their schools during the other half. Typically, their secondary roles involved small-group instruction or coaching. Infrequently, Reading Recovery teachers doubled as classroom teachers during the rest of the day. Teachers often spoke of the many demands on their time and the difficulty of finding adequate time to plan for lessons or meet with colleagues. "I would love to do [Reading Recovery] full time and not be the instructional coach," one teacher explained. "But we don't have any positions like that in our district. I feel very split in having [to be both] Reading Recovery and an instructional coach."

Qualitative findings also corroborate our survey finding of high fidelity in the area of treatment dosage: The provision of a lesson to every Reading Recovery student every school day is an important component of the program model. Teachers described frequent obstacles to adhering to the schedule of daily one-to-one lessons with each student. For instance, student availability and Reading Recovery teachers' other responsibilities can interfere with Reading Recovery lessons. However, most found ways to surmount these obstacles. We heard from many respondents about Reading Recovery teachers who came in early, stayed late, or altered their schedules to ensure that students received their daily lessons.

DISCUSSION

CPRE/CRESP's evaluation of implementation fidelity represents a rigorous, sustained, systematic, and successful attempt to measure and assess wide-ranging implementation activities. While our measurement was done via statistical analysis of survey data, the qualitative inquiry served both to corroborate the survey findings and to shed light on the few departures from fidelity that were observed. A consistent finding is that Reading Recovery was implemented with high fidelity to the program model over the course of the i3 scale-up. We found that the intervention was delivered as designed to the students in the scale-up, and that teachers delivering Reading Recovery lessons were properly trained. All in all, this finding suggests that the impacts we observed were, indeed, the result of faithful implementation of the intervention, and that the i3 scale-up successfully replicated Reading Recovery in the schools involved in the expansion.

This evaluation of implementation fidelity is also important for what it does not offer. We discovered that there were some important questions about Reading Recovery's implementation that our fidelity analysis could not answer. For example, while the effects of Reading Recovery overall were positive and medium to large (See *Chapter 2: Immediate Impacts of Reading Recovery*), we observed considerable variation in school-level impact estimates that is not explained by our main finding that the program was consistently implemented with high fidelity across the scale-up. Thus findings from this analysis offer no clues as to why some Reading Recovery implementations appear to be more effective than others. In addition, our examination of fidelity yielded no insights on program sustainability—the extent to which the schools that implemented Reading Recovery under the i3 grant are likely to maintain use of the program. These questions are relevant to our inquiry of what makes Reading Recovery successful, and are addressed in the following two chapters.

Lesson-Level Implementation: Instructional Strength in Reading Recovery

This chapter examines the lesson-level implementation of Reading Recovery as part of an independent evaluation conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania and the Center for Research on Education and Social Policy (CRESP) at the University of Delaware. CPRE/CRESP's independent evaluation was funded by a 2010 i3 Scaling Up What Works grant to The Ohio State University (OSU) for the expansion of Reading Recovery in the U.S.

In examining lesson-level implementation, this chapter focuses on Reading Recovery instruction. More specifically, we discuss the characteristics of Reading Recovery instruction that implementers associate with large gains in students' reading skill, and that stood out in our observations of Reading Recovery lessons. In doing so, we offer a unified theory of what characterizes instructional strength in Reading Recovery, which we define as the extent to which a teacher instructs for maximum learning in every lesson. We also discuss school- and district-level factors that support and facilitate strong instruction.

The findings presented here emerged from four years of in-depth mixed-methods research, during which we learned that some one-to-one lessons are more effective than others, and that the strongest Reading Recovery lessons are similar in particular ways. This insight resonates with a key finding of the randomized experiment: While nearly all Reading Recovery students outperformed control students, some students made especially large gains.

RESEARCH QUESTIONS: UNDERSTANDING IMPLEMENTATION IN DEPTH

At the outset of the Reading Recovery evaluation, the implementation research was focused on two primary questions:

1. How did the i3 scale-up process unfold?
2. Did i3 schools implement Reading Recovery with fidelity?

Our findings in response to these questions are presented in *Chapter 1: Taking Reading Recovery to Scale: Processes, Challenges, and Outcomes* to the program model. Early in our investigation into these two questions, however, we observed that implementers of Reading Recovery often emphasized the importance of some aspects of program implementation that were not fully captured by our research on scale-up and implementation fidelity. These aspects—specifically instructional strength and school- and district-level administrative support—were not reflected in the Standards and Guidelines or in our fidelity measures. In Years Two through Four of the evaluation, we investigated these aspects via a third implementation research question:

3. Are there meaningful differences in schools' implementations of Reading Recovery that are not captured in our analysis of implementation fidelity and, if so, what do these variations look like in practice?

This chapter and *Chapter 6: The School-Level Implementation of Reading Recovery* present our findings in response to this third question.

RESEARCH METHODS: AN EVOLVING MIXED-METHODS DESIGN

The implementation study used both qualitative and quantitative methods. Initially, we implemented these methods in parallel. In Year One, this triangulation approach (Creswell, Plano Clark, Gutmann, & Hanson, 2007; Thurmond, 2001) yielded fidelity data in the form of survey responses, as well as rich qualitative data on how the Reading Recovery model unfolds in practice.

With the addition of the third implementation research question in Year Two, our approach shifted toward an exploratory instrument-development approach (Creswell, 2013; Creswell & Plano Clark, 2007) in which qualitative data collection preceded survey administration in each annual research cycle. This enabled us to explore emerging questions in depth in the first part of the year, and then study those questions at scale toward the end of the year via survey instruments. In Years Three and Four, we began to combine our qualitative and quantitative efforts using a more emergent approach, a fully integrated mixed design (Teddlie & Tashakkori, 2009) that was iterative in nature and that assigned equal priority to qualitative and quantitative methods (Hall & Howard, 2008).

Table 5.1. Data Sources for CPRE/CRESP's Research on Reading Recovery Instruction

		Year One	Year Two	Year Three	Year Four	Total
Interviews*	University trainers	181	18	-	13	49
	Reading Recovery teachers	46	36	24	4	110
	i3 project staff and leadership at OSU	4	3	-	3	10
	Teacher leaders	9	32	24	4	69
	1 st -grade classroom teachers	8	8	18	4	38
	Principals	17	31	21	4	73
Annual surveys**	Reading Recovery teachers	813	1511	1567	696	4587***
	Teacher leaders	190	221	215	160	786**
	Site coordinators	114	136	124	97	471**
	1 st -grade teachers	608	601	1297	-	2506**
Focus groups	Teacher leaders	34	-	-	49	83
Instructional strength survey		-	-	-	127	127
Field-based lesson observations		36	32	24	16	108
Student record reviews		50	-	-	-	50

*Interview totals include those conducted in the course of the 23 field-based case studies discussed in the following chapter.

** Response rates for the survey instruments ranged from 50% to 89%. Response rates for specific instruments are reported in Chapter 5.

*** Survey totals do not represent unique respondents, as many respondents remained in the study and completed surveys in multiple years.

¹One UTC director declined to participate in interviews for the evaluation.

The data collection and analysis processes that resulted from this mixed-methods research design are detailed below.

DATA

Our research on Reading Recovery's implementation proceeded through four annual cycles of data collection and analysis. In each cycle, researchers interviewed key implementers and Reading Recovery personnel. These individuals included Reading Recovery trainers, who hold faculty positions at the 19 regional partner universities involved with the i3 scale-up and who oversee all Reading Recovery activities in their regions. In addition, we spoke with teacher leaders who train and support teachers; trained and in-training Reading Recovery teachers; district-level Reading Recovery site coordinators; and principals and 1st-grade classroom teachers in schools that implemented Reading Recovery under i3 the scale-up. We distributed annual surveys to Reading

Recovery and 1st-grade classroom teachers, teacher leaders, and site coordinators. We conducted focus groups with teacher leaders around particular topics in Years One and Four, and reviewed Reading Recovery lesson documentation in Year One. In each year of the evaluation, we also conducted comprehensive field-based case studies, each of which included multiple interviews, lesson observations, and, where possible, observations of behind-the-glass teacher training sessions. The results of the case-study research are discussed in Chapter 6; however, the data from the lesson and training observations that were conducted as part of the case-study research informed the findings discussed here.

Qualitative sample and data collection

With the exception of the focus groups in Years One and Four, qualitative data were collected from participants via 1:1 semi-structured interviews. Interviews ranged in length from 30 to 90 minutes, and were conducted face-to-face or via telephone. All interviews and focus groups were transcribed by an outside contractor prior to analysis.

Sampling for the interviews was conducted through a combination of random sampling, convenience sampling, criterion sampling, and theory-guided sampling. The sampling approach evolved in parallel with the implementation study's research design: In Year One, we interviewed randomly selected Reading Recovery teachers and principals in order to gather general information about implementation in the schools in the scale-up. In Year Two, we randomly selected Reading Recovery teachers for interviews, and then also interviewed the principals and teacher leaders associated with the selected teachers' schools wherever possible. The decision to consolidate our interviews around particular schools in this way—essentially as mini case studies—was theory-driven: The findings from our first year of research echoed prior observations about the role of school- and district-level forces in bolstering or hindering instructional programs (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Honig, 2013). Capturing implementation from a school-level perspective was therefore essential.

In Year Three, we repeated this mini case study data-collection approach, adding 1st-grade classroom teacher interviews and sampling schools for the mini case studies based on the prior year's school-level impact estimates from the randomized trial (See Chapter 2). Schools were randomly selected from a frame of schools that had shown large treatment effects in the prior year's randomized trial, and from a second frame of schools whose prior-year treatment effects were small relative to the aggregate impact estimates. This approach supported the validity of our findings by ensuring heterogeneity of the qualitative sample, and allowed us to study implementation in schools with a range of estimated effects.

READING RECOVERY LESSON OBSERVATIONS

The findings presented here were also informed by 108 field-based observations of Reading Recovery lessons. Lesson observations were conducted by the 12 researchers who collected data for the school case studies discussed in Chapter 6. All of these researchers were experienced in conducting instructional observations, and each was trained to use

an observation protocol designed by project researchers, in collaboration with a Reading Recovery trainer, specifically for use in Reading Recovery lessons. At least four lessons were observed by a single researcher at each case study site. These lessons were conducted by a single Reading Recovery teacher as she taught the four students on her caseload. This approach allowed researchers to observe continuities and differences in a given teacher's instruction as she worked with different students.

Using the lesson observation protocol, researchers recorded the sequence of activities that made up each lesson; when and how the teacher prompted students; techniques the teacher used to instruct students on specific skills; and the proportion of the lesson time dedicated to each activity. Each observation was followed by a short interview in which the Reading Recovery teacher was asked to explain why the lesson proceeded in the particular way it had, what she noticed about the student during the lesson, and what choices she made in response to what she saw the student do. In addition, the teacher was asked to reflect on what had gone well during the lesson and what challenges she had encountered (Emerson, Fretz, & Shaw, 2011). When possible, researchers also discussed the lessons they had observed with the teacher leader who supported that particular teacher. Teacher leaders were frequently present during case study data collection and were able to observe the lessons alongside our researchers.

Five members of the research team also observed a combined total of 3 to 6 behind-the-glass lessons in each year of the evaluation. As detailed in the Year One report (May et al., 2013), behind-the-glass lessons are a core component of Reading Recovery's training model in which groups of teachers and/or teacher leaders observe a lesson conducted behind a two-way mirror. The observers discuss the lesson in real time, offering nuanced commentary on the execution of the lesson and later providing detailed feedback to the teacher. Researchers attending behind-the-glass sessions took fieldnotes focusing on the aspects of instruction that drew comments from observers; the language observers used in their discussion during and after the lesson; and the suggestions observers offered as to how the instruction could be strengthened.

SURVEY SAMPLE AND DATA COLLECTION

The sample for the annual surveys included all Reading Recovery teachers, teacher leaders, and site coordinators involved with the i3 scale-up¹. Where possible, we recruited 1st-grade classroom teachers in these schools to complete the surveys as well.

The evaluation team administered all survey instruments online using Qualtrics, a secure online platform. The instruments included both fixed-response and open-ended items, which we developed through a rigorous process that included field piloting and revision. Survey response rates ranged from 70 percent to 79 percent for Reading Recovery teachers; from 74 percent to 89 percent for teacher leaders; from 63 percent to 74 percent for site coordinators; and from 50 percent to 62 percent for 1st-grade teachers.

1. In Year Four, the survey sample included all Reading Recovery teachers and teacher leaders from schools in the impact evaluation, which is a random sample of all i3 schools.

Just as our qualitative data collection activities changed in response to our evolving research questions, so too did our survey instruments. The instruments were originally designed to collect fidelity data; however, beginning in Year Two, we expanded them to include items related to other key aspects of implementation we had identified qualitatively. Specifically, our surveys in Years Two through Four included items designed to gather data about variations in ways school and district staff support or hinder Reading Recovery's implementation.

INSTRUCTIONAL STRENGTH AS AN EMERGING THEME

The issue of Reading Recovery teachers' instructional strength arose as a theme in our implementation research in Years One and Two. Echoing much recent scholarship (Raudenbush, 2008; Rivkin, Hanusjek, & Kain, 2005), respondents consistently identified the quality of individual Reading Recovery teachers' instruction as a critical and variable determinant of student learning. In response, we launched several data-collection efforts specifically focused on instructional strength in Reading Recovery. Our goal was to understand how university trainers, teacher leaders, and Reading Recovery teachers themselves understand and recognize instructional strength, and to reconcile their perspectives with our own observations. In addition, we hoped to identify factors that can support or hinder its expression.

Probing questions about Reading Recovery instruction were added to our interviews with university trainers and teacher leaders in Year Three. Respondents were asked to provide detailed descriptions of how Reading Recovery instruction varies. For example, trainers and teacher leaders were asked to reflect on specific lessons they have observed that were particularly strong, to relate in detail what they saw, and to compare those observations with what they have noticed in weaker lessons.

Five geographically dispersed teacher leader focus groups were conducted in the fall of Year Four. These yielded specific information about instructional strength in Reading Recovery; for instance, we asked the teacher leaders to describe the manifestations of Reading Recovery teachers' instructional strength as concretely as possible. Each focus group was approximately one hour long, and included between six and twelve teacher leaders, each of whom served as teacher and coach for a caseload of Reading Recovery teachers. Most teacher leaders are experienced Reading Recovery teachers themselves, and all have been through extensive training and observe multiple one-to-one lessons each week. They are intimately familiar with the ways instruction varies from one teacher to the next.

Combined with our accumulated interview data, the focus-group data led us to begin development of a survey instrument designed to explore Reading Recovery teachers' instructional strength at scale. We piloted this instrument in Year Four of the evaluation; however, more work is needed to further refine this survey.

QUALITATIVE DATA ANALYSIS

Interview transcripts were coded by five members of the research team using Dedoose, a secure, cloud-based platform. Codes were developed through an interactive process of identification, definition, and refinement. Inter-rater reliability was established by double-coding and comparing code applications on multiple transcripts until 80 percent agreement was reached for any given code. Once reliability was established, each interview was coded by one member of the research team. We then used a structured process of constructing and comparing research memos to identify key themes and concepts. Coder memos were read by the team and dialogic engagement exercises at strategic intervals pushed towards theme refinement (Ravitch & Carl, 2016). We identified emerging themes via bracketing and constant comparative analysis; new observations were continually compared with prior findings until patterns emerged in the data. At the end of each annual analysis cycle, we refined our data-collection instruments to facilitate deeper and more nuanced investigation of key concepts in the following round of data collection (Miles, Huberman, & Saldaña, 2014).

Themes related to instructional strength and the contextual factors that support strong Reading Recovery instruction first surfaced in the interviews from Year One. In Year Two they emerged distinctly, so researchers coded across the transcripts and wrote analytic memos around dimensions and nuances within these concepts (Miles, Huberman, & Saldaña, 2014). This process led to the further refinement of our interview protocols for Year Three in order to facilitate deep exploration of instructional strength in Reading Recovery.

Also in Year Two, as our shift to a mini-case study data-collection approach focused on specific schools suggests, our unit of analysis for the qualitative inquiry shifted from respondent role groups (e.g., Reading Recovery teacher, teacher leader, principal) to individuals nested within schools within districts. This added an additional layer of analysis as researchers analyzed the full set of transcripts associated with each school (three to four transcripts per school). In Year Four, we inductively developed a rubric to assess the extent to which teachers' instructional strength, and school- and district-level administrative support for instruction, were evidenced in each of the mini-case study schools. The rubric identified concrete indicators for each level of each of these three factors. The levels ranged from 1 to 4, with 1 indicating very low levels of a given factor, and 4 indicating very high levels. Three researchers then read across all of the mini-case studies, school by school, and used the rubric to characterize the level at which each concept was manifested at each individual school. Each researcher independently analyzed every case and then the three researchers compared their findings. Where disagreements arose, they were resolved through discussion until each school had a "score" for each of the key concepts. This process produced not only a strong understanding of how Reading Recovery teachers' instructional strength is manifested and supported at the school level, but also insights into how it is expressed in a wide range of school contexts.

FINDINGS: INSTRUCTIONAL STRENGTH IN READING RECOVERY

In the sections that follow, we present the findings of our four years of research on instructional strength in Reading Recovery, and on two contextual factors that facilitate or hinder strong instruction—school- and district-level administrative support. We discuss both general findings and detailed observations about the ways these important features of Reading Recovery’s implementation are manifested. In addition, we share representative quotes from our interviews with implementers, and we locate our findings with respect to relevant research.

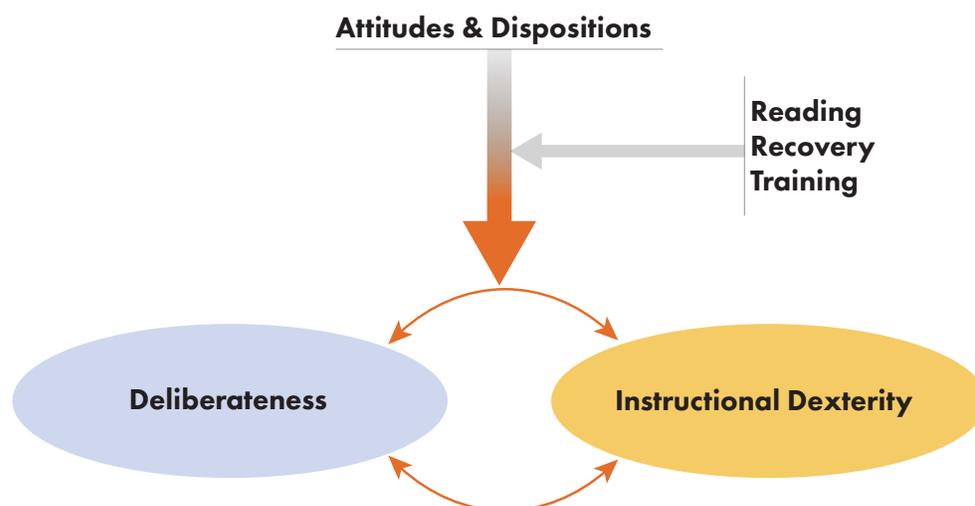
Reading Recovery teachers’ instructional strength ultimately rose above all other findings of the implementation study as *the* most important issue in the effectiveness of lessons and a likely source of the school-to-school variation in impacts we observed in the randomized experiment. Teacher quality has long been recognized as a key factor in students’ success (Aaronson, Barrow & Sander, 2007; Goldhaber & Anthony, 2007; Hanushek, Kain, O’Brien & Rivkin, 2005; Rivkin, Hanushek & Kain, 2005; Vandevoort, Amrein-Beardsley & Berliner, 2004; Berliner, 2004; Darling-Hammond & Sykes, 1999). Indeed, Raudenbush (2008) observes that decades of research on school performance increasingly identify instruction as “the proximal cause of student progress” (2008, p. 207). We observe that Reading Recovery is no exception; respondents point to the quality of Reading Recovery teachers’ instruction as the pivotal factor in the effectiveness of one-to-one lessons in facilitating students’ progress.

Our examination of instructional strength in the context of Reading Recovery builds upon a growing research base on teacher quality (Lampert & Gaziani, 2009; Fenstermacher & Richardson, 2005, 2008; Louden, Rohl et al., 2005; Richardson, 2005). More specifically, it answers recent calls from within that field of scholarship for greater focus on “instructional effectiveness, or the *what* and *how* of quality teaching” (Ingvarson & Rowe, 2007, p. 2). Ingvarson and Rowe (2007) and other scholars (Huang & Moon, 2009) call for studies that rely not on “proxy ‘measures’ of quality in terms of teachers’ qualifications, experience, and students’ academic outcomes” (Ingvarson & Rowe, 2007, p. 1), but rather offer conceptualizations of “what teachers *should know* (subject-matter knowledge) and *be able to do* (pedagogical skill)” (p. 2).

Instructional Strength in Reading Recovery: Deliberate and Dexterous

We define instructional strength in Reading Recovery, again, as *the extent to which a teacher instructs for maximum learning in every lesson*. In this definition, *learning* is operationalized as a student’s progress within and between the Reading Recovery text levels. Our analyses of teachers’ daily activity logs, lesson observations, and lesson documentation revealed that teachers involved in the i3 scale-up adhered closely to the lesson structure and procedures that are emphasized in Reading Recovery training and

Figure 5.1. Instructional Strength in Reading Recovery



codified in the *Standards and Guidelines of Reading Recovery in the United States* (6th ed.) (RRCNA, 2012). However, all instructional moves and decisions within those structures and procedures are at the discretion of individual teachers, who operate in response to individual understandings of Reading Recovery’s theoretical principles and individual interpretations of students’ needs. Thus, instruction is both interpretive and idiosyncratic, and instructional strength as we conceptualize it is not reflected in measures of fidelity to the Reading Recovery program model.

We find that those Reading Recovery teachers whom practitioners regard as strongest, and those whose lessons stood out to our researchers as particularly effective, demonstrate both *deliberateness* and *instructional dexterity*. In our conceptualization of instructional strength in Reading Recovery, *deliberateness* is understood as *an encompassing commitment to thoughtful practice*; *instructional dexterity* is defined as *the flexible application of deep skill*. These two components of instructional strength are complementary and interrelated, but manifest in different ways and at different times. For instance, *deliberateness* is manifested primarily before and after one-to-one lessons, whereas *dexterity* is evident within the lesson itself.

We also observe that underlying both *deliberateness* and *instructional dexterity* is a set of attitudes and dispositions that facilitate the development of instructional strength. By and large, respondents described these attitudes and dispositions as personal attributes—or aspects of individual teachers’ personalities—that interact with Reading Recovery training to produce *deliberateness* and *dexterity*.

Figure 5.1 illustrates this conceptualization.

Our organization of Reading Recovery teachers’ instructional strength into two distinct components and a set of underlying attributes reflects the manner in which Reading Recovery practitioners talk about teachers’ instructional skill—in terms of concrete

FINDINGS: INSTRUCTIONAL STRENGTH IN READING RECOVERY

manifestations that are both closely interrelated and notably distinct. In addition, it is informed by our own observations of both lessons and Reading Recovery training.

Our research suggests that both deliberateness and dexterity to exist on continua, with different teachers exhibiting each to different degrees. In addition, we find that instructional strength often develops over time, as a result of Reading Recovery's rigorous training process and accumulated professional experience; novice Reading Recovery teachers rarely exhibit high levels of all these characteristics. This finding is consistent with prior research on teaching quality and teachers' experience (Berliner, 1988).

DELIBERATENESS: AN ENCOMPASSING COMMITMENT TO THOUGHTFUL PRACTICE

In our conceptualization of instructional strength in Reading Recovery, we understand deliberateness—defined, again, as *an encompassing commitment to thoughtful practice*—to include those aspects of a teachers' practice that extend into the Reading Recovery lesson itself, but that are situated primarily outside that space, in the time between encounters with a student. Deliberate teachers engage in a particular set of behaviors, including:

1. purposeful analysis of students' progress that is guided by close, carefully documented observation;
2. ongoing reflection on their own instruction; and
3. active engagement with their own continual learning, both individually and through participation in a community of practice.

A description of each of these manifestations of deliberateness is provided below, along with illustrative excerpts from interviews and a brief discussion of the theory our conceptualization draws upon and extends.

1. Purposeful analysis of student progress that is guided by observation

The research team observed that strong Reading Recovery instruction is driven by an assumption on the part of the teacher that students' individual assets—be they developing skills, personal attributes, and/or family or academic resources—must be harnessed to maximize the impact of instruction. While this understanding is an expectation of Reading Recovery instruction, deliberate teachers are particularly skilled in noticing and identifying students' individual strengths and resources; they are particularly skilled, one university trainer noted, at “observing thoughtfully.” On a basic level, a teacher leader explained, this means “looking at what the child did well [in a lesson] and how to scaffold them to the very next thing that they need to learn.”

Through their lesson documentation, deliberate teachers convert their subtle observations into data, and rely on that data heavily in instructional planning. Daily running records inform the creation and continual revision of a “prediction of progress” that guides

all instruction for every student. While research suggests that teachers in general increasingly rely on data to guide their instruction (Little, 2012; Young & Kim, 2010; Boudett et al., 2005; Barrett, 2009), numerous studies attest that strong Reading Recovery instruction offers a particularly vivid and concrete example of data-informed instruction (Fried, 2013; Gibson, 2010a, 2012). The seminal writings of Marie Clay (1983; 1993; 2005) emphasize the centrality of data to the Reading Recovery approach. And, indeed, we observe that strong Reading Recovery teachers' reflection is tightly integrated with their collection and analysis of student data via lesson documentation and formative assessment tools like running records.

For deliberate Reading Recovery teachers, we observed, instruction becomes a cycle of continual collection, reflection on, and response to observations about each student. These observations pertain to their successes and missteps during lessons; their responses to particular books or activities; feedback from classroom teachers and parents; and the stories students bring from their lives outside the Reading Recovery room. It is through this constant cycle of observation during the lesson and analysis afterward, respondents said, that strong Reading Recovery teachers form instructional plans that enable them to teach for maximum impact in every lesson. As one teacher leader explained, when deliberate teachers prepare for each lesson...

... they think carefully about what new book they're going to select for that child... It's 'what does this child need? Is this book what I want to choose next?' It's 'do I see growth or not and why? Are they starting to change the behaviors when they attack an unknown word? What do I see?'

Deliberate teachers' observations of their students extend beyond the Reading Recovery context; they build relationships with parents and other teachers, and continually mine those connections for information about how their students are progressing, what motivates them, what is happening in their lives, and what barriers may need to be addressed. "[Strong teachers] are usually very skillful at bringing parents into the mix," said one teacher leader.

2. Ongoing reflection on their own instruction

As a second manifestation of deliberateness, strong Reading Recovery teachers engage in active, constant reflection on their own work as teachers. One teacher leader said: "They think 'Hey, how did it go today? Did I make some [decisions] that didn't turn out right? Did my teaching fall flat? If so, why? What did I miss?'" Deliberate teachers, another added, "are in a constant state of inquiry and refinement."

In our study of deliberateness as a component of Reading Recovery teachers' instructional strength, we draw upon the literature on reflective practice that descends from the work of Dewey (1933), Schön (1983, 1987, 1991), and Kolb (1984). In this literature, reflective practice is understood as a recursive process of giving consideration to past, present, and future actions in order to make decisions about what to do next. Reflective practitioners, according to this literature, give conscious, focused, in-depth consideration not only

to actions themselves, but to the beliefs and assumptions that underlie those actions. Reflection therefore facilitates continual improvement of practice, as one considers and reconsiders actions and beliefs in light of their results. Strong Reading Recovery teachers, a university trainer explained, “reflect on what went well and what you need to tweak and then you go again. So you have this constant reflective cycle.”

In our discussion of deliberateness, we focus specifically on those aspects of reflective practice that take place outside the lesson, as teachers contemplate past lessons in preparation for the future instruction. Our concept of deliberateness therefore resonates most closely with the literature focused on reflection on action (Schön, 1983). This theorized type of reflection stands in contrast with reflection in action, which is understood to take place in the moment (Schön, 1983). Our research reveals that, in the context of Reading Recovery, reflection on action and reflection in action are closely related but distinct kinds of activity. (Ideas about reflection in action inform our findings on instructional dexterity. We discuss this connection later in this chapter).

Despite the ubiquity of references to reflective practice in research on teaching, recent scholars note that the term is often applied too broadly to be practically useful (Mann & Walsh, 2013). Korthagen and Wubbels (1995) note that “reflection is a concept which is simply too big, too vague, and too general for everyday application” and point to a need for “coherent theories” that explicitly describe the relationship between good teaching and reflection (1995, p. 53). Mann and Walsh (2013), similarly, call for discussions of reflective practice that address its practical implications in concrete ways. The role of reflection as a concrete component of deliberate instruction offers one such theory.

3. Engagement with their own continual learning

In what respondents describe as a natural extension of this reflectiveness, deliberate teachers are active learners who continually seek opportunities to expand their knowledge of their own practice and of theories of literacy development. “You never stop learning in Reading Recovery,” a teacher leader explained. “We are always learning and as soon as someone says they know it all then we’ve got a problem.” The strongest teachers, another teacher leader said, “understand that to teach is truly to learn.”

As a result of this commitment to learning, teacher leaders report that strong teachers look for support when questions arise, drawing frequently and purposefully on their network of what Vygotsky (1978) described as “more knowing others”—trainers, teacher leaders, and more experienced teachers. One teacher leader explained:

They’re the ones who, all through the year, I probably have twenty or thirty communications with ... ‘I’ve tried this. I’ve tried this. Do you think such and such?’ We go, ‘Well, what do you think? What things could you try next?’ They don’t get tired of that.

A university trainer added: “there’s an inner tension that goes on with that particular teacher that you don’t normally see with other teachers. They’re constantly challenging

themselves.” Indeed, another university trainer explained, one mark of a strong Reading Recovery teacher-in-training is the desire to acquire as much knowledge as possible in order to provide every student with the best instruction she can:

It’s a sign of a good teacher. They feel regretful over some of the things they didn’t know at the beginning of the year, for the students they first had. [They say] ‘I wish I... you know, I could have done a better job for that one. That student would have been on a higher level if I’d known [what I know now].’

As part of this commitment to learning, we observe that deliberate teachers are continually engaged with other Reading Recovery practitioners. The importance of these connections is emphasized through Reading Recovery training. While all must participate in these training and professional development activities, deliberate Reading Recovery teachers are particularly invested in building and maintaining long-lasting practitioner networks.

In conceptualizing deliberateness in Reading Recovery we therefore draw upon the research on communities of practice—defined as “groups of people who share a concern or a passion...and learn how to do it better as they interact regularly” (Wenger-Trayner, 2011). In this literature, individual learning is understood to occur as a result of an ongoing process that is inherently social (Lave & Wenger, 1991; Lave & Chaiklin, 1993; Wenger, 1998; Rogoff, Turkkanis, & Bartlett, 2001; Wenger, 1999). Through participation in a community of practice, this research asserts, individuals who have a “shared repertoire” engage together in the process of “negotiating and learning practices that contribute to the satisfaction of a common goal” (Wenger, 1998). With their commitment to seeking and responding to feedback, and to regularly engaging with other Reading Recovery practitioners around issues of practice, deliberate teachers exemplify this process.

Findings like these lead us to extend prior research on both reflective practice and communities of practice to conceive of deliberate Reading Recovery teachers as engaged in *community-enhanced reflective practice*—a sort of reflection on action that is deepened and made more practically applicable as a result of community membership.

INSTRUCTIONAL DEXTERITY: FLEXIBLE APPLICATION OF DEEP SKILL

The second component of Reading Recovery instructional strength we identify is instructional dexterity, which we define as *the flexible application of deep skill*. The expressions of instructional dexterity we identify take place within the lesson itself, and include:

1. supportive rapport that continually pushes the student toward maximal growth;
2. in-the-moment decision-making that draws on both prior understandings and real-time observations;
3. judicious use of language; and
4. a sense of urgency that is evident in the pace of the lesson and the efficiency of instructional moves.

A description of each of the behaviors we associate with dexterity is provided below.

1. Supportive rapport that pushes students toward maximal growth

A key aspect of strong Reading Recovery instruction is the teacher's development and maintenance of a supportive rapport that continually pushes the student toward maximal growth. Through this rapport, strong Reading Recovery teachers can simultaneously convey expectation and encouragement, firmness and warmth. One teacher leader described it as an ability to keep the student engaged and motivated in the lesson while "staying on the child's cutting edge" instructionally.

Strong teachers' ability to maintain this momentum hinges in large part on their skill at connecting with their students. "They have great conversations with that child," a teacher leader said. "They build a level of trust that is just unmatched." Strong teachers are skillful at leveraging this trust to simultaneously push and support their students, moving with urgency toward instructional goals while keeping the lessons interesting and fun. "It can be as simple as the book selections," observed a teacher leader. "If you lay out [a strong teacher's] book selections across a child's program, you would see some of the personality of the child through those selections. As opposed to just 'this is my favorite book and I'll help you learn to like it.'"

Respondents told us, and we observed, that knowing when to assist a learner and when to push him toward independence can be a challenge for newer or less dexterous Reading Recovery teachers. This is a key area we saw emphasized in teacher leaders' feedback to teachers, and in behind-the-glass training sessions. Indeed, helping a student become self-sufficient by equipping him with strategies and knowledge he can later apply in any context is the central goal of the program. One university trainer explained:

If what [the student] needs to do is go back and re-read and think about the story and then get his mouth ready for that first sound ... The most amount of support would be to model that for him. The least amount of support would be "try that again." And there are amounts of support in between that. So it's being able to determine [what level of support is appropriate]. What you want to see is the least amount of support that will be useful.

Facilitating the development of this self-sufficiency demands that the teacher quickly assess each opportunity to intervene during a lesson and make strategic decisions about when intervention is necessary and when a student can learn most by problem-solving on his own. The strongest teachers "don't waste time teaching what the child already knows," reported a teacher leader. "They don't get stuck praising and prompting the things that are already under [the child's] control."

In considering this aspect of instructional dexterity, we find it useful to reflect on Reading Recovery's basis in constructivist learning theory (Piaget, 1967; Dewey, 1916; Bruner, 1966, 1983). This theory views learning as an active process in which students

are responsible for constructing meaning based on their own beliefs and experiences. Constructivist theorists—Marie Clay among them (Clay, 1991, 1998)—regard the teacher’s role in the learning process as primarily concerned with providing the student with opportunities to learn. A skillful teacher, from this perspective, is adept at anticipating, identifying, and arranging learning opportunities. Consistent with this scholarship, teacher leaders and university trainers spoke of the teacher as the *facilitator* of learning; her task is to provide the student with opportunities to learn that are carefully calibrated specifically for him, or, in the words of Clay, to “know how to open doors to learning” (Clay, 2005, p. 47).

2. In-the-moment decision-making based on knowledge and observations

A teacher’s adeptness at calibrating instruction to students’ actions is closely related to the second identified expression of Reading Recovery teacher dexterity: the ability to revise and redirect instruction in the moment. Reading Recovery teachers are trained to implement individualized teaching plans that target the specific skills critical for student progress (Clay, 2005). As teacher leaders and university trainers explain, however, a strong teacher’s plan is always tentative—continually revised in the moment based on her nuanced observations about the student. Strong Reading Recovery teachers have a plan for each lesson, but in the words of one teacher leader, “they know when to depart from that plan and do something that the child needs right then.” Another teacher leader added: “They are very keen observers [who can make] highly skilled decisions on the run.” Yet another said: “They can pick up on what the child is doing and let their body of knowledge switch gears if they have to.”

For example, a teacher’s in-the-moment-observation about a child’s sight-word recognition may lead her to question the book she pre-selected for him to read that day. As one university trainer explained: “If they’re reading a familiar book with a child, and they have some preconceived ideas of where they were going next, they are willing to adjust in the moment because of something they observed the child doing.” A university trainer further noted that:

The weaker teacher tends to follow the Reading Recovery framework tick by tick and is more concerned about checking off each lesson component than they are about being very, very responsive and flexible enough to adjust a component or a sequence when it’s obviously necessary.

An additional feature of this in-the-moment instructional flexibility is the ability to capitalize on students’ errors or partial errors during the lesson. As one teacher leader explains, a strong teacher “will acknowledge a student’s partial attempt and work with that.” For these teachers, errors are as useful as successes, as each informs the ongoing revision of the teacher’s theory in every moment of the lesson. A teacher leader reflects:

She understands that this is where this child is [in his literacy development]. If he’s not getting it this way, she knows how to move from one point to the other. Those are the ones who make the best teachers.

FINDINGS: INSTRUCTIONAL STRENGTH IN READING RECOVERY

The extent to which a teacher exhibits the flexibility to make on-the-run adjustments in response to subtle observations depends both on her level of experience and on the depth of her knowledge about her students, about literacy learning, and about the instructional strategies at her disposal. In this way, we observe that a teacher's ability to exhibit instructional dexterity depends on the degree to which she is deliberate. Said differently: *We find that deliberate teachers are not necessarily dexterous, but those who are dexterous were deliberate first.*

Our observations about instructional dexterity in Reading Recovery teachers is supported by research on teacher thinking and decision-making (Ball & Forzani, 2009), and some Reading Recovery scholarship (Rodgers, D'Agostino, Harmey, Kelly, & Brownfield, in press; Estice, 1997). Clay (2005) speaks to this aspect of strong instruction, writing: "The teacher's close supervision will allow her to detect an interfering or handicapping type of response [from the student] when it creeps in, and to swiftly arrange for a better response to occur" (p. 21).

In conceptualizing this aspect of instructional dexterity, we draw connections with research from multiple fields on expertise (Gottlieb, 2012; Salas et al., 2010; Berliner, 2001). Our findings are consistent, for example, with Berliner's (2001) assertion that "experts recognize meaningful patterns" (p. 464) and Dreyfus and Dreyfus' (2005) discussion of the "automaticity" of experts' decision-making: "The expert not only sees what needs to be achieved," they write, "Thanks to a vast repertoire of situational discriminations, he or she also sees immediately how to achieve the goal" (Dreyfus & Dreyfus, 2005, p. 787). Other scholars emphasize the role of intuition in decision-making (Styhre, 2011; Gobet & Chasse, 2009). Styhre identifies intuition as the feature of expertise that enables an individual to "apprehend the fluid and fluxing nature of the world" and to bring to bear "all the information and know-how required to accomplish a specific practice that is not provided by formal instructions or descriptions" (2011, p. 122). We find aspects of each of these ideas in respondents' descriptions of, and our own observations about, instructionally dexterous Reading Recovery teachers.

Our observations about the importance of in-the-moment decision-making for Reading Recovery instruction also draw on the scholarship on reflective practice. More specifically, we find parallels with reflection in action—that particular kind of reflection that takes place in real time, as an individual makes sense of and quickly responds to a situation (Schön, 1987; Mann & Walsh, 2013; van Manen, 1991; Richert, 1992). This type of reflection occurs, Yanow and Tsoukas write, "in the midst of action, without interrupting what one is already doing, and reshaping it at the same time" (2009, p. 1340). Griffiths and Tann (1992) call this "rapid reflection," which they characterize as "instinctive and immediate" (in Zeichner, 1994, p. 13). Tremmel describes it as "the capacity to reach into the center of confusing situations, to see itself, and to shift the base of its operations or pull up stakes altogether and follow the flow of the action" (Tremmel, 1993, p. 437).

We find particular resonance with this aspect of dexterity in a sometimes-overlapping descendant of the reflection-in-action scholarship that examines the process of

improvisation. Some compelling discussions in this area are grounded in analyses of the work and skills of musicians, surgeons, and improv comedians (Yanow & Tsoukas, 2009; Louth, 2010; Prouty, 2002; Berliner, 1994). In their analysis of improv performers, Yanow and Tsoukas (2009) observe that the actors draw on a repertoire that develops through participation in a community of practice, and that in performance they “are in ‘dialogue,’ in the moment, with the available tools and materials ... and the surprises these generate” (p. 1346). This too, we believe, is an apt description of dextrous Reading Recovery teachers as well.

3. Judicious use of language

We found that strong Reading Recovery teachers also make efficient—or, in Reading Recovery parlance, “crisp”—use of language during one-to-one lessons. The strongest teachers, a teacher leader explained, “are masterful at prompting. They don’t talk when they shouldn’t.” Another teacher leader added that the strongest teachers “don’t do a lot of teaching. They do a lot of setting up situations so that a child has to try something, and then [the student] learns from that.” A university trainer added:

[Strong teachers] have this sort of sense of just when to say the right thing, to say a prompt that’s very brief and succinct but it elicits an immediate positive and successful response in the child.

In the lessons we observed that were particularly strong, our researchers noted that the teachers spoke remarkably little by comparison with other kinds of instruction they had seen in the past. The teachers did very little explaining, and their instructions and prompts were short, clear, and focused on redirecting the students toward their established strengths. A typical redirection might be: “No. Try that again.” Or “You know a word that looks like that word.”

In coaching Reading Recovery teachers, teacher leaders and university trainers explain that they are attuned to the balance of teacher and student talk during a lesson. As that balance shifts toward more teacher talk, they say, it’s an indication that the instruction is not well matched to the student’s needs, or that he or she may not be fully engaged and therefore not receiving the full benefit of the lesson. “Who is doing the work? Who is doing the initiation? Who’s doing the problem-solving? Is it the child or the teacher?” asked a teacher leader. Another said: “Where there’s talking, there’s thinking. So if the child is talking, he’s thinking. If the teacher is talking, she’s thinking. So [strong teachers] are very economical with words. Very.”

4. A sense of urgency

The final expression of instructional dexterity we identified is a sense of urgency that is manifested in the pace of the lesson and the strategic selection of efficient instructional moves.

FINDINGS: INSTRUCTIONAL STRENGTH IN READING RECOVERY

“I always tell [teachers] ‘You are this child’s one chance. They get one chance to get caught up,’” a teacher leader reported. The urgency of this message, respondents said, is evident in strong teachers’ relentless push for rapid progress. One teacher leader reflected on a Reading Recovery teacher she supports:

She has an agenda: You’ve got to learn to read. You’ve got to do it fast and I have another kid waiting. It’s not that she’d rush them, just that she knows, you’re not getting away. You’re not allowed to not try. She demands that they give a hundred percent.

Teacher leaders and university trainers report that a teacher’s press for rapid growth is reflected in her students’ progress through the Reading Recovery text reading levels. Strong teachers “don’t keep children at the low levels for too long. They move them right along,” one teacher leader observed. This ability to quickly move students through the levels is a function, respondents say, of smart, strategic instruction. “What would be the most powerful [instructional move] in this moment?” asked a teacher leader. “We only have 30 minutes. I really have to think about what’s the most powerful.”

While weaker teachers often struggle to move students through the daily lesson’s instructional sequence in time, we observed that the strongest teachers adhere tightly to Reading Recovery’s 30-minute lesson block. “You feel the pace,” a teacher leader explained. “The lesson seems to flow smoothly and you’re amazed that it’s done in thirty minutes.” Another reflected: “It’s like watching a beautiful piece of music just happen. Their language is connected. The focus is connected to the language. It’s not disjointed. It flows.”

ATTITUDES AND DISPOSITIONS: THE BUILDING BLOCKS OF STRENGTH

Our research on instructional strength in Reading Recovery reveals that the strongest teachers who bring a set of “background” attitudes and dispositions to their practice. These attributes interact with Reading Recovery training to facilitate the development of deliberate, dexterous practice. Specifically, we find that these teachers exhibit:

1. openness to change;
2. excellent interpersonal skills;
3. a strong work ethic; and
4. belief in the ability of all students to learn.

These characteristics, respondents told us, are “must-haves” for any successful Reading Recovery teacher, and are the keys to hiring the best candidates for Reading Recovery positions. In addition, they report that these attitudes and dispositions cannot be presumed, and that they cannot necessarily be taught. Indeed, we were struck by trainers’ and teacher leaders’ tendency to characterize the personal dispositions and attitudes that

set the stage for deliberate, dexterous practice as “innate.” Though not all respondents held this view, many referenced a particularly strong teacher who “just gets it,” or spoke of a teacher’s “inherent” coachability. One teacher leader remarked:

If you go to the sports analogy, they will say, “Oh, you can’t coach that.” That athlete just does it because it’s in their soul, and because they understand and they know how to move. It’s the same thing. When you talk about a natural gift, there are teachers that understand certain things ... and they just do it. We can’t coach that.

Much research on teacher preparation and development refutes the notion of teaching ability as innate, offering evidence of the impacts of various learning experiences on teachers’ instructional skill (Naik & Ball, 2014; Darling-Hammond, Hammerness, Grossman, Rust, & Shulman, 2005; Sun & van Es, 2015; Rice, 2003). The teacher leaders and trainers we spoke with unanimously stressed the profound effects of Reading Recovery training on teachers’ instructional strength. At the same time, they often observed that people with certain attitudes or dispositions ultimately make better teachers, a position that also has basis in the literature. Building on extensive prior scholarship on teachers’ beliefs and thinking (Richardson, Anders, Tidwell, & Lloyd, 1991; Clark & Peterson, 1986), recent studies have highlighted the importance of “profession-specific teacher attributes,” including work-related motivation and self-regulation as “key aspects that determine teachers’ success” (Kunter et al., 2013, p. 807). Other recent research has demonstrated the effects of non-cognitive attributes such as a growth mindset (Dweck, 2006; Dweck & Molden, 2005) or “grit” (Duckworth, Peterson, Matthews, & Kelly, 2007) on individuals’ performance in a variety of realms. Even Dewey (1933) identified open-mindedness, whole-heartedness, and responsibility as three precursors to the reflectiveness good practitioners embrace.

Our findings suggest that the following four attitudes and dispositions are key ingredients of instructional strength.

1. Openness to change

Nearly unanimously, teacher leaders and university trainers identified openness to change as a critical attribute for strong Reading Recovery teachers. The ideal teacher, one teacher leader explained, is “that kind of person that’s open to inquiry, open to input, open to growth always.” This fundamental openness, said one university trainer, “sets the stage” for the reflectiveness and learning mindset that characterize deliberate Reading Recovery teachers.

Teacher leaders indicated that a teacher’s attitude toward new ideas is evident, as nearly all Reading Recovery teachers are challenged through the training experience to re-consider much of what they thought they knew about literacy instruction. For some teachers—frequently those who have considerable classroom experience and ingrained ideas about teaching—this change can be unwelcome. That lack of desire to change, a teacher leader said, “is hard to overcome... I think that is something that some teachers may never overcome.”

2. Strong interpersonal skills

Strong Reading Recovery teachers also possess excellent relational and communication skills. Teacher leaders and university trainers reported that strong teachers are skilled at managing relationships with their colleagues at school—in particular, the 1st-grade teachers who work with the students most of the day. First-grade teachers can provide critical insights about students’ reading behaviors in the classrooms that shape the direction of one-to-one lessons—but only if Reading Recovery teachers maintain open lines of communication. “I’ve had some really good teachers in the past but they didn’t always communicate with the staff around them,” a teacher leader recalled. “Just being an expert in individual teaching isn’t going to be a cure-all, no matter how great you are.

In addition, respondents report that these teachers are skilled at interacting not only with children but with their families as well. A teacher leader reflected on the strongest teachers she has known:

I think that they honor homes more, so they’re not focused on what the family isn’t doing. It’s more ‘every home is valuable and we’re going to honor where they come from.’ And so we don’t hear so much about the problems at home. It’s more... trying to work with the family and [tap into whatever resources] they have at home.

3. Strong work ethic

Respondents also spoke consistently of strong teachers’ willingness to “do whatever it takes” to come to each lesson fully prepared—having documented and carefully analyzed the prior lesson, reflected on the success of the previous lesson’s instruction and on the child’s trajectory overall, and prepared a plan for the new lesson. For most Reading Recovery teachers, who have back-to-back lessons and other responsibilities during the non-Reading Recovery half of the day, this requires time outside of school hours. Deliberate practice is therefore facilitated in part by a commitment to doing the work of a strong teacher, regardless of the time demands.

“It means taking work home. It means spending a lot of extra time on this,” a teacher leader said. Another observed: “If you have a strong work ethic and you believe all kids can learn, that’s the type of person we’re looking to train.”

One crucial manifestation of this work ethic is a teacher’s commitment to delivering a one-to-one lesson to every child she serves, every day. As the evaluation has noted in the past, Reading Recovery teachers report that frequent disruptions, schedule changes, and student absences can make it challenging to meet the program’s five-lessons-a-week goal (May et al., 2013, 2014). Over the course of the evaluation, we regularly heard stories of Reading Recovery teachers who come in early, stay late, rearrange their schedules, or double-up on lessons in order to make up lost time with a student. It is these teachers, respondents report, who most effectively facilitate their students’ progress.

4. Belief that all students can learn

Finally, we find that strong Reading Recovery teachers bring to their practice a deep belief that every child can learn to read and write well. A teacher leader observed: “The teachers that are good are the ones that have hope for all kids... like a hope and an urgency for them.”

A university trainer explained that strong teachers understand that every student learns differently, and that some may face greater challenges in becoming proficient readers than others. “A great teacher recognizes that every child is a new puzzle to solve,” she said. “A lesson with one child is going to look different than a lesson with another child because of that.”

Our finding that teachers’ belief in the ability of all students to learn translates to more effective instruction is consistent with decades of research that links teachers’ expectations for student learning with student resilience and academic achievement (McKown & Weinstein, 2002, 2008; Good, 1987; Benard, 1995; Robinson & Lubienski, 2011).

School- and District-Level Support for Strong Instruction

Even a very strong Reading Recovery teacher cannot be maximally effective without the benefit of strong leadership. Our research suggests that some aspects of school and district leaders’ roles directly impact the effectiveness of lessons by creating the conditions under which Reading Recovery teachers’ instructional strength can be best developed and expressed. These findings are consistent with a large body of research linking school leadership to the impact of instructional activity in schools (Goldring, Porter, Murphy, Elliot, & Cravens, 2007; Waters, Marzano, & McNulty, 2003; Leithwood, Patten & Jantzi, 2010). Leithwood, Seashore Louis, Anderson and Wahlstrom (2004) note that “the total (direct and indirect) effects of leadership on student learning account for about a quarter of total school effects” (p. 5)—a large effect when compared to those of other school-level variables.

Indeed, respondents consistently shared that the extent to which a principal prioritizes Reading Recovery has a direct impact on the strength of instruction. More specifically, implementers reported that principals who prioritize the program are more apt to select strong teachers for Reading Recovery training; to collaborate with teacher leaders to ensure that teachers receive the support they need to develop their skills; and to protect Reading Recovery teachers’ lesson schedules. Principals who exhibit this kind of support must do so despite competing priorities: For example, placing a strong teacher—one who exhibits the attitudes and dispositions we discuss above and brings a strong basis in literacy—in the Reading Recovery role often means removing a high performer from the regular classroom for part or all of the day (May et al., 2013). It’s a tradeoff many principals are unwilling to make.

FINDINGS: INSTRUCTIONAL STRENGTH IN READING RECOVERY

Despite the critical role principals play in ensuring strong Reading Recovery instruction, we find that principals' prioritization of the program is highly variable. Thirty-two percent of Reading Recovery teachers surveyed in Years One and Two of this study ($n=2,324$) disagreed with the statement: "When there are competing demands on my time, my principal makes it clear that Reading Recovery comes first." Nearly 25 percent reported that administrators at their schools assigned them duties that created direct scheduling conflicts with their one-to-one Reading Recovery lessons.

Similarly, our research reveals that district-level advocates for Reading Recovery play an important part in maximizing the strength of instruction. This advocacy role is most often filled by a district administrator—typically, an assistant superintendent or an instructional supervisor—who, among other roles, serves as the Reading Recovery site coordinator. From the vantage point of Reading Recovery teachers and other school-level implementers, the site coordinator's is a less visible role than the principal's. However, district-level administrators can play an important role in protecting the program from potential threats (Rodgers, 2016). Specifically, respondents told us, district-level advocates can help structure district hiring processes to ensure that Reading Recovery staff are carefully vetted—ideally with a teacher leader's input—and that the district seeks to hire principals who will prioritize the program. In addition, they can work with teacher leaders and principals on an ongoing basis to troubleshoot issues that hinder strong instruction by disrupting the consistency of Reading Recovery lessons.

Our survey findings suggest great variation in the extent to which site coordinators engage in school-level issues: Nineteen percent of the site coordinators who responded to our surveys in Years Two and Three ($n=236$) indicated that they had a direct role in hiring Reading Recovery personnel, pointing to an very direct role for some district administrators in ensuring strong instruction. Thirty-five percent of the 236 respondents indicated that they communicate with school administrators at least monthly about issues related to Reading Recovery. Twenty-seven percent indicated that they do so either never, or just once or twice a year. This finding is consistent with Honig's (2013) finding that district administrators are pulled in many directions, and districts often lack the capacity to provide instructional programs with the robust support school-level implementers feel they need.

Furthermore, respondents report that school and district leaders who are knowledgeable about literacy instruction generally, and Reading Recovery in particular, are most effective at hiring effective Reading Recovery teachers; at protecting lesson time; at drawing on the resources of teacher leaders; and at ensuring communication in support of Reading Recovery students. Respondents reported that the ideal Reading Recovery implementation would be overseen by leaders at both school and district levels who have personal experience with the program. "[When] the principal was Reading Recovery trained and their site coordinator was Reading Recovery trained, that's the best combination," said one teacher leader. Again echoing with prior research (Nelson & Sassi, 2005; McGhee & Lew, 2007; Quint et al., 2007), this finding suggests that school and district leaders' understanding of instructional content—a construct Nelson and Stein (2002) call "leadership content knowledge"—is a meaningful and vital component of effective Reading Recovery instruction.

In response to the emergence of this theme in Year Two, we asked site coordinators completing the Year Three survey ($n=123$) whether they had previously held a different role with Reading Recovery—teacher leader, teacher, or principal of a school that had the program. We found that 41 percent of respondents had some prior experience with Reading Recovery, and that 22 percent were actually Reading Recovery trained.

DISCUSSION

Our work on instructional strength in Reading Recovery offers insights for scholars interested in the characteristics and practices of excellent teachers, both in Reading Recovery and more generally. First, the two constructs at the heart of our discussion—deliberateness and instructional dexterity—offer a useful lens for differentiating the concrete demonstrations of instructional practice that take place *during* instruction from those that occur outside that context, while simultaneously acknowledging their interconnectedness and mutual reinforcement. Our focus on the behaviors that characterize strong Reading Recovery teachers makes our findings concrete and applicable; however, we avoid the pitfall of divorcing teachers’ actions from the perspectives and knowledge that underlie them. In contrast, a highly behavioral perspective like that advanced in Lemov’s (2010) work on effective teaching practices focuses on dexterity absent deliberateness.

Secondly, this work contributes to research on teaching by offering new ways to think about and apply the vast reflective practice literature, which is referenced widely in the scholarship on teachers and teaching. We connect reflection, as a component of our deliberateness construct, directly to teachers’ instructional strength; and second, we isolate the specific aspects of reflective practice we see in deliberate Reading Recovery teachers—those encompassed in theories of reflection *on* action. By parsing the reflective practice literature into these two representations and highlighting the ways these are manifested in both deliberateness and dexterity, we offer a grounded and practical conceptualization that warrants further study.

Our conceptualization further contributes to the reflective practice literature in its emphasis on the role of communities of practice. Mann and Walsh observe that reflective practice “is often presented as an individual process that does not foreground collaboration or participation in a community of practice” (p. 296). Reading Recovery, we find, offers a concrete model for community-enhanced reflective practice—one that is rich and informative from any number of perspectives: For instance, drawing on recent research by Ball, Ben-Peretz, & Cohen (2014), Reading Recovery offers an illustration of the role of “records of practice” in the cultivation of “collective professional knowledge” (2006); indeed, a more perfect instantiation of this theory is difficult to imagine.

The findings we present here also represent a contribution to the Reading Recovery literature, and potentially to its practice. Many articles on Reading Recovery have addressed the importance of the program’s own supportive network, and have discussed how it can embed itself in the school context (Jasmine, 2005; Dorn & Schubert, 2008; Jackson, 2008; Rodgers & Fried, 2009; Smith-Burke, 2010). However, the role of school

DISCUSSION

and district players is addressed only infrequently (Kent, 1998; Jones, 2004), and no study has looked deeply at the complex systemic factors, spanning both Reading Recovery and the schools it inhabits, that delineate those implementations that serve Reading Recovery students best. Furthermore, while volumes of literature examine best practices, our study is the first to advance a unified theory about what, precisely, characterizes the strongest Reading Recovery instruction. As such, our findings and any subsequent research in this vein may be useful for the training of new Reading Recovery teachers and the support and professional development offered to experienced teachers.

Finally, the findings presented in this report highlight the importance of administrators to the success of instruction. We argue, after Fenstermacher and Richardson (2005), that “‘quality teaching’ calls for not only certain teaching practices but also a set of contextual characteristics supportive of student learning” (p. 191). In the case of Reading Recovery, we find, principal prioritization and district-level protection of the intervention are vital and necessary supports. While the significance of leadership to instructional outcomes is well-trodden terrain, much of the existing literature focuses broadly on the myriad ways leaders impact schooling. We train the lens more tightly, working to isolate the *specific* behaviors principals and district administrators engage in that really matter to Reading Recovery instruction. For researchers and implementers interested in understanding how context can support effective instructional programs, this specificity is useful. Our findings also contribute to the literature that connects school leaders’ content knowledge with their leadership of instructional programs (Stein & Nelson, 2003; Prestine & Nelson, 2005). Research on this topic is relatively sparse; however, our finding that school and district administrators with knowledge of Reading Recovery are better equipped to support effective instruction supports the conclusions of past research, and points to a need for more study in this area.

Despite the prevalence of these understandings in the education policy and leadership literature, our findings suggest that the principal’s direct impact on Reading Recovery lesson effectiveness is under-appreciated. Teachers and teacher leaders clearly regard principal support as essential. However, while the Reading Recovery Council of North America—the national membership group for Reading Recovery practitioners—publishes a guidebook for principals (RRCNA, 2002), the principal’s role is largely unaddressed in the *Standards and Guidelines*, as well as most scholarship on Reading Recovery. Our findings also indicate that principals’ prioritization of Reading Recovery lessons varies widely and cannot be presumed. To help schools’ receive maximum impact from their investments in Reading Recovery, the program’s national leadership may need to consider new ways of ensuring principal buy-in and prioritization of the program. These themes are explored in more depth in the following chapter.

School-Level Implementation of Reading Recovery: A Cross-Case Analysis

This chapter presents the findings of the case-study research on the school-level implementation of Reading Recovery. This research is part of an independent evaluation of the i3-funded scale-up of Reading Recovery conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania and the Center for Research on Education and Social Policy (CRESP) at the University of Delaware. CPRE/CRESP's independent evaluation was funded by a 2010 i3 Scaling up What Works grant to The Ohio State University (OSU) for the expansion of Reading Recovery in the U.S.

The case-study research presented here was designed to closely examine how educational systems across the scale-up enacted and supported program implementation in context. The goal of this component was to explore the processes and successes of implementation at the school level, particularly as a way to complement our work to understand program processes and successes at the lesson level (discussed in *Chapter 5: Lesson-Level Implementation*). In addition, empirical work on how Reading Recovery operates 'on the ground' at the school and system levels is relatively limited, particularly compared to the volume of quantitative research available on the impacts of Reading Recovery lessons on student literacy skills (Munn & Ellis, 2005; Shanahan & Barr, 1995; Smith-Burke, 2010).¹ From our perspective, this was an important gap in contextualized knowledge about Reading Recovery that case studies could help address.

Throughout the course of our evaluation, we have met with differing opinions about the relationship between Reading Recovery and the contexts in which it is enacted. In our interviews with a range of Reading Recovery personnel we have heard various theories about whether and how Reading Recovery is affected by—and in turn affects—school settings; whether system-wide buy-in and support for the program is necessary for it to succeed in helping students; and whether system-level change is, or should be, a goal of program implementation.

1. For what does exist, see Glynn, Bethune, Crooks, Ballard, & Smith, 1992; Munn & Ellis, 2005; Rodgers & Fried, 2009

DISCUSSION

For several reasons, we take the stance that Reading Recovery cannot be thoroughly understood without attention to school-level enactment. First, we draw on a lengthy and significant stream of research on program and policy implementation. This literature argues that enacting a policy or program is rarely a simple, unidirectional process of inputs, throughputs, and outputs (impacts); instead, implementation is a recursive process of action, interaction, and interpretation in context (Berman & McLaughlin, 1977; Datnow, Hubbard, & Mehan 1998; Elmore, 1978; Fullan, 1993; Honig, 2006; Lin, 2002; Pressman & Wildavsky, 1984; Sabatier & Mazmanian, 1979; Spillane, 2000; Spillane, Reiser, & Reimer, 2002; Weatherly & Lipsky, 1977; Yanow, 1996). We echo Datnow et al. (1998) in asserting that teachers and administrators in school sites “do not simply respond to mandates; they are active agents, both responding to and enacting policy (p. 4)” through face-to-face interactions in real social settings.

Second, the foundational literature of Reading Recovery, particularly the work of Marie Clay (1987, 1994, 2005, 2009) states that successful implementation of Reading Recovery requires taking a system-wide view of the program, paying attention to how the program fits—or does not—in each implementation setting. Clay referred to the necessity of “problem-solving” Reading Recovery into each unique context, and argued that “in an effective intervention, the interdependence of variables demands a systemic implementation plan, for an innovation cannot move into an education system merely on the merits of what it can do for children” (Clay, 2009, p. 228). Other scholars of Reading Recovery (Askew, Fountas, Lyons, Pinnell, & Schmitt, 1998; DeFord, Lyons, & Pinnell, 1991; Lyons, Pinnell, & DeFord, 1993; Pinnell, 1989; Pinnell, Fried, & Estice, 1990; Rodgers & Fried, 2009; Schmitt, Askew, Fountas, Lyons, & Pinnell, 2005; Smith-Burke, 2010) echo Clay, and explain that implementing Reading Recovery with the highest quality and sustainability requires collective, coordinated effort, ownership, and commitment at all levels of the host system (e.g., school and district).

A third reason we chose to focus on school-level enactment of Reading Recovery is the content of materials produced by both the Reading Recovery Council of North America (RRCNA), a membership organization for Reading Recovery practitioners, and the North American Trainers Group (NATG), which consists of working groups of university-based Reading Recovery trainers. RRCNA “Implementation Briefs” describe Reading Recovery as a systemic intervention, and it is these materials that prospective adopters of the program (e.g., teachers, principals, district administrators) would likely access when seeking information about Reading Recovery. These briefs make clear that “Reading Recovery implementation, when done well, represents a serious effort to reform and improve the educational system (RRCNA, 2001, p. 1).” They also explain that to do Reading Recovery implementation well, a school must adhere to the program *Standards and Guidelines*, serve all students who are eligible for Reading Recovery, have administrator support, and encourage everyone in the building to feel ownership over the success of the program (RRCNA, 2004). The *Standards and Guidelines for Reading Recovery*, written by the NATG, also describe the program as a partnership—“a concentrated, continuous, united effort” between teachers, administrators, parents, and others (RRCNA, 2012, p. 1). Thus, it is clear that prospective implementers are meant to

understand Reading Recovery as a system-wide intervention that requires system-wide cooperation.

For all of these reasons, we chose to explore Reading Recovery in depth, and in contextualized ways, in order to better understand how the program is practiced in schools and districts, and what factors shape and mediate its enactment within and across specific school contexts. This report focuses on the 23 case studies completed over the course of the four-year evaluation. With its focus on Reading Recovery's implementation at the school level, it is a parallel analysis to the preceding chapter on lesson-level implementation.

RESEARCH METHODS

Case studies are an ideal research method for studying both organizational processes of program implementation, and the program-related attitudes and experiences of individuals within those organizations (Yin, 2012). This is because case studies promote the development of a contextualized understanding in a research or evaluation context by gathering data from many sources across multiple roles (Greene, 1998).

Between spring 2012 and winter 2014, members of the research team completed 23 case studies on individual schools implementing Reading Recovery. These study sites were randomly selected from within geographically balanced, purposive groups (i.e., urban schools, rural schools, low-performing schools, schools with large English Language Learner populations). If a school declined to participate in a case study, another school with a similar profile was selected. At least one school from each University Training Center (UTC) region was studied. Table 1 summarizes descriptive data about the final sample of case study schools.

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. A total of 12 researchers collected case-study data. Twenty of the case study schools were studied in the course of a two- to three-day site visit. In an effort to explore changes in implementation over time, three schools were studied at multiple times over two years, through a combination of site visits and phone interviews. Individual data collection and analysis processes were supported by ongoing dialogue and exchange between research team members, which was structured into the research design at strategic moments.

Data Collection

The first point of contact for each site was the 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, Reading Recovery teacher(s), 1st-grade classroom teachers, and the site coordinator. Only Reading Recovery teachers whose training had been funded by the i3 grant were selected for data collection; in cases

RESEARCH METHODS

where a school had more than one i3-funded teacher, one was randomly chosen for both interviews and observations, though attempts were made to observe all Reading Recovery teachers at the school.

A majority of the data for each case were collected during a two- or three-day site visit to the school conducted by the assigned researcher. Site visits included observation of Reading Recovery lessons accompanied by the recording of detailed field notes; shadowing the Reading Recovery teacher; observing instruction in 1st-grade classrooms; observing teacher leader visits or teacher training sessions taking place during the scheduled visit; and a series of one-to-one, semi-structured interviews

OBSERVATION AND FIELDNOTES

Researchers observed at least one lesson conducted by the Reading Recovery teacher with each of her current students (four lessons total per Reading Recovery teacher). Researchers also conducted observations of and recorded detailed fieldnotes on literacy instruction in 1st-grade general education classrooms. Two general education classrooms and teachers were chosen for data collection at each school, based on the presence of at least one current Reading Recovery 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 Reading Recovery students participated in the instructional settings, and how the teacher interacted with those students compared to how they interacted with other, non-Reading Recovery students.

SHADOWING

Following the observation of lessons, the researcher shadowed the Reading Recovery teacher as she went about her other duties for the rest of the day. The goal of shadowing was to understand the teacher's daily routine and interactions with students and adults at the school, both related and unrelated to Reading Recovery. Researchers recorded fieldnotes that carefully documented the teacher's process of preparing for lessons, the content and tone of conversations she had with teachers and administrators, and with whom in the building she interacted. If the Reading Recovery teacher was in her training year, when possible the researcher also shadowed the Reading Recovery teacher as she attended her weekly professional development session conducted by her teacher leader. If the Reading Recovery teacher was visited by her teacher leader at school for coaching sessions, these activities were also observed.

INTERVIEWS

Researchers conducted one-on-one, semi-structured interviews with individuals at both the school and district levels about their involvement with Reading Recovery

implementation, and about how the program functioned at the case study school. Each interview lasted approximately 45 minutes. Reading Recovery teachers were interviewed independent of shadowing and observation. The goal of this interview was to probe the teacher's understanding of her role in the school, her perception of how Reading Recovery was impacting students and the school overall, and to map her support network both inside and outside the school. More specifically, the interview included questions about:

- » the teacher's background as a teacher, and her choice to be trained in Reading Recovery;
- » the training she received as part of the program;
- » the response she, and Reading Recovery in general, had received from the school community;
- » perceptions of institutional change at the school related to the implementation of Reading Recovery;
- » challenges to implementing the program with fidelity, as defined by the *Standards and Guidelines*;
- » how well Reading Recovery aligned with the school's overall approach to literacy instruction;
- » what school- and district-level supports and oversights were in place for Reading Recovery, and those needed; and
- » prospects for the future of the program in the school.

Researchers also conducted interviews with the principals of each assigned school. These interviews focused on the history of Reading Recovery adoption and implementation in the school; the influence of i3 funds on adoption and implementation; the performance of students who participated in Reading Recovery; the attitudes toward Reading Recovery throughout the building; the structures in place to support the program; the alignment of Reading Recovery with the literacy goals of the school; the perceived "value" of Reading Recovery as an investment (e.g., whether or not the significant financial commitment is worth the level of impact on student achievement); and future plans regarding the program.

Researchers interviewed the teacher leader responsible for each site both to learn about his or her role in the implementation of Reading Recovery in the school and district, and to triangulate comments from the Reading Recovery teacher and principal. From their unique position, being both involved in the school and removed from it at different moments, teacher leaders are well positioned to verify or question the structures and interactions the Reading Recovery teacher and principal describe. In addition to questions about the history of Reading Recovery adoption and student achievement at the case study school, the teacher leader was asked questions about how well the principal of the school understood the program, what level of collaboration existed between the Reading Recovery teacher and other teachers, and to what extent Reading Recovery had become part of the school's approach to teaching and learning. When possible, the site coordinator

who oversaw 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 Reading Recovery and plans for the program going forward.

Interviews with general education teachers focused first on their overall familiarity with Reading Recovery, such as how students were chosen to participate in the program, and what happens during a Reading Recovery lesson. These interviews also included questions about structures and relationships surrounding Reading Recovery in the building, including how the teacher thought about her individual role in implementation, her interactions with the Reading Recovery teacher, what Reading Recovery supports were available, how Reading Recovery was generally regarded, and the level of understanding and support of Reading Recovery provided by the principal. Finally, 1st-grade teachers were asked to comment on what impacts on pedagogy and student achievement they saw resulting from Reading Recovery, particularly whether their own literacy instructional practices had shifted with their increasing familiarity with Reading Recovery, and how the Reading Recovery students in their class were progressing in comparison to other students.

Data Analysis

The analysis of case study data occurred in two stages. In the first stage of analysis, each case was treated as a stand-alone entity. The goal of within-case analysis at this stage was to develop a contextualized understanding of Reading Recovery implementation in each unique setting, and to allow the important aspects of each individual case to emerge before the cases were compared (Eisenhardt, 1989).

INDIVIDUAL CASES

After interviews were transcribed, each of the 12 researchers involved in data collection analyzed the qualitative data collected from his or her assigned site. The overarching questions used to guide this analysis were:

1. What are the different ways people think and talk about Reading Recovery?
2. What contextual characteristics influence the way Reading Recovery is implemented in this school, and do they affect the school's ability to implement Reading Recovery with fidelity?
3. What constitutes the Reading Recovery support network at the school?

Each researcher open coded transcripts from the interviews they conducted, developed analytical memos, and used this analysis to craft a written case study. Researchers were given a list of potential issues to write about based on the data they were able to gather at their case study site. This list was not meant to be restrictive, but to aid researcher analysis;

Table 6.1. Case Study School Descriptive Data*

	Locale	Student Population	% ELL**	% FRPL	Years of Reading Recovery***
School A	Rural: Fringe	629	0.46	73.29	10
School B	Suburb: Large	555	5.57	29.37	3
School C	City: Mid-size	521	1.67	27.83	2
School D	Suburb: Large	568	19.96	50.00	13
School E	Suburb: Large	745	Data not available from IDEC	42.15	Data not available from IDEC
School F	Suburb: Large	707	3.18	7.78	8
School G	Suburb: Large	410	0.17	26.34	6
School H	Town: Distant	604	6.42	81.62	12
School I	Rural: Distant	344	0.60	71.80	7
School J	Rural: Distant	84	Data not available from IDEC	73.81	Data not available from IDEC
School K	City: Small	394	8.16	35.79	1
School L	City: Small	277	0.69	55.60	12
School M	City: Large	450	6.30	47.33	Data not available from IDEC
School N	Suburb: Large	518	19.96	83.20	10
School O	City: Large	700	2.80	40.43	11
School P	City: Small	630	0.13	42.06	Data not available from IDEC
School Q	City: Mid-size	775	19.96	58.97	13
School R	Rural: Distant	334	1.92	67.07	2
School S	Suburb: Large	420	1.64	9.05	3
School T	Suburb: Large	645	2.92	38.76	3
School U	Suburb: Mid-size	567	1.26	5.82	12
School V	Suburb: Large	590	19.96	55.42	13
School W	Rural: Fringe	520	4.28	89.23	11

*Note. The data presented here represent the year the school was visited for a case study

**Note. %ELL data are from the district level

***Note. Refers to number of years the school had been implementing Reading Recovery at the time it was visited for the case study

RESEARCH METHODS

each researcher was also asked to include any additional information they gathered that was not covered by the topics on this list. These issues for potential focus included:

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

This list was developed from the case study interview protocols, and was intended to capture a broad picture of the history of implementation in the school, as well as a more nuanced vision of how Reading Recovery fit in each unique context. Based on their individual analysis, each researcher wrote a detailed case study memo of approximately 7,500 words.

CROSS-CASE SYNTHESIS

In the second stage of analysis, the individual cases were compared and contrasted. Using a variable-oriented strategy (Huberman & Miles, 1998), the lead author of this chapter read and synthesized the 23 individual case study memos. The goal of analysis at this stage was to identify themes and patterns in modes of Reading Recovery implementation both within and across the cases. She compared and contrasted the individual cases, guided—but not constrained—by the themes listed above, looking for similarities and differences in how Reading Recovery was understood, experienced, and implemented. The lead author discussed the analysis with the project’s core qualitative team and the project co-PI overseeing the implementation study through regular, ongoing meetings. The findings of this cross-case comparison were shared with the research team (the members of which had all conducted case study data collection) as a form of triangulation.

FINDINGS

Overall, the basic processes of Reading Recovery implementation looked very similar across the 23 case study schools. Almost all appeared to be implementing the program with fidelity to the best of their ability and understanding, and the practices of daily instruction were largely the same. However, if we take a view of implementation as an ongoing interaction of roles and routines playing out in specific contexts, we can see that there was a great deal of variation in how Reading Recovery was operating in schools. In particular, there was variation in the types of relationships formed around Reading Recovery, how well the program was understood by those outside of it, how deep a commitment teachers and administrators expressed to supporting and sustaining the program, and how much it appeared to be part of the school's broader approach to literacy. In each case study school, these and other factors combined to create the conditions of implementation, and these conditions influenced the level of success (i.e., fidelity, support, and sustainability) the program achieved in each context.

In some of our cases, Reading Recovery operated as a largely isolated intervention; in these schools the program remained a mystery to most people in the building, as there was little communication about Reading Recovery, and little or no connection between Reading Recovery and classroom instruction. In other cases, Reading Recovery was the centerpiece of a comprehensive Response to Intervention program, was influencing both classroom instruction and school-wide relationships, and was clearly valued by both teachers and administrators. One goal of the following analysis was to look closely at the different conditions of implementation in these cases—the different ways people enacted Reading Recovery in context. A second goal was to tease apart how those conditions are related to the successful implementation of Reading Recovery at the school level.

A wide variety of issues appeared to influence how Reading Recovery operated in the contexts of these case study schools. Seemingly small differences in relationships, resources, perceptions, and actions created very different conditions for implementation. At their root, however, most factors that affected how Reading Recovery operated in a school can be loosely grouped into two categories: issues of understanding and issues of commitment.

In the case studies, issues of understanding were related to the extent to which those outside the Reading Recovery program—i.e., teachers (particularly 1st-grade teachers) and administrators—were knowledgeable about different aspects of Reading Recovery. These included Reading Recovery philosophy, processes, and instructional strategies; how Reading Recovery can benefit the whole school; what roles support the program; and what are reasonable expectations of the program. When teachers and administrators in our case study schools had a strong, shared understanding of how and why Reading Recovery operates as it does, they were better prepared—and more likely—to protect program time, advocate for it, and contribute to successful implementation.

Issues of commitment to Reading Recovery implementation were related to how well the teachers, administrators, and Reading Recovery professionals in case study

FINDINGS

schools communicated and collaborated to collectively ensure the program operated with efficiency, fidelity, and broad support. Independent of level of understanding, commitment to successful Reading Recovery implementation may suffer due to continuous budget concerns, administrator preference for alternative programs, or barriers erected by classroom teachers. When teachers and administrators shared a strong, collective commitment to high-quality implementation, they problem-solved to protect Reading Recovery time, positioned the Reading Recovery teacher as a literacy expert and leader in the building, and organized classroom instruction to support Reading Recovery.

The schools in our sample of case studies demonstrated different levels and combinations of understanding and commitment related to Reading Recovery. Some schools had both a shared understanding of the program and a clear commitment to it, some seemed to have neither, and other schools had one but not the other. These different combinations of understanding and commitment can be organized as *schemas of implementation*, or different ways of thinking about, talking about, and doing Reading Recovery in context. The four schemas we identified through in-depth analysis of our 23 case studies are presented here. These are Isolation, Obstruction, Endorsement, and Integration.



Table 6.2. Characteristics of Schemas of Implementation

	ISOLATION	OBSTRUCTION	ENDORSEMENT	INTEGRATION
	"Reading Recovery is its own island."	"Reading Recovery doesn't always meet our needs."	"We don't necessarily know what goes on in Reading Recovery, but we're glad to have it."	"Reading Recovery is part of and enhances what we do here."
# of schools in case study sample	2	4	7	10
School-wide shared understanding of Reading Recovery	No	Variable	Variable	Yes
School-wide problem-solving to ensure program fidelity	No	Variable	Yes	Yes
Principal engaged with Reading Recovery	No	Variable	Passively	Actively
Reading Recovery lesson time protected	No	No	Yes	Yes
Frequent Reading Recovery teacher/teacher collaboration and communication	No	Variable	Yes	Yes
Reading Recovery teacher viewed as leader/resource	No	No	Yes	Yes
Classroom instruction supportive of Reading Recovery	No	Variable	Variable	Yes

Schemas of Implementation

The following analysis ascribes schemas of implementation to case study schools based on the perspectives available in the school at the moment in time (or moments, in longitudinal cases) that they were studied. It is important to note that the schemas identified are not fixed, rigid categories, but are fluid groupings. Depending on who is doing the describing, an implementation may exhibit characteristics of different schemas, and schools may demonstrate different schemas over time. Table 2 summarizes some of the defining characteristics of the four schemas, and indicates some of the primary ways the schemas differed from each other.

SCHEMA OF IMPLEMENTATION: ISOLATION

LOW UNDERSTANDING/LOW COMMITMENT

“Reading Recovery is its own island.”

- Little or no shared understanding of Reading Recovery
- Little or no problem-solving to ensure program fidelity
- Principal not engaged in Reading Recovery implementation
- No emphasis on school-wide impact of Reading Recovery
- Reading Recovery instructional time not well protected
- Little Reading Recovery teacher/classroom teacher communication
- Reading Recovery teacher not positioned as literacy leader or resource in the building
- Classroom instruction generally not supportive of Reading Recovery

We identify the first schema of implementation we observed among the case study schools as Isolation. In the schools exhibiting this schema, Reading Recovery was ‘isolated’ in that it was not positioned as part of the school’s broader approach to literacy, nor was it aligned with (i.e., influencing or being influenced by) classroom literacy instruction. Our interviews with teachers and administrators revealed that no one outside of Reading Recovery in these schools had much knowledge of either the specific instructional strategies of the program, or its underlying theory or philosophy. Nor did anyone outside of Reading Recovery indicate that they felt they had a role to play in successful implementation. Reading Recovery had very clear boundaries in these cases, and was not being integrated into the broader context of the school via exchanges of information and expertise.

Only two of our case study schools demonstrated elements of isolation, and we hypothesize two reasons so few schools in our sample exhibited this schema. First, it is possible that Reading Recovery does not remain in schools very long when it is implemented in isolation, and fails to become integrated into the school’s approach to literacy. This would mean that many more schools implementing Reading Recovery at any given time are exhibiting other schemas. Second, while we made every effort to

compile a representative sample of case study schools, there was an unavoidable element of self-selection. Schools were able to decline participating in our research, and schools exhibiting isolation may have been less inclined to agree to participate.

The school-wide lack of understanding of Reading Recovery that characterizes the isolation schema can be framed as both a cause and consequence of how little communication there was between Reading Recovery teachers and classroom teachers in the case study schools. Conversations usually occurred as quick interactions during Reading Recovery student pick-up and drop-off from class, and topics were restricted to the student's progress. There were very few (sometimes no) formal opportunities for knowledge sharing between the Reading Recovery teacher and classroom teachers (e.g., in faculty meetings or scheduled professional development). Classroom teachers in these cases did not describe Reading Recovery teachers as a resource for literacy expertise, conceivably because they had so little interaction with them.

There was little effort on the part of teachers or administrators to support Reading Recovery in case study schools demonstrating the isolation schema, suggesting these people did not see roles for themselves in program implementation. This means that there was no expression of commitment to making sure the program was implemented with fidelity and sustainability, from either the classroom teachers or administrators. Reading Recovery time was not highly prioritized, and there was no collective problem-solving to protect Reading Recovery time when conflicts arose. As a result, Reading Recovery lessons were sometimes sacrificed in order to meet competing needs.

It was clear in the case study schools where Reading Recovery remained isolated that the program was not highly valued by either the administration or the classroom teachers. Several Reading Recovery teachers mentioned hearing negative comments about the program from people within their building; the comments they mentioned hearing most frequently from teachers and administrators were about the cost of the program, and how the staff position dedicated to Reading Recovery would be better allocated to someone who served more students. In interviews, principals and teachers in these schools were not shy about speaking up regarding their concerns. One principal commented, "I mean, really? We're going to save eight starfish and leave the rest of them on the shore?" and another performed calculations of the Reading Recovery teacher's salary-per-student during her interview. A classroom teacher talked at length about how Reading Recovery is "such an expensive program." She remarked:

When you think about how much... if the teacher only sees eight kids a year, and divide their salary by those eight kids, that is so expensive. I'm not saying Reading Recovery isn't good, but isn't there something else that's just as effective that wouldn't be so expensive to implement?

This sentiment was repeated often in cases across the schemas. When administrators or classroom teachers expressed concerns or reservations about Reading Recovery, they were rarely based in feelings that the program does not work, or that it does not significantly improve students' literacy skills. Their hesitation to commit to Reading Recovery was

FINDINGS

generally presented in terms of a cost-benefit analysis; they feared that the per-pupil cost was disproportionate to the limited reach of the program (e.g., eight students per Reading Recovery teacher). In schools demonstrating isolation, this view was reinforced by the fact that administrators and teachers generally thought of Reading Recovery as impacting only the few students served, not as benefiting the whole school.

So while program cost was an issue discussed frequently in schools across all of the schemas of implementation, this issue presented a particularly tricky stumbling block in those demonstrating isolation. In these cases, administrators and classroom teachers viewed Reading Recovery as something separate from other literacy instruction—a program that operated without the need for, or benefit of, their input.

In the two schools demonstrating isolation, no groundwork had been laid before implementation began; that is, neither the administrators nor classroom teachers were given any guidance as to how the program could benefit the whole school, or what roles they could play to contribute to the program's success and facilitate the school-wide spread of expertise. Without a more nuanced understanding of Reading Recovery, the administrators and teachers in isolation schools relied on their perceptions of the high cost and limited reach of the program when making judgments about the program's value. Reading Recovery teachers and teacher leaders in these schools remarked that, if administrators better understood the school-wide impact Reading Recovery could have, they may be more comfortable with the cost-per-student, and more apt to foster collaboration between classroom teachers and Reading Recovery teachers. They also commented that if classroom teachers better understood how Reading Recovery teachers could serve as a resource for them, they would also be more comfortable with the cost of the program. However, concerns about the cost of Reading Recovery seemed hard to overcome once the isolation perspective was established, leaving administrators and teachers hesitant to embrace and commit to the program. Teacher leaders stated that, without commitment to making the program successful, Reading Recovery would never be integrated into the school, and understanding would never improve. This cycle left the program, along with the Reading Recovery teacher, isolated in the Reading Recovery classroom, facing persistent challenges to fidelity, support, and sustainability.

In School K, where the program was particularly isolated, the lack of commitment to Reading Recovery manifested as classroom teachers creating obstacles for the Reading Recovery teacher. The Reading Recovery teacher in this school described the principal and classroom teachers as lacking any understanding of Reading Recovery, and having no clear commitment to supporting the program or ensuring it was implemented with fidelity. She knew from conversations with others in her building that their primary objection to the program was cost. She repeatedly commented that these conditions created a situation in which “it has always been a struggle to actually do [her] job.” She explained that this “struggle” included facing what she perceived as resistance from classroom teachers, particularly about excusing students from class to attend their Reading Recovery lessons:

They're just not making time for me. I have this three-and-a-half hours that I need to teach, and this is the intervention program for first grade, so there should be something in place that says "[The Reading Recovery teacher] needs three and a half hours." But it's always been a struggle. Always.

In addition to negative reactions to program cost, this Reading Recovery teacher felt there was another reason classroom teachers were not making time for her: the lack of understanding and commitment on the part of the principal. Though the principal voluntarily brought Reading Recovery into School K, the Reading Recovery teacher perceived he had done so rather grudgingly, and had not taken any steps to situate the program. For instance, the principal had not made any statements of support or endorsement of the program, had not created any structures by which the Reading Recovery teacher could share her literacy expertise with the classroom teachers, and had not indicated to the staff in any way that Reading Recovery was a welcome and important addition to the school's approach to literacy instruction.

For the principal's part, he did not prioritize Reading Recovery, nor did he see himself as having a role in implementation. He felt it was important to "respect [the classroom] teachers' professional judgment" in how they chose to interact with the Reading Recovery program. He maintained this opinion even following a decision by the teachers to rearrange the daily schedule in a way that was detrimental to Reading Recovery. The Reading Recovery teacher felt that this lack of demonstrated support by the principal was the reason the classroom teachers questioned the value of Reading Recovery, and therefore actively isolated both her and the program. Because of the pushback she received from classroom teachers, the Reading Recovery teacher felt it was particularly important for the principal to take responsibility for bringing the program into the school. It was inappropriate and ineffective, she thought, for her to try justifying her own position:

I feel like each year I'm just trying to make myself clear, and I get it. It's just one person, four kids at a time. But I keep saying "Reading Recovery impacts the whole class, because if these guys can participate at all close to their peers, it's going to help the whole class." So I just keep saying what I say... People need to understand why Reading Recovery is important, and I think that comes from the higher-ups who made the decision to implement it.

In this school, lack of understanding and commitment manifested as classroom teacher pushback and limited administrative support, and these conditions resulted in isolation of Reading Recovery and its impact.

Reading Recovery also remained entirely isolated in School J. While there was no active pushback from teachers or administrators, still neither group had any understanding of Reading Recovery, or any clear commitment to supporting successful implementation. The Reading Recovery teacher in this school explained that neither the principal nor any classroom teachers had observed her during a lesson, or come to her with questions regarding literacy instruction. When asked about her perspective on Reading Recovery, a 1st-grade teacher at School J indicated her limited understanding of the program by her

FINDINGS

description of it: “I know it is an early literacy program, primarily for K-1, and sometimes 2, if they’re still needing it.”

These cases of isolation reinforce the importance of taking a system-wide view of Reading Recovery implementation. If we were to only focus on what was happening inside the Reading Recovery classroom at these schools, we would likely see high-quality lessons and student growth. However, this would not give us any indication that the program was facing threats to fidelity and sustainability. By taking a broader view, as we have with these case studies, we can see the school- and system-level factors influencing implementation, and are better able to understand the barriers to success Reading Recovery faces in context.

SCHEMA OF IMPLEMENTATION: OBSTRUCTION

VARIABLE UNDERSTANDING/VARIABLE COMMITMENT

“Reading Recovery does not fit our needs.”

- Variable understanding of Reading Recovery
- Variable commitment to program fidelity
- Limited principal engagement with Reading Recovery
- No emphasis on school-wide impact of Reading Recovery
- Reading Recovery instructional time not generally protected
- Variable Reading Recovery teacher/classroom teacher communication
- Reading Recovery teacher not positioned as literacy leader or resource in the building
- Classroom instruction variably supportive of Reading Recovery
- Classroom instruction generally not supportive of Reading Recovery

The second schema of implementation we identified was **Obstruction**. In contrast to schools demonstrating isolation—where there was neither understanding nor commitment to Reading Recovery—understanding and commitment were variable in schools demonstrating obstruction. What schools demonstrating obstruction had in common was that their contextual conditions were, or became, unable to support successful Reading Recovery implementation. That is, one or more obstacles interfered with Reading Recovery implementation in a way that threatened some aspect of the program’s success. Schools demonstrating the obstruction schema provide evidence that seemingly small components of the implementation context can have important implications for the success of the program. Four schools in our sample of case studies exhibited aspects of this schema.

Obstructions to Program Fidelity

In some schools demonstrating obstruction, the conditions of implementation were unsupportive of Reading Recovery due to contextual constraints. For instance, there were several cases where the obstacle to successful implementation emerged as a result of how Reading Recovery was staffed in the building. Reading Recovery teachers generally devote

half of their day to Reading Recovery lessons, and play a variety of roles during the rest of their day. In various case study schools, Reading Recovery teachers doubled as math resource teachers, literacy coaches, special education teachers, English language teachers, and administrators-in-training. These roles were generally combined as a way to make Reading Recovery economically feasible for the school. However, combining these roles with the responsibilities of the program was challenging, and several Reading Recovery teachers mentioned that when their roles conflicted, it was often Reading Recovery that was de-prioritized. “Administrators aren’t always protecting that Reading Recovery time,” explained one teacher leader. “They’re saying if they’re short on teachers or they don’t have a substitute for the day, they might use a Reading Recovery teacher or two.” In these cases, the role combinations were seen as necessary to meet the needs of the school, but were also obstructions to high-fidelity implementation.

The evaluation has previously reported that fewer than half of Reading Recovery teachers also work as classroom teachers or co-teachers (see Year One report, May et al., 2013). This Reading Recovery teacher/classroom teacher role combination was present in two of our case study schools, where it proved problematic in several ways for implementing Reading Recovery with fidelity. First, teacher leaders and Reading Recovery teachers explained that the classroom teacher role offered less flexibility than some other roles frequently held by Reading Recovery teachers. Those who doubled as classroom teachers therefore reported more difficulty in making up missed lessons. Second, the Reading Recovery teacher/classroom teacher combination of roles threatened lesson fidelity and quality in these schools when the teachers found it difficult to balance their attention and obligations to both Reading Recovery and their classroom. One of the Reading Recovery teachers who was also a 1st-grade teacher explained that it was difficult to “let go” of her classroom students “because, as a teacher, this is my class, and these are my babies... accepting that I wasn’t going to teach them writing, or social studies, or science. It was very hard, even though I have this passion for Reading Recovery.”

This issue of role balance is where previously discussed School J could be thought of as overlapping the isolation and obstruction schemas. As mentioned, Reading Recovery was totally isolated in School J; neither the administrators nor teachers had any understanding of the program, and none expressed any interest in interacting with the program or capitalizing on the Reading Recovery teacher’s expertise. There was also no expressed commitment to maintaining the program past the three years required to receive i3-funded training. The Reading Recovery teacher at this school was also a 1st-grade teacher, and chose to conduct her lessons in the back of her classroom so that she could keep watch over her other students as they worked with a classroom aide. This very frequently led to a distracted teacher as well as distracted students. This teacher explained that she felt responsible for what was happening with the rest of her class while she was delivering Reading Recovery lessons. “I still have that ownership over my class that I feel like I need to be the one to oversee it all, even if I have [an aide] in there helping me,” she said. “Does that make sense?” She continued, “I would like to be the one that is wandering around looking at their journals and checking their writing. And I want to be the one who’s doing the guided reading with everybody.”

FINDINGS

In other cases of obstruction, the conditions of implementation became unsupportive of Reading Recovery not because of staffing resource constraints, but because the value of the program was contested in some way. In these case study schools, someone in the building—administrators or other teachers—perceived a lack of fit between Reading Recovery and the needs of the school. In some places this perception was borne of poor understanding of Reading Recovery. More often, however, it was an issue of competing priorities and an associated lack of commitment to protecting the program at the expense of other needs. For example, several teacher leaders in our case study schools discussed the challenge of negotiating with principals who decide to prioritize other programs, and choose not to protect Reading Recovery. They explained that this is particularly difficult because principals hold the ultimate decision-making power about how their staff will be used. As one teacher leader explained “Principals pretty much have free reign of using their people however they see fit.” In one of the schools she oversaw, this meant the principal restricted the number of students the Reading Recovery teacher could work with. This teacher leader worked with her university trainer to make sure the principal understood that the Reading Recovery teacher needed to meet the standard of working with four students per semester, but to no avail. “What it came down to,” the teacher leader explained, “was the principal said ‘Okay, well then the teacher will not be teaching Reading Recovery.’ Because they wanted her to do other things.” It seems that from this administrator’s perspective, protecting Reading Recovery—and emphasizing fidelity to the standards of implementation—did not meet the needs of the school.

Reading Recovery also faced barriers to fidelity and quality in case study schools when administrators or classroom teachers felt Reading Recovery did not fit with other school or district priorities for curriculum or instruction. For instance, in School H, the principal felt it was important for Reading Recovery to contribute to meeting other, non-literacy-related district goals. Reading Recovery had been operating in this school for over 10 years. Because it had been in place so long, classroom teachers and the principal were familiar with the Reading Recovery approach to literacy instruction, and had knowledge of the instructional strategies it employs. The barrier to successful implementation emerged with the introduction of a new teacher evaluation system in the state. The Reading Recovery teacher explained that her principal expected to see certain elements of instruction in Reading Recovery lessons when she was observed for evaluation, in order to comply with the new system. In her understanding, not including these elements would lead to a negative evaluation. The Reading Recovery teacher explained:

In each lesson we have to include parts of [other school-wide, non-literacy programs]. It’s not really what Clay has in her 30 minutes of Reading Recovery... the lesson didn’t last 30 minutes, it lasted like an hour. I was like “You know that I can’t do that for a whole hour every day.” And Clay says choose your words carefully on something that is going to help the child. And I said “I really don’t think this is choosing my words carefully. It’s just confusing.”

Obstructions to Program Sustainability

In other case study schools demonstrating the obstruction schema, the obstacle was not to fidelity of implementation, but to sustainability or longevity of Reading Recovery in the school. In such situations, classroom teachers and administrators frequently remarked that that Reading Recovery did not meet the literacy needs of their school. In the case of School T, the classroom teachers did not see Reading Recovery as the best choice for their students. The principal of School T commented that while she herself felt there were “no strategies better than Reading Recovery strategies,” the classroom teachers felt differently. “I wouldn’t say that people are dying to get their kids into Reading Recovery,” she explained. The principal attributed this lack of teacher interest and commitment to Reading Recovery, in part, to the fact that the school was not serving all students who needed the intervention. She also explained that other non-Reading Recovery literacy specialists in the building did not have clear confidence in the program. “The reading teachers do have reading degrees, so they are specialists in their field,” she commented. “Whereas our Reading Recovery people are not necessarily reading specialists. Although now they have Reading Recovery training, they just didn’t have the prior experiences that our reading staff had.” The principal continued:

Quite often when the child is exiting after 22 weeks in Reading Recovery, they are still in need of support. They’re not really ready to go back to the classroom and just keep going, so we’ve often had to transfer them into a reading support group... which doesn’t create the best viewpoint of Reading Recovery for the reading teacher. So they’re feeling like “If I had this child from the beginning, I’d be building continuity, experiences. Now I’m getting this child in February and they have been taught in a similar, but not exactly the same way.” So that’s been a sticky thing.

With a lack of “teacher buy-in to the model,” the principal felt the future of Reading Recovery at her school was uncertain.

In another case of obstruction, School E, Reading Recovery had been in place for over 20 years. Because it had been in the school for so long, the teaching staff was familiar with the program’s approach to literacy instruction, and many teachers were using Reading Recovery-based instructional strategies in their own classrooms. The relatively new principal, however, had little understanding of Reading Recovery, and made clear that she was considering discontinuing the program. She explained that she was committed to providing an intervention for struggling first graders, but was not certain Reading Recovery was the best choice, despite the value the classroom teachers clearly placed on it.

This principal gave several reasons for her hesitation. For instance, she felt the available data on long-term student gains was not conclusive. “I really want to look at that data for those kids leaving sixth grade,” she explained. “Have they been able to maintain their reading, that trajectory [after Reading Recovery]?” The principal also suggested that perhaps Reading Recovery was over-diagnosing students as in need of the intervention. “I think sometimes children come out looking like they require Reading Recovery when they

don't actually," she explained. "I think there are actually fewer children who need it." The Reading Recovery teacher in this school commented that she did not think the principal had "any knowledge of how powerful [Reading Recovery] is," and the teacher leader made similar comments. In this case, though there was understanding and commitment on the part of classroom teachers, the principal's perception of the value of Reading Recovery was an obstruction to the future sustainability of the program in School E.

SCHEMA OF IMPLEMENTATION: ENDORSEMENT

Variable Understanding/High Commitment

"We don't necessarily know what goes on in Reading Recovery, but we are glad to have it."

- Variable understanding of Reading Recovery
- Active school-wide problem-solving to ensure program fidelity
- Principal passively engaged with Reading Recovery
- Little emphasis on school-wide impact of Reading Recovery
- Reading Recovery instructional time protected
- Frequent Reading Recovery teacher/classroom teacher communication
- Reading Recovery teacher positioned as literacy leader or resource in the building
- Classroom instruction variably supportive of Reading Recovery

The third schema of implementation we identified was **Endorsement**. In cases of endorsement, there was widespread outward commitment to Reading Recovery at the school level, despite the fact that few people outside the program had any extensive knowledge of it. People in these schools did not know much about how Reading Recovery works or what takes place in lessons, but they were glad to have it, and were generally committed to making sure it operated successfully in the school context. One Reading Recovery teacher's description of her principal well sums up the schema of endorsement: "I don't know how much of what we [Reading Recovery teachers] actually do she really knows. She sees the good results of what we do, but not necessarily what goes into it."

Seven of our 23 cases study schools exhibited aspects of endorsement. In these schools, understanding of Reading Recovery was absent, but commitment to the program was present. Commitment to making Reading Recovery successful seemed to emerge a few different ways in these schools. In many cases, administrators and classroom teachers had simply seen over several years how Reading Recovery positively impacted students' literacy skills, confidence, and enthusiasm, and based their endorsement on the effects they observed. Classroom teachers in endorsement schools frequently made comments that they saw "amazing" and "tremendous" growth in their students who received Reading Recovery:

In literacy skills, I think they're able to attack a word better... In writing I see that they're able to get their thoughts down easier... Their enthusiasm for reading has increased tremendously... It gives you goose bumps.

I have personally seen children who have been pulled from my class for [Reading Recovery] excel, where I was not sure they were reachable. You have that doubt, and then they blossom, so it is worthwhile.

In other endorsement schools, the Reading Recovery teachers were what could be termed “high leverage” members of the staff—teachers who were veterans in the building, and whose strength and expertise were acknowledged and admired by both administrators and classroom teachers. Because the staff trusted the person trained in Reading Recovery, they were willing to endorse and commit to the program, even when it was in its first years of implementation, under-implemented (i.e., not all students eligible for the intervention are being served) or both. One teacher leader explained that “a big piece of success of implementation anywhere” is a school “choosing wisely” who they will train for Reading Recovery. A UTC director made a similar comment:

If [schools] choose the right person to train as a Reading Recovery teacher and that person is a part of that community and wanting to stay in that community, they will have someone there for 15, 20 years. And the person quite often takes on a role much like a literacy coach or facilitator. Because the teacher is not only successful with a Reading Recovery student, but also supports the classroom teachers in helping them help the students in the classroom so that they continue to be successful, but also, sharing information from the conferences they attend, and from the readings that they do in their classes. So I think that there's something about that that is more stable.

In schools where Reading Recovery was endorsed by the administrators and staff, Reading Recovery teachers were generally looked upon as valuable resources with much expertise to share. Classroom teachers in these schools described Reading Recovery teachers as “helpful,” “valuable,” and “appreciated,” and there was relatively frequent communication between the two groups. Most often these communications were about Reading Recovery students; classroom teachers sought information from Reading Recovery teachers about their students' progress, and Reading Recovery teachers sought information about how students were performing in the regular classroom. One Reading Recovery teacher commented that it was particularly useful to get information about what kind of work her students were doing in the classroom:

I ask [the classroom teachers] to see student samples from the classroom, just to see what they're working on. And then I also ask them about what an average first grader does, because I see all the low students.

A majority of this communication occurred as teachers were passing in the hallways, or over email. Both classroom teachers and Reading Recovery teachers in endorsement schools explained that they had few opportunities to communicate at length—let alone

FINDINGS

collaborate around instruction—because, as one Reading Recovery teacher put it, “there’s just no set-aside time.” This meant that, while they were supportive of Reading Recovery, classroom teachers were generally not very familiar with the philosophy or instructional strategies of the program. Many classroom teachers commented on how little they knew about Reading Recovery, or made comments that reflected their limited understanding. “They basically do [in Reading Recovery] what I do in guided reading, but one-on-one,” explained one 1st-grade teacher. Others made similar comments:

I really don’t know what goes on in a [Reading Recovery] lesson. She sits with them, they would go over vocabulary, then introduce their book, and maybe they would do some writing during that time. I never did sit down with [the Reading Recovery teacher] to see what exactly they do every single day.

I have no idea what it is. I mean, it’s a half hour of reading. That’s what I’m thinking it is.

While they were glad to have Reading Recovery teachers in their buildings, the lack of formal opportunities (i.e., “set-aside time”) to communicate and collaborate with them led many classroom teachers in endorsement schools to feel that they were not fully benefiting from the Reading Recovery teacher’s expertise. This frustrated some classroom teachers, who felt that Reading Recovery was a valuable resource they were not being given the chance to access. One classroom teacher explained that she and her colleagues wanted to improve their literacy instruction through collaboration with the Reading Recovery teacher, but “We haven’t been offered enough [information] to make an impact on the classroom.” She continued:

I just think there’s a lot of valuable tools that she could be helping train us and others in the district... And I don’t really feel that she has been able to share any of her knowledge... by no fault of her own.

This disappointment about missed opportunities for collaboration was echoed by many Reading Recovery teachers; they knew they had expertise to share, just no opportunity to share it. “I have all this knowledge,” lamented one Reading Recovery teacher, “and I’m not using it the way I could.” Others made similar comments:

I just wish there was more of a chance to share with [the classroom teachers] what I’m doing because I think there’s so many things that we do that would be really positive in their classroom.

Without time to share expertise and collaborate, it was difficult for Reading Recovery teachers and classroom teachers in endorsement schools to develop and coordinate instructional plans that would support students across their day (e.g., by using similar language, prompts, and instructional strategies in both the Reading Recovery and regular classrooms). This meant that classroom instruction did not necessarily reinforce the gains students were making in their Reading Recovery lessons.

Reading Recovery teachers, as well as many classroom teachers, attributed the lack

of opportunities for communication and collaboration to limited involvement by their principals. Indeed, comments by many Reading Recovery teachers in schools demonstrating endorsement revealed that they considered their principals to be more passive—rather than active—supporters of Reading Recovery. “[My principal] is supportive if I go to him,” explained one Reading Recovery teacher. “But he’s not the kind of guy that’s going to swing by and ask how it’s going. He won’t attend the meetings. He’s never seen a behind-the-glass lesson.” Another Reading Recovery teacher commented that the program was being implemented in her school with “a lack of direction from above.” Classroom teachers also noticed the lack of involvement in Reading Recovery on the part of their principals:

Nope, he has not even talked about it except for... maybe saying that [the Reading Recovery teacher] had gotten the grant and that she would be doing it. That’s about all the communication he has had with us.

I’m not sure why [our principal] doesn’t promote [Reading Recovery]. My only thought is that she just doesn’t know enough about it.

For their part, principals in schools demonstrating endorsement described their role as that of a “champion,” “cheerleader,” or “advocate” for Reading Recovery. However, their comments generally revealed that they did not perceive their role to include creating formal structures for knowledge sharing; instead they saw themselves as behind-the-scenes supporters of the program. One principal even addressed this issue directly, commenting that he had purposefully taken “a hands-off approach.” He explained that he does not make suggestions about how classroom teachers should interact with Reading Recovery teachers “because if it comes from me, that kind of takes away [the Reading Recovery teacher’s] leverage.”

Unfortunately, many Reading Recovery teachers commented that the lack of facilitative support by their principal was the most challenging aspect of implementation in their context. They felt that, if the principal put more formal structures in place to support collaboration, the whole school could benefit from their expertise:

I think it would be nice to have more support from my principal... I would love to do some professional development with our lower elementary teachers, kindergarten, first, and second grade. I’ve found that a lot of those early reading behaviors that are addressed in Reading Recovery are getting skipped [in the classroom]... So I feel like if I could do some professional development with early reading behaviors and ways to analyze data and specific prompts you could use, I think that our general student growth would be higher in reading.

I think it would be helpful if the principal would say to teachers “you need to meet on this day, or you need to set aside a meeting and [the Reading Recovery teacher] will be there to do some professional development in Reading Recovery strategies... I would venture to say he will tell you that I need to take on more of a leadership role... but I don’t quite feel comfortable doing that yet. So, I kind of need him to help me with that.

FINDINGS

As these comments indicate, many Reading Recovery teachers in case study schools felt it was the principal's role to create formal structures for collaboration. This opinion is in contrast to some teacher leaders, and clearly some principals, who explained that they felt it was the responsibility of the Reading Recovery teacher to situate herself as a literacy leader in the school, and to initiate collaboration with classroom teachers.

There are several reasons this sometimes proved difficult for Reading Recovery teachers, however. One was the persistent lack of time. As one teacher leader explained “Administrators really want Reading Recovery teachers to be literacy leaders, yet they don't really give them the time to plan for professional development.” Another issue was, as the quote above indicates, that not all Reading Recovery teachers had the confidence to assert themselves as experts, particularly if they were younger than other members of the teaching staff. “I could have a little better communication with the teachers,” explained one Reading Recovery teacher, “but I think I've been timid in that area. I don't want to come off sounding like I'm an expert or anything because I don't really know it all and they're so much more seasoned than I am.”

In the case study schools demonstrating endorsement, Reading Recovery implementation proceeded with fidelity, support, and sustainability; though people across the school context had limited understandings of Reading Recovery, their commitment to successful program implementation did not seem to suffer because of it. The only negative consequence of this limited understanding seemed to be recurring confusion—and sometimes tension—between classroom teachers and Reading Recovery teachers over which students were selected to receive the intervention. Because they did not understand the reasoning behind Reading Recovery's selection of the very lowest-scoring readers (regardless of many other factors, such as language acquisition or behavior), classroom teachers in these schools sometimes raised objections if they disagreed with the selection of students. Sometimes teachers objected to the fact that a student they felt needed Reading Recovery was not chosen; other times the objection was to the selection of a student they felt could not benefit from Reading Recovery—generally due to language acquisition—thereby “wasting a spot” on a student they did not think would improve through Reading Recovery instruction. One classroom teacher expressed her frustration about the students selected to receive Reading Recovery in her school this way:

That's one of the things I don't understand about Reading Recovery. There's one little boy we tried to get [transferred from Reading Recovery to Special Education] because he has a processing disorder... He'll test really high in [Reading Recovery], but then he comes in here [to my classroom], and... maybe the [reading] levels are off because he's nowhere near the level the Reading Recovery tests say he is... I don't know if it's because of the techniques that [the Reading Recovery teacher] uses, but I don't see a lot of carry over into the classroom.

Several Reading Recovery teachers also made comments that suggested they felt pressure from both classroom teachers and administrators around student selection. One Reading Recovery teacher comment that, at her school, “the teachers start to get wiggly about it, and there's a lot of pressure.” She continued:

They're asking, "Why aren't you taking this kid, too?" And they'll call the principal in, and I'll have these meetings and all this kind of stuff... And then you don't feel very supported because people are breathing down your neck and wanting to know "Why are you only working with eight children?"

This is an issue where limited understanding of Reading Recovery's approach and processes presents potential challenges to support and commitment from classroom teachers and administrators, as well as potential setbacks to Reading Recovery teacher perceptions of support.

SCHEMA OF IMPLEMENTATION: INTEGRATION

HIGH UNDERSTANDING/HIGH COMMITMENT

"Reading Recovery is part of the way we do things here."

- School-wide shared understanding of Reading Recovery
- School-wide active problem-solving to ensure program fidelity
- Principal actively engaged with Reading Recovery
- Highly emphasize school-wide impact of Reading Recovery
- Reading Recovery instructional time protected
- Frequent Reading Recovery teacher/classroom teacher communication
- Reading Recovery teacher positioned as literacy leader or resource in the building
- Classroom instruction generally supportive of Reading Recovery

The fourth and final schema of implementation exhibited by our case study schools was **Integration**. This schema was the most common; 11 of our 23 case study schools exhibited aspects of integration. These schools might be considered ideal implementations—in them, we observed widespread understanding of Reading Recovery, widespread commitment to making the program successful in the school context, and a clear sense that Reading Recovery was an embedded component of the school's approach to literacy instruction. In schools exhibiting integration, Reading Recovery was widely described as a valuable part of the overall Response to Intervention plan, sometimes influencing what other programs and interventions were brought into the school. It was also in these schools that both principals and classroom teachers made comments such as "[Reading Recovery] is just part of what we do," and "[Reading Recovery] is part of how we do business."

A hallmark of schools demonstrating integration was a clear commitment to using Reading Recovery to "build capacity" across the school. Principals in these case study schools commented that increasing the literacy expertise of classroom teachers was not just a happy by-product of having Reading Recovery—it was one of the program's main selling points. One principal explained that, in her school, Reading Recovery was unquestionably expected to improve classroom literacy instruction. "The biggest thing," she said, "is we just have to have the capacity for really understanding how kids approach

FINDINGS

unfamiliar text, or how they learn all the strategies that they need to become successful readers and writers.” From her perspective, Reading Recovery helped her classroom teachers build this knowledge. Another principal commented that she wanted to be sure classroom teachers were learning about Reading Recovery instructional strategies “so they can apply it to their practice all day long.” A teacher leader explained that the principal in one of her schools put emphasis on using Reading Recovery to affect classroom literacy instruction:

[The principal wants to make sure] that Reading Recovery is embedded in their system. That it’s not only the Reading Recovery teachers that are the knowledgeable ones, but that the other knowledgeable teachers are the classroom teachers... [The principal] wants all of them to understand literacy at the acquisition stage, especially for struggling readers.

Several principals in case study schools demonstrating integration explained that, in order to extend the influence of Reading Recovery to classroom teachers, they—as leaders in the building—needed to play an active role in the program. While several mentioned that they did not know what “the defined role” for a principal was with regard to Reading Recovery, they described roles for themselves that involved active participation in implementation. In some cases this meant sitting in on all Reading Recovery-related meetings, including student selection, having Reading Recovery students read to them, reviewing student progress reports, or frequently observing lessons. In other cases, it meant being less involved in the day-to-day aspects of the program, but always vigilant about protecting and providing for Reading Recovery. One principal explained that his role was “to provide encouragement, and also the resources and scheduling and time and structures that are necessary to make [Reading Recovery] work.” “My role is to champion those teachers who are doing [Reading Recovery],” said another principal. “It’s to lead them, but it’s more so walking beside them, and rooting for them, and asking them “Are there any other things that you need?”

When discussing what structures are “necessary to make Reading Recovery work,” principals in case study schools demonstrating integration emphasized frequent communication between Reading Recovery teachers and classroom teachers. To encourage this communication, several principals felt it was important for them to first position their Reading Recovery teachers as literacy leaders in the school. One principal emphasized that it was her job to make sure the classroom teachers understood the Reading Recovery teachers to be “literacy leaders, and people who the teachers in first, and even second grade, should be coming to when they have a stumbling block.”

Many principals also explained that it was specifically their responsibility to ensure that there were frequent, designated opportunities for Reading Recovery teachers and classroom teachers to communicate. “It is critically important for the Reading Recovery teacher and the classroom teacher to be able to communicate,” explained one principal. He continued:

Every week it’s important for that classroom teacher to see what are they doing in

Reading Recovery and saying “This student is making progress there, so I need to make sure that I’m pushing him there, too.” My job as principal is to create the opportunities for those conversations to happen.

Communication between Reading Recovery teachers and classroom teachers was facilitated by both informal and formal structures in these case study schools. In addition to informal hallway or lunchtime conversations about Reading Recovery students, classroom teachers in these schools talked about seeking out a Reading Recovery teacher for advice and guidance on their own literacy instruction. “[The Reading Recovery teacher] is very helpful,” commented one classroom teacher. “If I have any questions, she always has information, or she will get it to me—very detailed information.” Other classroom teachers made similar comments:

It won’t even have to be a struggling kid. I had [the Reading Recovery Teacher] in here the other day to watch me teach one of my higher groups, because I felt like I wasn’t exactly sure what they needed to be working on. And she sat there and watched me, and could immediately see “Oh, well this kid, the first thing he does when he gets a hard word, he looks at the picture...”

I’ve been fortunate that the Reading Recovery teachers here are open to letting us go observe lessons so we can pick up their language, the way they structure their lesson, and the prompts that they’re using... I think it’s really affected my teaching in terms of being able to support students.

More formally, the case study schools where Reading Recovery was well integrated had the “set-aside time” for teacher collaboration mentioned in the previous discussion of endorsement. The nature of this time varied across schools; in cases new opportunities were created, such as literacy team meetings organized around specific students, or shared planning time. In other cases, Reading Recovery has been integrated into existing structures, such as 1st-grade team meetings. One Reading Recovery teacher explained that she and her Reading Recovery colleagues “are seen as members of the 1st-grade team. And I would say that we’re seen as experts with not only struggling readers, but an expert in... what happens in first grade with all readers.”

Across all case study schools exhibiting integration, classroom teachers described Reading Recovery teachers as valuable resources that were “fabulous,” “wonderful,” and—as one classroom teacher explained—“amazing to watch and work with.” In School N, a classroom teacher explained that Reading Recovery “is very highly regarded here. It’s one of our high priorities, and the Reading Recovery teachers are regarded with the utmost respect.” Many Reading Recovery teachers also commented that they felt well supported by both teachers and administrators; one mentioned that she felt “well respected and honored for being a Reading Recovery teacher,” and another commented that she felt like she was “treated like gold.” Yet another explained “If I bring something to one teacher, they’ll make copies for the whole team and say ‘[the Reading Recovery teacher] said to do this.’ So I feel like they’re valuing everything that I say.”

FINDINGS

Collegial and collaborative relationships between Reading Recovery and classroom teachers made several things possible in case study schools exhibiting integration. First, they were able to collectively problem-solve around Reading Recovery to make sure it was implemented with fidelity, and with the least disruption possible. Several classroom teachers and administrators remarked that it was the norm for Reading Recovery teachers to collaborate with others around scheduling student lessons. One classroom teacher explained:

[The Reading Recovery teachers] do a really good job of asking us at the beginning of the year to provide a schedule for what a general classroom day would look like. And then they line up all the 1st-grade teachers, their schedules, and then they select the times when we're in the classroom... And then they do the best they can with creating a rotating schedule.

Several principals commented that this collaboration had influenced how classroom teachers perceived Reading Recovery. One explained that “the classroom teachers have stopped treating Reading Recovery as separate” because “the lines [between Reading Recovery and the classroom] have been blurred” through frequent communication. Another principal explained that collaboration had been a key to developing classroom teacher enthusiasm for Reading Recovery:

The Reading Recovery teachers are very flexible, which has significantly increased 1st-grade teacher buy-in, because they are instantly seen as willing participants of the first grade, and willing partners of the first grade.

The outcome of this collective problem-solving was ensuring Reading Recovery lessons could happen every day. Several Reading Recovery teachers in integration schools talked explicitly about how well protected Reading Recovery is. One explained “I don't have to give up my lunch. I don't have to give up my conference period. I don't do the lessons before or after school. It's within the regular day. So that's the ideal.”

A second outcome made possible by good relationships and supportive structures in schools demonstrating integration was extensive knowledge sharing between classroom teachers and Reading Recovery teachers. Classroom teachers in these schools not only understood Reading Recovery's approach and instructional strategies, but frequently used them in their own classrooms. Teacher leaders, Reading Recovery teachers, and classroom teachers all described classroom literacy instruction in these schools as well aligned with, and able to support, Reading Recovery. A classroom teacher commented:

I think our Reading Recovery teachers have worked really hard to work with us to help us make sure we're using similar language and strategies with kids to try to make it seamless, and try to make it so that even the other children in our class are still benefiting from the Reading Recovery teachers even if they're not their students.

Many classroom teachers described using prompts, sound boxes, letter sorting, cut-up sentences, and the “slow check” method in their classroom literacy instruction—all strategies they learned from Reading Recovery teachers. They also described structuring their guided reading based on their understanding of Reading Recovery:

A lot of the Reading Recovery lesson is very similar to what we do in the classroom, because we’ll be doing some of those same things in our guided reading groups. We re-read a previously read book, we introduce a new book, read that book... we work on writing afterwards.

Reading Recovery teachers in these schools corroborated that classroom teachers are doing literacy instruction in a way that is aligned with the program. One explained: “[A focus on] meaning-making, having understanding and comprehension of the story... writing continuous text... is definitely what’s going on in [classrooms].”

Classroom teachers in integration schools also frequently commented that, in addition to learning new instructional techniques, they were thinking differently about teaching due to their interactions with Reading Recovery teachers:

I really thought it was interesting how little [the Reading Recovery teacher] said when the student was reading. And I think that’s something I learned about my own instruction. Sometimes I have to actually sit on my hands and say nothing, because you want to be the person that comes in and rescues them. The best thing for them is to just let them be. I think also really focusing on one thing in their reading that you want them to be focusing on, instead of trying to fix everything.

I think as classroom teachers, we tend to really focus on “Okay, what can’t [the students] do?” and I think [learning from the Reading Recovery teacher] has really changed the way I’m looking at my whole class... “Okay, what skills does this kid have?” and really observing and seeing how to untangle things and not make assumptions.

It is interesting to note that schools exhibiting integration had been implementing Reading Recovery anywhere from two to over 20 years. In School B, the newest adopter in our case study sample, there was high-level district support for the program, but principals in each building decided whether to implement Reading Recovery. The principal of School B was described by the teacher leader as a “strong instructional leader,” and a “strong advocate for Reading Recovery;” the teacher leader remarked on the principal’s willingness to give Reading Recovery teachers the freedom and flexibility they need to perform their duties and collaborate with classroom teachers. The classroom teachers were described by both the teacher leader and principal as “very supportive” of Reading Recovery. The classroom teachers themselves described the program as able to affect “enormous changes” in students, leading them to “blossom” and become “motivated readers.” One classroom teacher in School B commented that Reading Recovery “is the best thing, bar none, that the county has done in the 28 years I have been teaching here.”

FINDINGS

The Reading Recovery teacher, classroom teachers, and principal of School B explained that there was a high level of communication and coordination across the building, and that Reading Recovery had become well integrated into the whole school's approach to literacy instruction. Classroom teachers observed Reading Recovery lessons, and used techniques shared by the Reading Recovery teacher—such as prompts and Running Records—to guide their classroom instruction. The teacher leader commented that “one reason [Reading Recovery] has worked so well at this school is because their [classroom] instruction matches what we have been doing [in the Reading Recovery classroom].”

School P had one of the longest-running implementations of all schools in our case study sample; Reading Recovery was in its 20th year during the time the school was studied. In this case, there was also high-level district support for Reading Recovery, but the district chose to implement the program in all elementary schools, regardless of principal choice. Here, as in other schools demonstrating integration, the principal and classroom teachers described themselves as “very supportive” of Reading Recovery. The Reading Recovery teacher corroborated these statements, commenting that she—and the program in general—were highly regarded.

What is particularly interesting about School P is the extent to which Reading Recovery was being used to build capacity across the school, and even across the district. Reading Recovery training, delivered by the teacher leader, was extended to teachers across grades at all schools. These teachers, called Reading Recovery Advocates, agreed to work with one student every day during the year. During the year School P was studied, the district extended the Advocate training to a group of principals who expressed interest in learning about the program. These administrators attended 15 weekly training sessions, and worked with one student three times per week. There was such high regard for Reading Recovery in this district that district administrators decided all literacy coaches needed to be Reading Recovery-trained as well.

As in the other schemas of implementation, the cost of Reading Recovery was mentioned by many respondents in schools demonstrating integration. However, in these schools, the comments were quite different than those in schools exhibiting other schemas. Principals and teachers in integration schools frequently remarked that, though they perceived Reading Recovery as an expensive program, they thought the cost was justified. One teacher commented that she felt differently about spending funds on Reading Recovery than she had about other programs in the past: “I have to say, having taught for as long as I have, there have been so many programs where you're just like ‘Oh my gosh, this is just one more thing, and it is not the best place to put the money.’ But, [Reading Recovery] helps, it works... I am a supporter of it.” A principal remarked that the teachers in his school seemed to feel the same way:

If we didn't have Reading Recovery, we would have 1.5 more [early intervention] teachers, so ten more segments of [early intervention] could be happening. But not a single 3rd-, 4th-, or 5th-grade teacher is willing to give up Reading Recovery to have [more early intervention] service. Not one of them.

Several principals also commented on how important they felt it was to maintain funding for Reading Recovery into the future. “If I got a phone call today that the budget was due tomorrow, and we’re not having Reading Recovery,” explained one principal, “I would do everything in my power to either get that funding back, or just find a way to fund it myself... by reallocating our school funds somehow.” Other principals made similar comments:

With budget constraints, there are always decisions that have to be made with respect to programming. But Reading Recovery has never been one that is up for debate.

[If the district cut funding for Reading Recovery] I would have tried to go beg, borrow, and then steal. I mean I would have gone to some of my community members to try to source the funds. I would have done that. It’s that important.

This view is in particularly stark contrast to those expressed by administrators and teachers in schools demonstrating the isolation schema, reinforcing the notion that communication and understanding are key to support of Reading Recovery.

DISCUSSION

In the course of our investigation of Reading Recovery implementation, we often heard informants in different positions within the program identify two main factors that—from their perspectives—most heavily influence the impact and sustainability of Reading Recovery in a school: 1) how long the program has been in place, and 2) level of “coverage” (i.e., whether all eligible students are being served by the program). These propositions certainly stand to reason, but—as the case studies indicate—they offer incomplete answers to what conditions lead to Reading Recovery implementations that are supported, successful, and sustainable.

Our 23 case study schools reveal that having Reading Recovery for many years does not necessarily mean the program is integrating into the school or being implemented with high fidelity, and it does not always lead to a level of commitment that would protect the program from termination. The cases also reveal that, while full coverage may increase the number of students who exit Reading Recovery reading at the average level of their peers, it does not buy commitment and cooperation from the administrators or classroom teachers in a building. Factors beyond time and level of implementation are clearly influencing whether Reading Recovery is implemented with fidelity, support, and sustainability in different contexts.

The findings presented thus far have delved into each schema, exploring understanding and commitment as two school-level factors that cannot be entirely accounted for by length of implementation and level of coverage, but that—in their variation—shape conditions of Reading Recovery implementation. The next analytical step is to look *across* the schemas of implementation to explore what seems to shape understanding and commitment, and the relationship between these two factors.

DISCUSSION

Taking this step, two important phenomena emerge as influences on understanding and commitment and, therefore, as indirect influences on how Reading Recovery implementation progresses in context: **communication around Reading Recovery**, and **principal engagement** related to the program. As the following discussion will explain, both communication and principal engagement varied across the case study schools in their relative presence and prominence, and in the particular shape they took.

Communication about Reading Recovery in case study schools ranged in frequency (from never to daily), and also in scope (from narrowly focused to comprehensive). The case studies suggest that the particular type of communication about Reading Recovery happening in a school is related to the particular type of understanding people in the building develop about the program.

Principal engagement with Reading Recovery also varied across the case study schools in shape and scope—some principals had almost no contact with the program or the Reading Recovery teacher, while others were heavily involved in multiple aspects of implementation. According to the case studies, the nature of a principal's engagement with Reading Recovery is related to how committed they, and classroom teachers, are to implementing the program with support, fidelity, and sustainability.

Communication

The kind of talk and conversation going on around Reading Recovery varied across the different schemas of implementation. In some schools there seemed to be relative silence—there was very little (sometimes no) communication about the program flowing in or out of the Reading Recovery classroom, or into the school from other sources. This was the case in both schools exhibiting isolation, and communication was also limited in some obstruction and endorsement schools. The absence of communication had important consequences in several case study schools. Most basically, when there was no talk within the building about Reading Recovery, there was no information circling. This meant administrators and classroom teachers were unable to build an understanding of Reading Recovery's theory, general approach to literacy, or instructional strategies. In the presence of an information vacuum, administrators and teachers in these schools relied on their assumptions about Reading Recovery, and whatever insights they may have gleaned from other sources, in order to develop their understanding of what Reading Recovery is and how it relates to them as participants in the school context. In schools demonstrating isolation and obstruction, the understandings of Reading Recovery developed within the school were not always well informed, and frequently resulted in implementation strategies that threatened the program's fidelity and sustainability. This was evident when Reading Recovery teachers also held roles that conflicted with their program duties, or when Reading Recovery time was not protected.

In addition, when there was little communication about the program in a context, people within the context could not develop a sense of how their positions in the building related to Reading Recovery. There were several cases where lack of understanding resulted in

role confusion; Reading Recovery teachers, administrators, and classroom teachers all expected certain provisions or supports from each other that went uncommunicated and, therefore, unfulfilled. Without clear understandings of who should do what in these instances, administrators stayed quiet about the program, Reading Recovery teachers felt unsupported, and classroom teachers were not given any guidance from either direction about what they could do to help Reading Recovery be successful in their school.

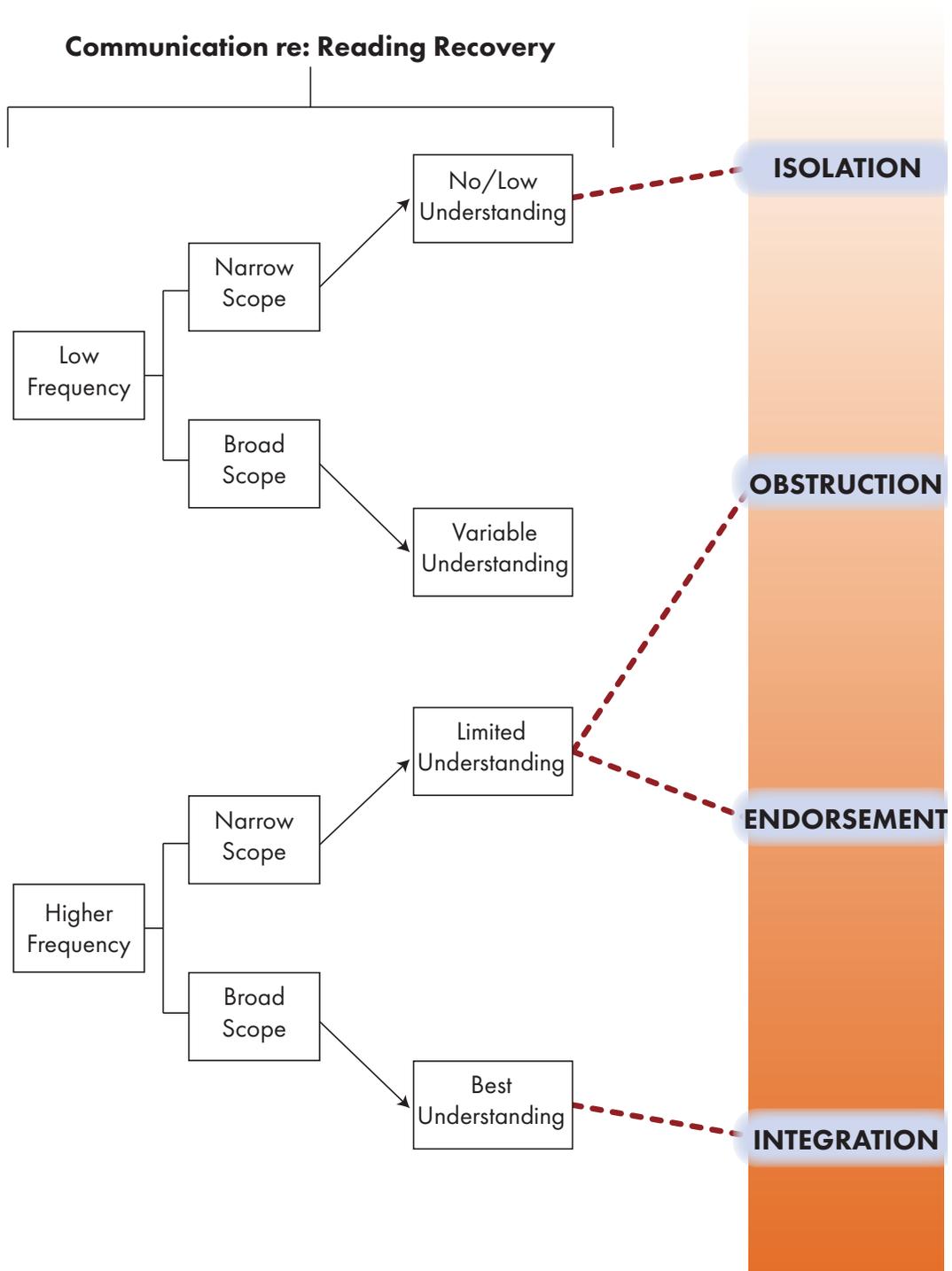
Additionally, without communication about Reading Recovery, it was very difficult for people within the context to develop a sense that the program had value for their building, either within the Reading Recovery classroom or beyond. And in cases where the only (or simply the loudest) message circulating about Reading Recovery was related to the perceived high cost of the program, cost quickly became the focus of administrators and teachers. Examples of such situations were evident in the previous descriptions of isolation and obstruction schools. In these cases, both administrators and classroom teachers discussed the cost of Reading Recovery as the program's most prominent feature, to the exclusion of any student- or school-level benefits it offers. When Reading Recovery did not hold value for administrators or teachers, commitment to maintaining the program was extremely difficult to engender. And without understanding or commitment, implementation faced challenges to fidelity, support, and sustainability.

In schools exhibiting isolation and obstruction, it was clear that lack of communication, lack of understanding, and lack of commitment to Reading Recovery were related. Schools demonstrating integration reinforce the idea that these three issues are connected, though in a more positive direction. In integration schools, communication was frequent, understanding was high, and commitment was unquestionable.

However, schools demonstrating endorsement suggest there is a more complicated view we should take of the relationship between communication, understanding, and commitment. In these schools, certain types of communication about Reading Recovery were frequent, but understanding of the program was low, suggesting there are different types and levels of communication that lead to different types and levels of understanding.

This nuance becomes clearer when comparing endorsement schools with integration schools. In schools exhibiting endorsement, the talk between Reading Recovery teachers and classroom teachers was frequent, but narrowly focused on Reading Recovery students; there were few opportunities for them to communicate or collaborate around literacy instruction more broadly. Communication between Reading Recovery teachers and administrators in these schools was generally focused on program results and funding or support, not on whole-school benefits of Reading Recovery. The result of the narrow scope of communication in schools demonstrating endorsement was a limited understanding on the part of classroom teachers and administrators with regard to the school-wide value of Reading Recovery and the roles they could play in extending the success of Reading Recovery beyond the program classroom. While they clearly saw the benefits of Reading Recovery for students served, classroom teachers in endorsement schools did not have much sense of how they themselves, or their other classroom students, could benefit

Figure 6.1 Communication Around Reading Recovery Within a School Context



from the expertise of the Reading Recovery teacher. They also did not have a sense of how they could support Reading Recovery students in their own classrooms. Likewise, administrators were happy to have the program, but did not seem to understand that they could (and arguably should) facilitate the sharing of literacy expertise between Reading Recovery teachers and classroom teachers, thereby extending the benefit of the program to the whole school. In endorsement schools, limited communication about Reading Recovery led to limited understanding of it.

By contrast, in schools exhibiting integration, communication between Reading Recovery teachers, classroom teachers, and administrators was not only about Reading Recovery students, but about literacy more broadly, and what contribution the program could make to increasing capacity for excellent literacy instruction across the school. In these schools, classroom teachers actively collaborated with Reading Recovery teachers, and used the understandings they developed through this collaboration to improve their classroom literacy instruction. Because the scope of communication extended beyond the Reading Recovery classroom and Reading Recovery students, classroom teachers understood not only what happened in Reading Recovery, but how they could parlay the program into school-wide improvement. This comparison suggests the scope of communication is an important contributor to level of understanding.

The comparison of endorsement and integration schools can also help us parse the complicated connection between understanding and commitment, which, when looking particularly at isolation and integration schools, appears fairly straightforward. Figure 6.1 below traces the flow from different kinds of communication to different types of understanding. In schools demonstrating endorsement, understanding of Reading Recovery was low. And while this did limit the extent to which classroom instruction could support and align with Reading Recovery, as well as how widespread the benefits of the program could be in the school, it did not seem to significantly dampen anyone's apparent enthusiasm for the program. Something other than understanding was behind commitment in these schools—something seemingly absent from schools demonstrating isolation and obstruction. Many things could be behind this commitment-without-understanding; one we can glean from the case studies is the important influence of principal engagement.

Principal Engagement

The nature of principal engagement with Reading Recovery also varies greatly across the schemas of implementation, and while it does overlap with communication, the overlap is not perfect. That is, principal engagement, or lack thereof, is visible in areas outside communication about the program. For instance, principals in schools demonstrating isolation, and many in schools demonstrating obstruction, had very low levels of engagement with Reading Recovery. These principals were not generally involved in any aspects of the program beyond providing space and materials for the Reading Recovery teacher to use to conduct her lessons. This was evident both in these principals' lack of communication about the program (as discussed above), and in other decisions, such as

DISCUSSION

choosing not to participate in discussions about student selection, or choosing to assign Reading Recovery teachers to duties that interfered with their lessons. Through their actions (and often lack thereof) regarding Reading Recovery, principals in isolation and many obstruction case study schools demonstrated their lack of commitment to, and prioritization of, the program.

Importantly, this lack of engagement by principals is not benign. Classroom teachers in these schools clearly received a message about Reading Recovery from the principal loud and clear. They observed their principal's lack of engagement with the program, realized there were no expectations that they would interact with or support the program, and likewise did not prioritize Reading Recovery. This lack of support by both administrators and teachers meant that the processes and routines that support successful implementation were never put in place—there were no opportunities for Reading Recovery teachers and classroom teachers to collaborate, the Reading Recovery teacher was not positioned as an expert resource in the building, and Reading Recovery time was not protected. Thus, through both passive and active means, principals who chose not to engage with Reading Recovery prevented the development of school-level understanding of, and commitment to, the program. In these schools, Reading Recovery was implemented with threatened fidelity, little to no support, and highly tenuous sustainability.

The relationship between principal engagement and commitment is perhaps easiest to see when discussing the integration schemas. In schools exhibiting these schemas, principals chose to actively involve themselves in more aspects of Reading Recovery. Many of them participated in student selection, observed lessons, attended behind-the-glass training sessions, and used Reading Recovery data in decision-making about the program and the school as a whole. The high level of active engagement by these principals manifested in several ways. First, they were visible and vocal in their enthusiasm about Reading Recovery—classroom teachers saw their principal frequently interacting with Reading Recovery teachers and students, and heard them talk about how important the program was to the school. Simply observing this level of enthusiasm about Reading Recovery sent the message to classroom teachers that it was important to support the program.

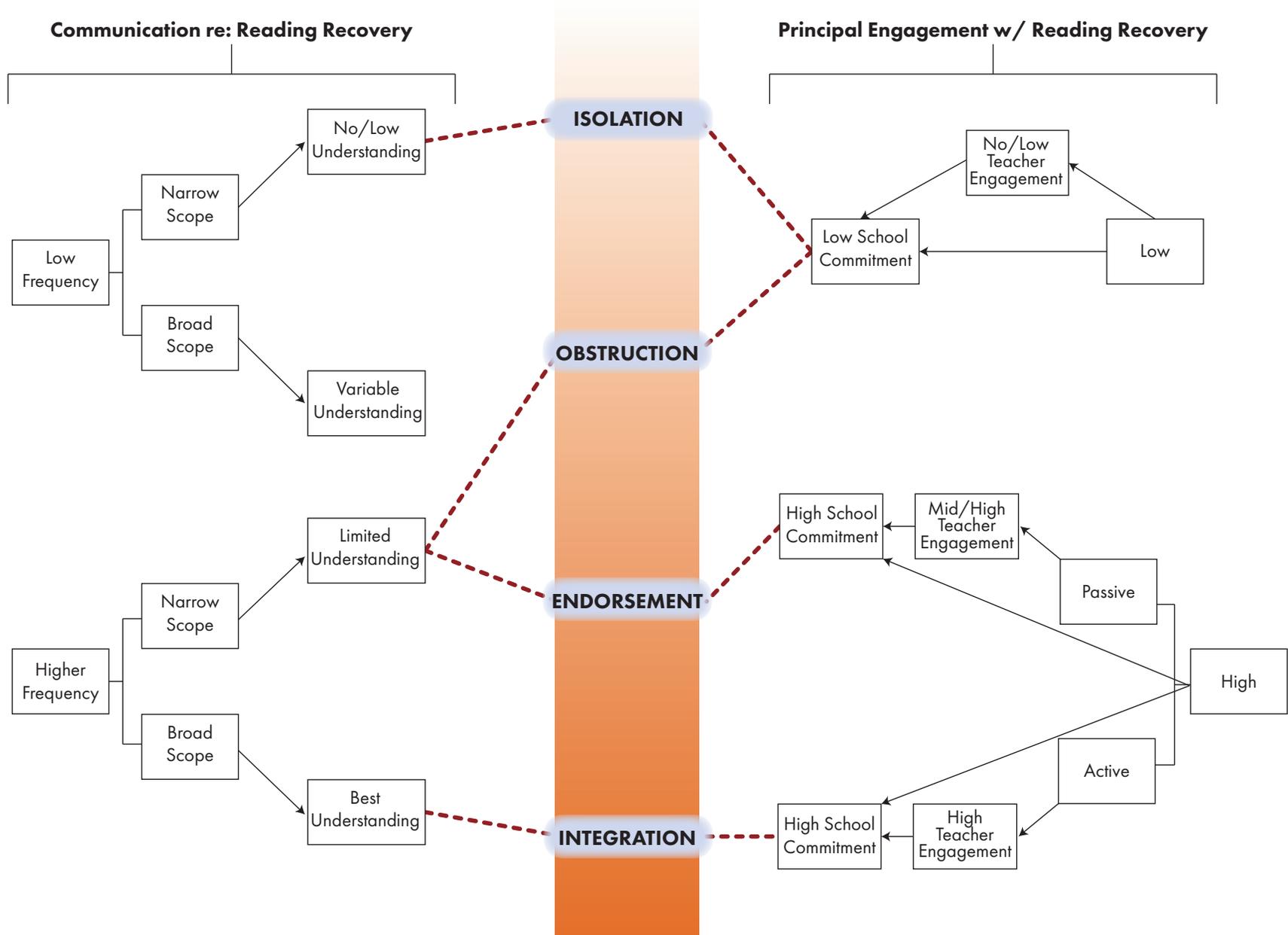
Second, actively engaged principals positioned the Reading Recovery teacher as a literacy expert and leader in the building. That is, they referred classroom teachers to the Reading Recovery teacher for guidance on literacy instruction, they highlighted the Reading Recovery teacher's expertise in staff meetings or professional development, and may have even stated formal expectations that their classroom teachers would observe Reading Recovery lessons and consult with the Reading Recovery teacher on issues of literacy instruction. These principals also put structures in place (e.g., the “set aside time” previously discussed) that made it possible for classroom teachers to interact with Reading Recovery teachers in ways that built capacity across the school. Through these moves, principals demonstrated the importance of Reading Recovery with both their words and actions, and enabled classroom teachers to develop a sense of the program's value.

It is important to note that this kind of principal engagement (i.e., highly involved and active) is not the definitive mark of principal commitment to Reading Recovery. In endorsement schools, for instance, principals were highly committed to implementing the program with fidelity, support, and sustainability, but were more passively engaged than those in integration schools. While they, too, were vocalizing the importance of Reading Recovery, were positioning the Reading Recovery teacher as an expert and leader in the building, and were encouraging classroom teachers to collaborate with the Reading Recovery teacher, they were not specifically creating times and spaces for knowledge sharing to happen. This was likely because their understanding of Reading Recovery was low, and they therefore did not realize the role they, or the classroom teachers, could play in successful implementation. It was this distinction in the particular shape a principal's high level of engagement took (i.e., passive versus active) that distinguished schools that were categorized as demonstrating endorsement from those categorized as demonstrating integration.

Where this distinction becomes most important is when considering whether a school can achieve building-wide benefits through their particular mode of Reading Recovery implementation. Schools demonstrating endorsement had frequent, but narrowly focused communication about Reading Recovery. They also had a high level of commitment to implementing Reading Recovery with fidelity, support, and sustainability, but a passively engaged principal that did not (for whatever reason) create structures for sharing information and building instructional capacity across the school. This resulted in schools where teachers and administrators were committed to doing everything they knew to do to support successful Reading Recovery implementation. However, the impact of the program, and the expertise about literacy instruction, was largely isolated to the Reading Recovery classroom in these schools. Still, this was an excellent outcome for Reading Recovery—it resulted in an implementation that was protected, prioritized, and supported.

The case study schools that had moved beyond excellent toward exceptional implementation made two important changes: to the scope of communication, and to the shape of principal engagement. Case study schools demonstrating integration had frequent communication that included exchanges about Reading Recovery students, as well as literacy instruction more broadly. Based on the scope of these exchanges, both classroom teachers and administrators understood what roles they could play to support implementation. These schools also had principals who demonstrated a high level of commitment to implementing the program with fidelity, support, and sustainability, and who also created structures to enable teachers to share expertise and build capacity. This resulted in schools where teachers and administrators understood there was more they could do to support implementation, and were committed to doing it all. This allowed the impact of Reading Recovery to move beyond the program classroom, benefiting teachers across the school. Figure 6.2 maps the flows of communication, understanding, principal engagement, and commitment that lead to each schema of implementation.

Figure 6.2. Theory of Action for Schemas of Implementation



CONCLUSIONS

The most fundamental and significant takeaway emerging from this analysis is that schools implement Reading Recovery with great range and variation, and that different modes of implementation enable different levels of program impact and institutionalization. The depth of this analysis also allows us to add texture to this statement, bringing to light the subtle ways implementation is shaped by school-level conditions. It is the fine-grained nuance enabled by this analysis that suggests important lessons for what contextual conditions enable Reading Recovery implementation to proceed with support, fidelity, and sustainability.

The first important lesson that emerges from the case studies is that enthusiasm for Reading Recovery is not the same as *support* for Reading Recovery. Reading Recovery generally enters a school or district setting based on the espoused commitment of administrators, many of whom devote budgetary resources to the program. However, as we saw in schools demonstrating isolation and obstruction, commitment—verbal or monetary—does not guarantee that Reading Recovery will be well supported in a setting. In some cases the lack of support is due to lack of understanding of Reading Recovery, in other cases it is due to shallow interest in the program. Whatever the cause, lack of support for Reading Recovery (e.g., lesson time not prioritized; little teacher communication and collaboration) threatens the impact, fidelity, and sustainability of the program in a setting.

This finding suggests that Reading Recovery personnel may find important information about the program's potential impact and sustainability in a particular setting by looking past what administrators say about their commitment to Reading Recovery, to what structures the school puts in place to actively support the program. When support structures are not in place, it will be important to determine whether this is because school-level actors do not understand Reading Recovery, or because there are other factors prohibiting them from highly valuing and fully supporting the program. This also implies that Reading Recovery personnel should be attentive to how Reading Recovery aligns or interacts with existing routines, structures, and norms in the building, and be prepared to deal with any misalignment that would present challenges to supporting or sustaining implementation.

A second, related lesson that emerges from the case studies is that Reading Recovery must gain value in some way—in the eyes of school-level actors—if they are going to support a high-quality, sustainable implementation. Fortunately, the case studies suggest that increasing understanding of Reading Recovery (which could prove difficult in some settings) is only one way to engender support for the program. As we see when contrasting schools demonstrating elements of endorsement and integration, support for the program does not require everyone in the building to understand the nitty-gritty details of the theory or pedagogy underlying Reading Recovery. Teachers and administrators may be enthusiastic supporters of the program, and do everything in their power to ensure it is implemented with fidelity and sustainability, without having a deep knowledge of its philosophy or procedures.

CONCLUSIONS

However, in order to be well supported, Reading Recovery must gain value somehow, and it is not always true that student impact data can make the case for the program alone. The case studies suggest that, in situations where school-level actors do not have a deep understanding of Reading Recovery, it is important that they have trust in either the Reading Recovery teacher(s), or the administrator who is asking them to support the program through cooperation and communication. As we saw in several case studies, questions about the effectiveness of the program, or concerns about the cost of Reading Recovery relative to the number of students it serves, were overcome by the level of confidence teachers and administrators had in the people associated with Reading Recovery in their school or district. This finding reinforces the notion that selection of teachers to be trained in Reading Recovery is a high-stakes matter; it also suggests that the position held by a teacher in a building (e.g., seniority, centrality to teacher networks) may be an important consideration when selecting teachers for training.

Finally, a third lesson that emerges from the case studies is that *if* it is the goal that Reading Recovery influence literacy instruction school-wide, certain supports must be in place. Most fundamentally, if the impact of Reading Recovery is to extend beyond the program classroom, the program—and personnel associated with the program—must be thoughtfully and deliberately positioned by administrators. It is not wise to rely on the efforts of the Reading Recovery teacher alone to facilitate a wide reach, as she may not have the time or flexibility, or carry enough social capital in her building, to engender support on her own. Instead, administrators must state the value of Reading Recovery, vocally advocate for the program, facilitate communication and collaboration between Reading Recovery and classroom teachers, identify the Reading Recovery teacher as a literacy expert, and clearly communicate each person's role in the program (including their own) if they hope to spread the impact of Reading Recovery across the school.

This suggests that positioning of the program and Reading Recovery teachers in a building should be conscious part of initial program adoption, as well as ongoing implementation—something that would require particular attention by administrators, teacher leaders, site coordinators, and even UTC directors. This would be a departure from the way implementation begins and proceeds in many settings, where Reading Recovery is introduced to a school without creating a supportive niche for the program from the outset. While this may allow Reading Recovery to produce significant gains with Reading Recovery students, it does not allow the program to demonstrate its full potential.

Learning from Scale-Up: Lessons from the Four-Year i3 Evaluation of Reading Recovery

The i3-funded scale-up of Reading Recovery was one of the most ambitious and well-documented expansions of an instructional program in U.S. history, and it was highly successful. Led by Jerry D’Agostino and Emily Rodgers at The Ohio State University (OSU), the scale-up met or exceeded all of its expansion goals. With four-year average treatment effects for participating schools ranging from .38 to .99 standard deviations—among the largest documented for a curricular intervention—across several outcome measures, the project demonstrates the feasibility of effectively scaling up an educational intervention.

Despite its ultimate success, the scale-up was not without its challenges. The tasks of recruiting and training nearly 4,000 new Reading Recovery teachers stretched the program’s existing infrastructure and taxed its personnel in unprecedented ways. Faculty trainers and training center staff at OSU’s 19 partner universities struggled to keep up with the volume of grant-related paperwork, and to navigate university bureaucracies that balked at the project reporting and reimbursement requirements. Teacher leaders and trainers sought support for recruitment efforts that altered their job responsibilities in ways few had anticipated. At the same time, the independent evaluation brought new scrutiny of program operations and raised the stakes on ensuring the quality of training and school-level implementation.

Navigating these challenges demanded a programmatic commitment to reflection and refinement that, in CPRE/CRESP’s view, was a major part of the scale-up’s success. This report documents multiple concrete examples of efforts on the part of program and project leadership to learn from and respond to both internal and evaluators’ feedback. Steps taken by the North American Trainers’ Group—the key leadership entity of Reading Recovery in the U.S.—to address inconsistencies in schools’ processes for selecting Reading Recovery students is one such example (See Chapters 1 and 4). In our judgment, Reading Recovery as a nationwide system is more nimble, organized, and responsive now, at the end of the scale-up process, than it was at the outset.

CONCLUSIONS

In detailing these accomplishments, the preceding chapters present the findings and methods of one of the most comprehensive evaluations ever implemented in the field of education. By offering an in-depth, implementation-focused, mixed-methods analysis of the scale-up process in the context of a major impact study, this report represents a contribution to the growing body of research on the conditions for and impacts of scale-up. In addition, with 6,888 student participants in more than 1,200 schools nationwide, CPRE/CRESP's randomized controlled trial (RCT) of immediate impacts in the scale-up schools is among the largest ever conducted. The study's matched-pairs design accommodated Reading Recovery's program model and ensured that all students identified for Reading Recovery were able to receive treatment during first grade. Because pairs were dropped if either the treatment or the control student was missing outcome data, the design produced no differential attrition—nearly unheard of in RCTs. The RCT's rigorous design and large sample offer strong evidence of the effects of Reading Recovery on the progress of struggling students, both in the study sample overall and in the two subgroups of interest—English Language Learners and students in rural schools (see Chapter 2).

The regression discontinuity (RD) study of Cohort 1 impacts that is discussed in Chapter 3 assessed program effects at three time points: at the mid-first grade conclusion of the randomized trial; at the end of first grade; and again when the cohort reached third grade. The RD analysis of mid-first grade impacts on treatment students replicated the RCT's large effect-size finding as well as its observed variability, with school-level impact estimates ranging from near zero to more than one standard deviation. The RD study also demonstrated that Reading Recovery's effects on treated students persisted at the end of first grade, months after they completed the intervention. The RD study did not find evidence of a significant overall impact on third grade state test scores; however, this analysis had low statistical power because of the limited amount of available data.

The evaluation paired these rigorous experimental and quasi-experimental investigations with an in-depth implementation study that established strong fidelity of program implementation in the scale-up schools in all four years of the study, supporting the validity of the experimental findings. The implementation study, which began with the general goal of documenting fidelity, evolved to examine the determinants of Reading Recovery's success at both lesson and school levels. By identifying key aspects of Reading Recovery's implementation that are not addressed by the program's well-documented model, our findings illustrate the importance of a deeper and more comprehensive investigation of implementation.

Extending beyond this evaluation of Reading Recovery's impacts and implementation under the i3 scale-up, a few questions linger. The RD study provides evidence of significant effects on treated students at the end of first grade. While impacts were not seen in the RD analysis of 3rd-grade state test scores, the data currently available for assessing impacts in 3rd grade are too sparse to produce conclusive findings. Future large-sample research will examine Reading Recovery's effects beyond first grade. Secondly, the RCT consistently revealed significant school-to-school variation in impacts. Findings

from our implementation research on instructional strength in Reading Recovery (Chapter 5) and on its integration at the school level (Chapter 6) offer clues as to potential sources of that variation. However, more research is needed to fully explore the relationship between these factors and program impacts.

Finally, while this evaluation provides a comprehensive picture of Reading Recovery at scale, questions remain as to the feasibility of sustaining this expansion following the i3 funding. This evaluation has documented several programmatic strengths—chiefly, operational responsiveness and evidence of impact—that Reading Recovery can leverage to support ongoing and effective programming at scale.

Equation A1. Statistical Model for the Randomized Control Trial

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

$$Y_{ijk} = \beta_{00k} + \beta_{1jk}(Pretest) + \beta_{20k}(Trt) + \sum_{i=3}^5 \beta_{ijk}(Year) + \sum_{i=6}^8 \beta_{ijk}(Trt * Year) + \gamma_j + \alpha_k + \varphi_k(Trt) + \varepsilon_{ijk}$$

With:

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

Where:

- Y_{ijk} is the posttest outcome score for student i from pair j in school k
- β_{00k} is the model intercept
- β_{1jk} is the slope coefficient for the pretest covariate
- β_{20k} is the main effect for treatment
- Trt is the treatment assignment indicator, with 1=treatment and 0=control
- $\beta_{\{3,4,5\}jk}$ are the main effects for Year, with SY 2014-2015 as reference
- $Year$ is the Year indicator
- $\beta_{\{6,7,8\}jk}$ are the interaction effects for Year and treatment
- 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 a separate variance estimate, $\sigma_{T,C}^2$ for treatment versus control groups

Table A1: Number of Schools Selected to Contribute Students to the RCT by School Year

	Recruited for RCT			Implemented RCT	
	New ^a	Rotation ^b	Total	Total	Percent
2011-2012	207	0	207	159	77%
2012-2013	148	200	348	267	77%
2013-2014	66	225	321	289	90%
2014-2105	138	476	614	539	88%
Total	559	931	1490	1254	84%

^aNew schools were randomly sampled from those newly recruited to the i3 scale-up.

^bRotation schools are i3 schools recruited to the i3 scale-up in previous years.

Figure A1. Consort Flow Diagram for the Reading Recovery i3 RCT

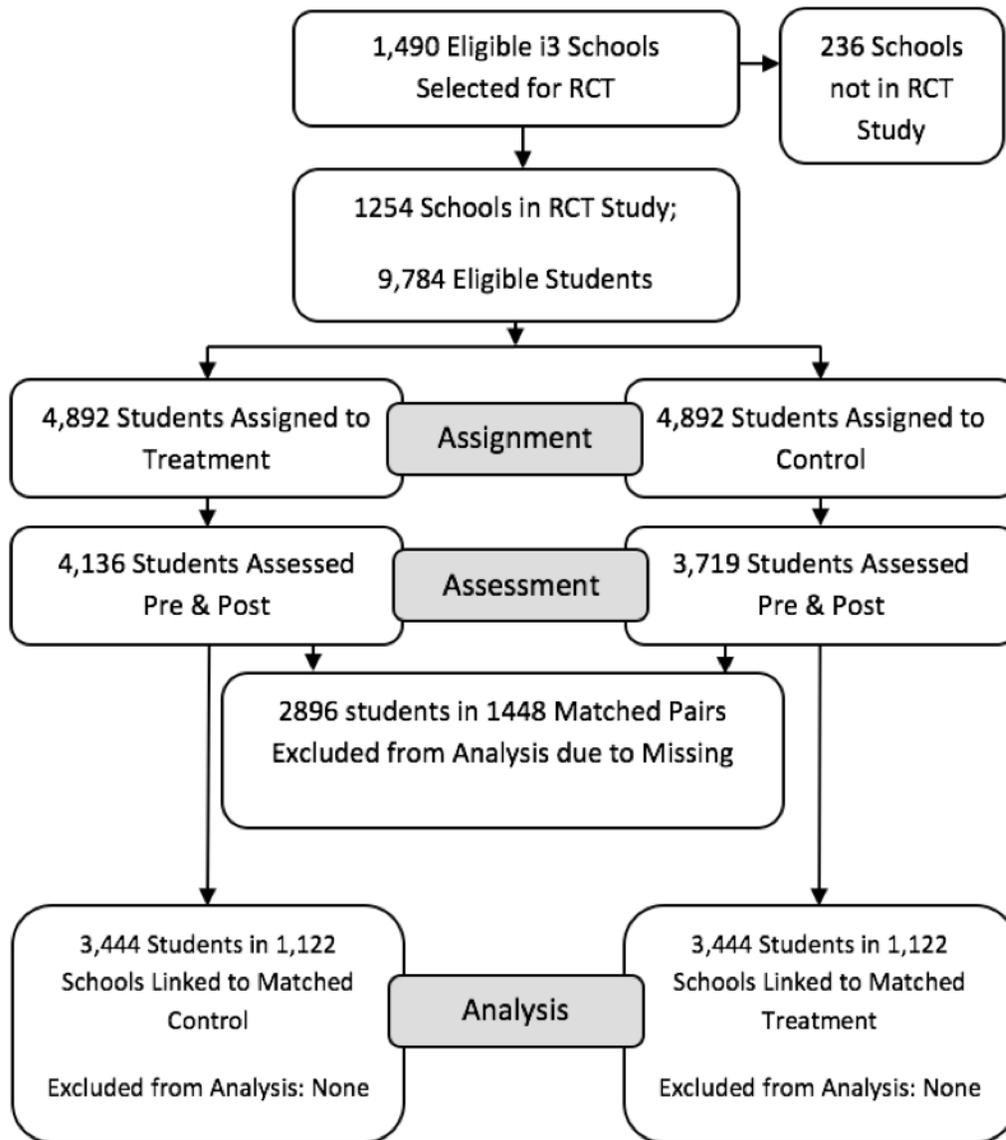


Table A2. CAP sample by year

	2011	2012	2013	2014	Total
Total students randomly assigned					
Treatment	624	1043	1128	2097	4892
Control	624	1043	1128	2097	4892
Total	1248	2086	2256	4194	9784
Missing assessment data (pre and/or post)					
Treatment	100	183	140	333	756
Control	153	260	220	540	1173
Total	253	443	360	873	1929
Missing matched pair with assessment data ^a					
Treatment	93	134	133	332	692
Control	40	57	53	125	275
Total	133	191	186	457	967
Students included in impact analysis					
Treatment	431	726	855	1432	3444
Control	431	726	855	1432	3444
Total	862	1452	1710	2864	6888

^aThe number of unmatched students with complete assessment data

Table A3. Analytic Sample Characteristics by Subgroup

Variable	Randomized Sample	Analytic Sample	Percent Attrition
English Language Learners			
Treatment	898	664	26.1%
Control	861	639	25.8%
All	1759	1303	25.9%
Students in Rural Schools			
Treatment	1720	1367	20.5%
Control	1720	1367	20.5%
All	3440	2734	20.5%
All students			
Treatment	4892	3444	29.6%
Control	4892	3444	29.6%
All	9784	6888	29.6%

Note. Attrition for ITBS Total, Reading Words, and Comprehension Impact Analyses

Table A4. Raw Means and Standard Deviations for ITBS Scores by Year and Treatment Group

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
	Treatment	Control								
N	431	431	726	726	855	855	1432	1432	3444	3444
TRL Pre-Test Scale Scores										
Mean	0.89	0.88	0.96	0.94	1.04	1.02	1.13	1.12	1.04	1.02
(Standard Deviation)	(1.15)	(1.13)	(1.16)	(1.13)	(1.16)	(1.15)	(1.59)	(1.44)	(1.36)	(1.28)
ITBS Total Scale Scores										
Mean	139.1	135.3	138.6	135.5	138.6	135.5	139	135.3	138.8	135.4
(Standard Deviation)	(7.6)	(6.8)	(7.4)	(7.4)	(7.6)	(7.2)	(7.5)	(7.1)	(7.5)	(7.2)
ITBS Reading Words Scale Scores										
Mean	141.1	137.1	140.6	137.1	140.4	137.3	140.8	136.9	140.7	137.1
(Standard Deviation)	(9.1)	(8.2)	(9.0)	(8.5)	(9.1)	(8.3)	(8.9)	(8.1)	(9.0)	(8.2)
ITBS Comprehension Scale Scores										
Mean	139.9	135.8	139.7	136.3	139.8	136.1	140.2	135.9	140.0	136.0
(Standard Deviation)	(8.9)	(8.3)	(9.4)	(9.0)	(9.4)	(8.9)	(9.7)	(9.2)	(9.5)	(9.0)

Table A5. Raw Means and Standard Deviations for OS Scores by Year and Treatment Group

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
	Treatment	Control								
N	419	391	710	706	845	843	1394	1363	3368	3303
Total OS Pre-Test Scale Scores										
Mean	368.5	366.1	366.4	366.2	372.9	370.8	371.5	368.8	370.4	368.4
(Standard Deviation)	(33.5)	(35.7)	(34.9)	(37.1)	(37.0)	(37.5)	(40.1)	(40.0)	(37.6)	(38.3)
Total OS Post-Test Scale Scores										
Mean	495.7	454.1	495.7	449.9	497.8	454.4	496.4	449.6	496.5	451.4
(Standard Deviation)	(40.2)	(45.4)	(41.9)	(51.9)	(44.2)	(46.2)	(46.6)	(50.2)	(44.3)	(49.1)

Table A6. Raw Means and Standard Deviations for ITBS Scores by Year and Treatment Group for ELL Students

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
	Treatment	Control								
N	75	76	154	150	134	133	301	280	664	639
TRL Pre-Test Scale Scores										
Mean	0.6	0.6	0.7	0.7	0.9	0.9	0.8	0.9	0.7	0.8
(Standard Deviation)	(0.8)	(0.9)	(1.0)	(1.0)	(1.1)	(1.1)	(1.2)	(1.2)	(1.1)	(1.1)
ITBS Total Scale Scores										
Mean	137.7	134.7	136.8	133.6	138.6	134.5	138.0	133.7	137.8	134.0
(Standard Deviation)	(7.6)	(5.9)	(7.1)	(7.6)	(7.0)	(7.0)	(7.7)	(7.2)	(7.4)	(7.1)
ITBS Reading Words Scale Scores										
Mean	139.1	136.1	138.6	135.3	139.7	136.3	139.6	135.2	139.3	135.6
(Standard Deviation)	(8.1)	(7.2)	(8.3)	(8.9)	(8.3)	(8.0)	(9.0)	(7.9)	(8.6)	(8.1)
ITBS Comprehension Scale Scores										
Mean	139.1	135.7	137.7	134.8	140.2	135.3	139.3	134.9	139.1	135.0
(Standard Deviation)	(9.4)	(7.4)	(9.4)	(8.6)	(8.6)	(9.1)	(9.4)	(8.9)	(9.3)	(8.7)

Table A7. Raw Means and Standard Deviations for OS Scores by Year and Treatment Group for ELL Students

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
	Treatment	Control								
N	68	68	153	147	132	31	290	264	643	610
Total OS Pre-Test Scale Scores										
Mean	361.5	364.1	360.6	359.5	365.8	362.8	359.8	356.5	361.4	359.4
(Standard Deviation)	(33.6)	(36.0)	(32.7)	(38.0)	(36.4)	(39.3)	(39.9)	(43.0)	(36.9)	(40.3)
Total OS Post-Test Scale Scores										
Mean	491.0	444.0	494.1	445.3	496.5	448.9	492.0	432.8	493.3	440.5
(Standard Deviation)	(39.4)	(48.9)	(40.5)	(57.2)	(44.7)	(47.6)	(48.8)	(51.3)	(45.1)	(52.1)

Table A8. Raw Means and Standard Deviations for ITBS Scores by Year and Treatment Group for Students in Rural Schools

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
	Treatment	Control								
N	152	152	331	331	395	395	489	489	1367	1367
TRL Pre-Test Scale Scores										
Mean	1.0	1.0	1.1	1.0	1.1	1.2	1.3	1.3	1.2	1.1
(Standard Deviation)	(1.3)	(1.3)	(1.2)	(1.1)	(1.2)	(1.2)	(1.8)	(1.5)	(1.5)	(1.3)
ITBS Total Scale Scores										
Mean	139.4	134.7	136.8	133.6	138.6	134.5	138.0	133.7	137.8	134.0
(Standard Deviation)	(7.3)	(5.9)	(7.1)	(7.6)	(7.0)	(7.0)	(7.7)	(7.2)	(7.4)	(7.1)
ITBS Reading Words Scale Scores										
Mean	141.6	136.7	140.9	136.9	140.8	137.9	141.4	136.8	141.0	137.1
(Standard Deviation)	(9.01)	(8.4)	(9.0)	(8.8)	(9.2)	(8.5)	(9.9)	(9.7)	(9.3)	(9.0)
ITBS Comprehension Scale Scores										
Mean	140.0	136.4	139.5	135.8	139.4	137.3	140.0	136.6	139.7	136.6
(Standard Deviation)	(8.4)	(6.2)	(6.7)	(7.2)	(7.4)	(7.1)	(7.3)	(6.9)	(7.2)	(7.0)

Table A9. Raw Means and Standard Deviations for OS Scores by Year and Treatment Group for Students in Rural Schools

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
	Treatment	Control								
N	152	142	331	329	395	395	488	488	1366	1354
Total OS Pre-Test Scale Scores										
Mean	369.8	370.4	369.8	368.6	379.7	378.8	381.0	380.0	376.7	375.9
(Standard Deviation)	(32.8)	(36.4)	(34.8)	(33.7)	(34.8)	(36.0)	(37.7)	(35.5)	(36.0)	(35.6)
Total OS Post-Test Scale Scores										
Mean	495.9	462.7	497.2	452.5	502.5	463.7	504.40	463.9	501.1	460.9
(Standard Deviation)	(37.3)	(44.1)	(37.0)	(44.0)	(36.9)	(50.4)	(40.6)	(42.8)	(38.4)	(45.8)

Table A10. HLM Parameter Estimates for Impacts on ITBS and OS Scores

	ITBS Total Score		ITBS Reading Words		ITBS Comp		OS Total Score	
Fixed Effects								
Intercept (β_0)	133.71	(0.24)	135.16	(0.27)	134.23	(0.30)	157.76	(4.88)
Pretest (β_1)	1.30	(0.06)	1.47	(0.08)	1.39	(0.08)	0.79	(0.01)
Treatment Effect (β_2)	3.70	(0.25)	3.83	(0.30)	4.32	(0.32)	44.51	(1.48)
Year								
SY 2011-2012 (β_3)	0.44	(0.48)	0.61	(0.54)	0.38	(0.59)	6.49	(2.56)
SY 2012-2013 (β_4)	0.48	(0.40)	0.55	(0.45)	0.70	(0.49)	2.14	(2.10)
SY 2013-2014 (β_5)	0.32	(0.38)	0.44	(0.43)	0.24	(0.47)	3.15	(2.00)
SY 2014-2015	Ref.		Ref.		Ref.		Ref.	
Treatment Effect * Year								
SY 2011-2012 (β_6)	0.07	(0.53)	0.21	(0.63)	-0.31	(0.67)	-4.44	(3.09)
SY 2012-2013 (β_7)	-0.60	(0.44)	-0.39	(0.52)	-0.98	(0.56)	1.03	(2.54)
SY 2013-2014 (β_8)	-0.70	(0.42)	-0.79	(0.50)	-0.73	(0.53)	-2.78	(2.42)
SY 2014-2015	Ref.		Ref.		Ref.		Ref.	
Random Effects								
School Intercept (τ^2)	2.66	(0.67)	16.20	(1.40)	18.44	(1.73)	338.21	(30.5)
School Impact (ξ^2)	14.20	(1.12)	11.18	(1.89)	13.53	(2.22)	389.77	(44.9)
School Intercept/Impact (ρ)	10.01	(1.35)	-0.25	(0.08)	-0.14	(0.09)	-0.55	(0.05)
Matched Pair (ω^2)	-0.24	(0.07)	2.69	(1.02)	5.21	(1.19)	48.19	(19.8)
Treatment Residual (σ^2)	30.28	(1.10)	51.02	(1.78)	51.80	(1.92)	825.02	(31.2)
Control Residual (σ^2)	29.61	(1.09)	42.98	(1.60)	52.55	(1.93)	935.48	(34.1)

Note. Significant parameter estimates ($p < .05$) are marked in bold type. Year One coded 0. The analytic sample for the ITBS consists of 6,888 students in 1,122 schools. The analytic sample for the OS consists of 6,644 students in 1,122 schools. See Appendix above for the mathematical / symbolic form of the HLM model of program impacts and definitions for each model parameter.

Table A11. HLM Parameter Estimates for Impacts on ELL Students' ITBS and OS Scores

	ITBS Total Score	ITBS Reading Words	ITBS Comp	OS Total Score
Fixed Effects				
Intercept (β_0)	131.85 (0.46)	133.08 (0.52)	132.97 (0.58)	132.88 (10.17)
Pretest (β_1)	2.11 (0.17)	2.33 (0.20)	2.18 (0.22)	0.84 (0.03)
Treatment Effect (β_2)	4.67 (0.55)	4.93 (0.63)	4.73 (0.68)	56.65 (3.12)
Year				
SY 2011-2012 (β_3)	1.69 (0.93)	1.82 (1.05)	1.40 (1.17)	4.92 (5.42)
SY 2012-2013 (β_4)	0.51 (0.74)	0.78 (0.83)	0.56 (0.93)	9.65 (4.21)
SY 2013-2014 (β_5)	0.66 (0.76)	1.04 (0.86)	-0.05 (0.97)	11.82 (4.37)
SY 2014-2015	Ref.	Ref.	Ref.	Ref.
Treatment Effect * Year				
SY 2011-2012 (β_6)	-1.61 (1.21)	-1.72 (1.39)	-1.39 (1.50)	-7.93 (6.98)
SY 2012-2013 (β_7)	-1.30 (0.94)	-1.37 (1.07)	-1.64 (1.16)	-8.03 (5.27)
SY 2013-2014 (β_8)	-0.47 (0.98)	-1.31 (1.12)	0.49 (1.22)	-11.70 (5.51)
SY 2014-2015	Ref.	Ref.	Ref.	Ref.
Random Effects				
School Intercept (τ^2)	10.65 (2.45)	12.48 (3.34)	18.30 (4.05)	410.80 (83.35)
School Impact (ξ^2)	17.03 (3.93)	17.82 (5.30)	23.55 (6.19)	539.46 (135.36)
School Intercept/Impact (ρ)	-0.23 (0.15)	-0.05 (0.21)	-0.31 (0.15)	-0.52 (0.11)
Matched Pair (ω^2)	5.66 (1.95)	4.18 (2.71)	10.22 (3.37)	65.65 (56.94)
Treatment Residual (σ^2)	21.28 (2.58)	33.52 (3.79)	38.52 (4.52)	695.70 (81.88)
Control Residual (σ^2)	27.38 (2.87)	40.20 (4.17)	40.29 (4.55)	847.67 (92.24)

Note. Significant parameter estimates ($p < .05$) are marked in bold type. Year One coded 0. The analytic sample for the ITBS consists of 1,303 students in 482 schools. The analytic sample for the OS consists of 1,253 students in 478 schools. See Appendix above for the mathematical / symbolic form of the HLM model of program impacts and definitions for each model parameter.

Table A12. HLM Parameter Estimates for Impacts on Rural Students' ITBS and OS Scores

	ITBS Total Score		ITBS Reading Words		ITBS Comp		OS Total Score	
Fixed Effects								
Intercept (β_0)	135.18	(0.40)	136.86	(0.45)	135.24	(0.51)	207.45	(7.87)
Pretest (β_1)	1.07	(0.10)	1.17	(0.12)	1.18	(0.13)	0.67	(0.02)
Treatment Effect (β_2)	3.39	(0.43)	3.22	(0.52)	4.66	(0.52)	39.18	(2.37)
Year								
SY 2011-2012 (β_3)	0.31	(0.79)	0.41	(0.89)	0.39	(1.01)	4.64	(4.18)
SY 2012-2013 (β_4)	-0.36	(0.60)	-0.79	(0.67)	0.58	(0.76)	-3.80	(3.12)
SY 2013-2014 (β_5)	0.73	(0.57)	0.72	(0.64)	1.11	(0.73)	2.00	(2.98)
SY 2014-2015	Ref.		Ref.		Ref.		Ref.	
Treatment Effect * Year								
SY 2011-2012 (β_6)	-0.45	(0.88)	0.05	(1.08)	-1.34	(1.08)	-5.40	(4.97)
SY 2012-2013 (β_7)	0.17	(0.67)	0.73	(0.83)	-0.78	(0.82)	4.50	(3.74)
SY 2013-2014 (β_8)	-1.29	(0.64)	-1.22	(0.79)	-1.91	(0.79)	-3.07	(3.56)
SY 2014-2015	Ref.		Ref.		Ref.		Ref.	
Random Effects								
School Intercept (τ^2)	1.67	(1.07)	2.20	(1.66)	1.99	(1.91)	4.89	(28.51)
School Impact (ξ^2)	11.80	(1.59)	12.72	(2.02)	17.31	(2.68)	270.60	(43.80)
School Intercept/Impact (ρ)	8.32	(2.04)	11.25	(3.09)	6.33	(3.17)	303.17	(63.67)
Matched Pair (ω^2)	-0.21	(0.12)	-0.26	(0.13)	0.09	(0.24)	-0.56	(0.08)
Treatment Residual (σ^2)	30.97	(1.77)	51.56	(2.88)	55.01	(3.15)	708.79	(43.47)
Control Residual (σ^2)	31.29	(1.77)	46.63	(2.68)	58.42	(3.27)	995.66	(54.83)

Note. Significant parameter estimates ($p < .05$) are marked in bold type. Year One coded 0. The analytic sample for the ITBS consists of 2,734 students in 443 schools. The analytic sample for the OS consists of 2,720 students in 443 schools. See Appendix above for the mathematical / symbolic form of the HLM model of program impacts and definitions for each model parameter.

Table A13. Impact Estimates on ITBS Scale Scores and OS Total Scores by Year

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
ITBS Total Scale Scores										
Impact Estimate (s.e.)	3.77	(0.46)	3.11	(0.36)	3.00	(0.33)	3.70	(0.25)	3.41	(0.16)
Glass's Δ	0.55		0.42		0.42		0.52		0.48	
Cohen's <i>d</i>	0.41		0.34		0.33		0.41		0.37	
ITBS RW Scale Scores										
Impact Estimate (s.e.)	4.04	(0.55)	3.43	(0.42)	3.04	(0.39)	3.83	(0.30)	3.57	(0.20)
Glass's Δ	0.48		0.37		0.41		0.47		0.43	
Cohen's <i>d</i>	0.39		0.33		0.35		0.42		0.35	
ITBS Comp Scale Scores										
Impact Estimate (s.e.)	4.01	(0.59)	3.34	(0.45)	3.60	(0.42)	4.32	(0.32)	3.90	(0.21)
Glass's Δ	0.49		0.41		0.37		0.47		0.43	
Cohen's <i>d</i>	0.40		0.34		0.30		0.38		0.38	
OS Total Scores										
Impact Estimate (s.e.)	40.07	(2.72)	45.54	(2.07)	41.73	(1.91)	44.51	(1.48)	43.49	(0.95)
Glass's Δ	0.88		0.88		0.90		0.89		0.89	
Cohen's <i>d</i>	0.91		1.04		0.95		1.01		0.99	

Note. Glass's Δ calculated using year specific year control group SD (see Table A4). The raw student-level correlations between pretest and posttest scores were .31, .25, .29, and .60 for ITBS Total, Comprehension, ReadWords, and OS Total Score respectively.

Table A14. Impact Estimates on ELL Students' ITBS and OS Scores by Year

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
ITBS Total Scale Scores										
Impact Estimate (s.e.)	3.06	(1.07)	3.37	(0.76)	4.20	(0.81)	4.67	(0.55)	4.08	(0.37)
Glass's Δ	0.52		0.44		0.60		0.65		0.57	
Cohen's <i>d</i>	0.34		0.37		0.46		0.51		0.45	
ITBS RW Scale Scores										
Impact Estimate (s.e.)	3.21	(1.24)	3.56	(0.87)	3.62	(0.93)	4.93	(0.63)	4.14	(0.42)
Glass's Δ	0.45		0.40		0.45		0.63		0.51	
Cohen's <i>d</i>	0.31		0.35		0.35		0.48		0.41	

APPENDIX A

ITBS Comp Scale Scores										
Impact Estimate (s.e.)	3.34	(1.34)	3.09	(0.94)	5.22	(1.01)	4.73	(0.68)	4.29	(0.46)
Glass's Δ	0.45		0.36		0.58		0.53		0.49	
Cohen's <i>d</i>	0.33		0.30		0.51		0.46		0.42	
OS Total Scores										
Impact Estimate (s.e.)	48.71	(6.24)	48.62	(4.25)	44.95	(4.54)	56.65	(3.12)	51.39	(2.08)
Glass's Δ	1.00		0.85		0.95		1.10		0.99	
Cohen's <i>d</i>	1.11		1.11		1.02		1.29		1.17	

Note. Glass's Δ calculated using year specific year control group SD.

Table A15. Impact Estimates on Rural Students' ITBS and OS Scores by Year

	2011-2012		2012-2013		2013-2014		2014-2015		Pooled	
ITBS Total Scale Scores										
Impact Estimate (s.e.)	2.93	(0.77)	3.56	(0.52)	2.09	(0.48)	3.39	(0.43)	3.00	(0.26)
Glass's Δ	0.47		0.50		0.30		0.49		0.43	
Cohen's <i>d</i>	0.32		0.39		0.23		0.37		0.33	
ITBS RW Scale Scores										
Impact Estimate (s.e.)	3.27	(0.95)	3.95	(0.64)	2.00	(0.59)	3.22	(0.52)	3.05	(0.31)
Glass's Δ	0.43		0.52		0.23		0.40		0.38	
Cohen's <i>d</i>	0.32		0.39		0.20		0.32		0.30	
ITBS Comp Scale Scores										
Impact Estimate (s.e.)	3.33	(0.94)	3.88	(0.64)	2.75	(0.59)	4.66	(0.52)	3.77	(0.31)
Glass's Δ	0.39		0.44		0.33		0.48		0.42	
Cohen's <i>d</i>	0.33		0.38		0.27		0.46		0.37	
OS Total Scores										
Impact Estimate (s.e.)	33.78	(4.36)	43.68	(2.88)	36.10	(2.66)	39.18	(2.37)	38.78	(1.43)
Glass's Δ	0.77		0.99		0.72		0.92		0.85	
Cohen's <i>d</i>	0.77		0.99		0.82		0.89		0.88	

Note. Glass's Δ calculated using year specific year control group SD.

Equation B1. Statistical Model for the Regression Discontinuity Study

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

$$Y_{ik} = \beta_0 + \beta_1(\text{Pretest}_{ik} - \text{Cutscore}_k) + \beta_2(\text{Trt}_{ik}) \\ + \alpha_k + \lambda_k(\text{Pretest}_{ik} - \text{Cutscore}_k) + \varphi_k(\text{Trt}_{ik}) + \varepsilon_{ijk}$$

$$\begin{pmatrix} \alpha_k \\ \lambda_k \\ \varphi_k \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2 & \rho_{\alpha,\lambda} & \rho_{\alpha,\varphi} \\ & \psi^2 & \rho_{\lambda,\varphi} \\ & & \xi^2 \end{pmatrix} \right), \text{ and } \varepsilon_{ijk} \sim N(0, \sigma^2)$$

Where:

- Y_{ik} is the posttest outcome score for student i from school k
- β_0 is the model intercept
- β_1 is the overall slope coefficient for the Pretest_{ik} covariate (i.e., the assignment variable: OS composite scores), which is centered around the school-specific RD cutscore (i.e., Cutscore_k)
- β_2 is the overall treatment effect of Reading Recovery
- Trt is a dummy variable which represents assignment based on the RD cutscore ($\text{Trt} = 1$ for students with pretest scores below the cutscore, and 0 for those above)
- α_k is the random intercept associated with school k , with variance τ^2
- λ_k is the random pretest slope associated with school k , with variance ψ^2
- φ_k is the random treatment effect associated with school k , with variance ξ^2
- $\rho_{\alpha,\lambda}$ is the correlation between random school intercepts and slopes
- $\rho_{\alpha,\varphi}$ is the correlation between random school intercepts and treatment effects
- $\rho_{\lambda,\varphi}$ is the correlation between random school slopes and treatment effects
- ε_{ijk} is the student-level residual, with variance σ^2

Table C1. Implementation Fidelity Activities

Ref	Standard	Survey	Population
Background and selection			
3.01	[RRT] Be employed in a school system that has a commitment to implementation.	RRT	All
3.02	[RRT] Hold teacher certification.	RRT	All
3.03	[RRT] Show evidence of successful teaching experience.	RRT	All
4.01	[TL] Hold a master's degree and teacher certification.	TL	All
4.02	[TL] Show evidence of successful teaching experience.	TL	All
4.04	[TL] Be nominated by an administrative agency making a RR application or by an established site.	TL	All
4.05	Have a commitment, following the completion of training, to remain employed as a RR TL with the agency.	TL	All
TL and Site capacity			
2.11	Maintain a training facility with a one-way glass and sound system for behind-the-glass-lessons	TL	All
2.14	Assure that a site report is developed annually and a copy is submitted to the university training center	TL	All
4.11	Participate in weekly class sessions and seminars (Reading Recovery teaching, leadership, theory)	TL	In training
4.13	Teach a child behind the glass a minimum of three times during the training year.	TL	In training
4.14	Teach four RR children per day individually for 30-minute sessions in a school setting throughout the SY.	TL	In training
4.16	Keep complete records on each child as a basis for instruction.	TL	In training
4.17	Receive at least four school visits from a university trainer.	TL	In training
4.18	Administer Observation Survey as appropriate TL& SC throughout the year.	TL	In training
4.24	Conduct colleague visits to other teacher leaders-in-training.	TL	In training
4.25	Make school visits to Reading Recovery teachers, initially with a teacher leader and then independently	TL	In training
4.36	Participate in an annual RRCNA-approved Reading Recovery conference.	TL	In training
4.37	Participate in the annual Teacher Leader Institute.	TL	In training
4.39	Teach four Reading Recovery children daily during the field year.	TL	Field year
4.40	Continue to teach a minimum of two children daily.	TL	Trained
4.49	Prepare an annual site report and submit it to the affiliated university training center	TL	Trained

Ref	Standard	Survey	Population
4.55	Monitor the selection and progress of children with reference to the teachers' records.	TL	Trained
4.59	Serve as a specialist to support Reading Recovery teachers in their work with challenging children.	TL	Trained
4.65	Participate in professional development opportunities for RR TLs sponsored by the university training center	TL	Trained
4.66	Receive a minimum of two site visits from a trainer during the teacher leader's first year(s) in the field.	TL	Field year
4.67	Participate in the annual Teacher Leader Institute.	TL	Trained
4.68	Participate in an approved RRCNA Reading Recovery conference each year.	TL	Trained
Training and PD			
2.19	[TL] Provide continuing professional development and support services for trained teachers.	TL	Trained
3.11	Participate in assessment training sessions.	RRT	In training
3.13	Teach a child behind the glass at least three times during the training year.	RRT	In training
3.15	Receive at least four school visits from the teacher leader over the course of the training year.	RRT	In training
3.43	Consult with the teacher leader about children not making satisfactory progress and other issues.	RRT	Trained
3.44	Participate in a min 6 PD sessions each year, including a min 4 behind-the-glass sessions each with 2 lessons.	RRT	Trained
3.45	Receive at least one school visit from a teacher leader annually.	RRT	Trained
4.42	Teach a training class of 8–12 Reading Recovery teachers during the field year.	TL*	Field year
4.43	Teach a training class of 8–12 Reading Recovery teachers in subsequent years as needed by the site.	TL*	Trained
4.44	Conduct a minimum of 24 hours assessment training (including practice with children).	TL	Trained
4.46	Ensure that teachers teach behind the glass at least three times during the year.	TL	Trained
4.47	Ensure that 80% of classes in the teacher training course include two behind-the-glass lessons.	TL*	Trained
4.48	Visit teachers-in-training four to six times during the year to provide guidance and instructional assistance.	TL	Trained
4.50	Provide a min 6 PD sessions each year, including a min 4 behind-the-glass sessions each with 2 lessons.	TL	Trained

Ref	Standard	Survey	Population
4.51	Visit trained RR teachers at least once each year to ensure quality control of the program.	TL	Trained
RR Lessons			
2.05	Select the lowest achieving children for service first (based upon Observation Survey) in all decisions.	RRT*	All
2.06	Administer Observation Survey and begin service to children within 2 weeks of school opening.	RRT	All
2.17	Assure consistent, daily, 30-minute individual lessons for all RR children on all days school is in session.	RRT	All
3.17	Teach at least 4 first-grade children per day for 30-min individually in a school setting throughout the SY.	RRT	In training
3.25	Submit data to the teacher leader as required.	RRT	In training
3.29	Teach at least 4 first-grade children per day for 30-min individually in a school setting throughout the SY.	RRT	Trained
3.32	Serve a minimum of eight children per year.	RRT	Trained
3.37	Submit data to the teacher leader as required.	RRT	Trained

Table C2. Percent of Standards Met by Applicable Respondent Group

Ref	Year 1	Y2	Y3	Y4	Pooled	Standard met all years
3.01	88	78*	70*	74*	77*	NO
3.02	100	100	100	100	100	YES
3.03	96	96	96	96	96	YES
4.01	100	100	100	100	100	YES
4.02	100	100	100	100	100	YES
4.04	94	100	100	100	97	YES
4.05	90	85	87	87	87	YES
2.11	94	96	98	98	97	YES
2.14	94	82	94	94	91	YES
4.11	100	100	100	100	100	YES
4.13	100	83	83	83	86	YES
4.14	100	83	83	83	86	YES
4.16	100	100	100	100	100	YES
4.17	100	83	83	83	86	YES
4.18	100	100	100	100	100	YES
4.24	100	83	83	83	86	YES
4.25	100	100	100	100	100	YES
4.36	100	100	100	100	100	YES
4.37	100	100	100	100	100	YES
4.39	100	84	82	80	83	YES
4.40	80	96	86	88	88	YES
4.49	82	84	100	100	91	YES
4.55	94	80	96	98	92	YES
4.59	98	96	100	100	99	YES
4.65	98	100	100	98	99	YES
4.66	89	78*	81	81	81	NO
4.67	98	98	98	98	98	YES
4.68	98	98	98	98	98	YES
2.19	100	98	98	100	99	YES
3.11	94	90	96	98	93	YES
3.13	86	90	88	86	88	YES
3.15	100	80	86	80	88	YES
3.43	94	94	94	94	94	YES
3.44	76*	76*	68*	68*	71*	NO
3.45	94	94	96	98	96	YES

Ref	Year 1	Y2	Y3	Y4	Pooled	Standard met all years
4.42	N/O	100	100	100	100	YES
4.43	N/O	98	98	98	98	YES
4.44	94	80	84	84	86	YES
4.46	100	98	96	94	96	YES
4.47	N/O	86	88	82	86	YES
4.48	98	94	84	98	94	YES
4.50	80	88	86	86	85	YES
4.51	88	82	100	100	93	YES
2.05	N/O	78*	84	84	81	NO
2.06	N/O	80	74*	76*	74*	NO*
2.17	92	98	82	84	90	YES
3.17	90	98	100	100	96	YES
3.25	100	100	100	100	100	YES
3.29	94	100	82	82	88	YES
3.32	84	94	100	100	97	YES
3.37	100	98	98	100	99	YES

*Did not meet benchmark for adequate fidelity. Note: N/O Not Observed

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95-135.
- Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2014). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation*, 1098214014557009.
- Allington, R. (2005). How much evidence is enough evidence? *The Journal of Reading Recovery*, 4(2), 8-11.
- Allison, P.D. (2012). *Logistic Regression Using SAS: Theory and Application, Second Edition*. Cary, NC: SAS Institute Inc.
- 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.
- Askew, B. J., Fountas, I. C., Lyons, C. A., Pinnell, G. S., & Schmitt, M. C. (1998). *Reading Recovery review: Understandings, outcomes and implications*. Columbus, OH: Reading Recovery Council of North America.
- Baker, E. L. (2006). Principles for scaling up: Choosing, measuring effects, and promoting widespread use of educational innovation. *Scale-up in Education*, 1, 37-54.
- Ball, D. L., Ben-Peretz, M., & Cohen, R. B. (2014). Records of practice and the development of collective professional knowledge. *British Journal of Educational Studies*, 62(3), 317-335.
- Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, 60(5), 497-511.
- Benard, B. (1995). *Fostering resilience in children*. ERIC Digest. Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.
- Berliner, D. C. (1994). Expertise: The wonder of exemplary performances. *Creating powerful thinking in teachers and students*, 161-186.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463-482.
- Berliner, B. A. (2004). Reaching unmotivated students. *Education Digest: Essential Readings Condensed for Quick Review*, 69(5), 46-47.
- Berman, P., & McLaughlin, M. W. (1977). Federal programs supporting educational

REFERENCES

- change. Volume III, Volume IV: Factors affecting implementation and continuation. Santa Monica, CA: Rand.
- Berman, P., & McLaughlin, M. W. (1976, March). Implementation of educational innovation. *The Educational Forum*, 40(3), 345-370. Taylor & Francis Group.
- Blom, G. (1958). *Statistical estimates and transformed beta variables*. New York: John Wiley & Sons, Inc.
- Bloom, H. (2009). *Modern regression discontinuity analysis*. MDRC Working Papers on Research Methodology. New York: MDRC.
- Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. W., & Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist*, 35(3), 149-164.
- Bodilly, S., Keltner, B., Purnell, S., Reichardt, R., & Schuyler, G. (1998). Lessons from New American Schools' scale-up phase. *Santa Monica, CA: Rand*, 79-84.
- Borman, G. D., & D'Agostino, J. V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*, 18(4), 309-326.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An experimental evaluation of Reading Recovery. *Reading Research Quarterly*, 30, 240-263.
- Century, J., Cassata, A., Rudnick, M., & Freeman, C. (2012). Measuring enactment of innovations and the factors that affect implementation and sustainability: Moving toward common language and shared conceptual understanding. *The Journal of Behavioral Health Services & Research*, 39(4), 343-361.
- Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, 31(2), 199-218.
- Chaiklin, S., & Lave, J. (Eds.). (1993). *Understanding practice: Perspectives on activity and context* (Vol. 78). Cambridge, UK: Cambridge University Press.
- Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed. Pp255-296). New York: Macmillan.

- Clay, M.M. (1987). Implementing Reading Recovery: Systemic adaptations to an educational innovation. *New Zealand Journal of Educational Studies*, 22(1), 35-58.
- Clay, M. M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.
- Clay, M. M. (1994a). *Reading Recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.
- Clay, M.M. (1994b). Reading Recovery: The wider implications of an educational innovation. *Literacy, Teaching, and Learning*, 1(1), 121-141.
- Clay, M. M. (1998). *By different paths to common outcomes*. Portland, ME: Stenhouse Publishers.
- Clay, M. M. (2005a). *Literacy Lessons Designed for Individuals: Why? when? and how?*. Portsmouth, NH: Heinemann.
- Clay, M.M. (2005b). *An Observation Survey of Early Literacy Achievement*, 2nd Edition. Portsmouth, NH: Heinemann.
- Clay, M. M. (2009). Implementing reading recovery internationally. In B. Watson and B. Askew (Eds.), *Boundless horizons: Marie Clay's search for the possible in children's literacy*. Rosedale, NZ: Pearson Education.
- Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies persistence of effects in the third year. *American Educational Research Journal*, 50(4), 812-850.
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3-12.
- Creswell, J. W. (2013). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Creswell, J. W., & Clark, L. L. P. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.
- D'Agostino, J., & Harmey, S. (2016). An international meta-analysis of Reading Recovery. *Journal of Education for Students Placed at Risk*, 21(1): 29-46.
- 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.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23-45.

REFERENCES

- Darling-Hammond, L., Hammerness, K., Grossman, P., Rust, F., & Shulman, L. (2005). The design of teacher education programs. In L. Darling-Hammond and J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 390-441). Jossey-Bass: San Francisco.
- Darling-Hammond, L., & Sykes, G. (1999). *Teaching as the learning profession: Handbook of policy and practice*. Jossey-Bass Education Series. San Francisco, CA: Jossey-Bass Inc.
- Datnow, A., Hubbard, L., & Mehan, H. (1998). *Educational reform implementation: A co-constructed process. Research Report 5*. Washington, DC: Office of Educational Research and Improvement.
- Day, C., Sammons, P., Hopkins, D., Leithwood, K., & Kington, A. (2008). Research into the impact of school leadership on pupil outcomes: Policy and research contexts. *School Leadership and Management*, 28(1), 5-25.
- Dede, C. (2006). Evolving innovations beyond ideal settings to challenging contexts of practice. In R.K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*, (p.551-566). Cambridge, UK: Cambridge University Press.
- Dede, C., Honan, J. P., & Peters, L. C. (2005). *Scaling Up Success: Lessons Learned from Technology-Based Educational Improvement*. Jossey-Bass: Indianapolis, IN.
- DeFord, D., Lyons, C., & Pinnell, G.S. (1991). *Bridges to literacy*. Portsmouth, NH: Heinemann.
- 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.
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York, NY: Macmillan.
- Dhillon, S., Darrow, C., & Meyers, C. V. (2015). Introduction to implementation fidelity. In C.V. Meyers and W.C. Brandt (Eds.) *Implementation fidelity in education research: Designer and evaluator considerations* (pp. 8-22). Routledge: New York.
- Dong, N. and Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness*, 6(1), 24-67.
- Dorn, L., & Schubert, B. (2008). A comprehensive intervention model for preventing reading failure: A response to intervention process. *The Journal of Reading Recovery*, 7(2), 29-41.
- Doyle, M. A., Gibson, S. A., Gómez-Bellengé, F., Kelly, P. R., & Tang, M. (2008). Assessment and identification of first-grade students at risk: Correlating the

- dynamic indicators of basic early literacy skills and an observation survey of early literacy achievement. In *57th Yearbook of the National Reading Conference* (p. 144). Austin, TX.
- Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral vision expertise in real world contexts. *Organization Studies*, *26*(5), 779-792.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087-1101.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237-256.
- Dweck, C. (2006). *Mindset: The new psychology of success*. New York, NY: Ballantine Books.
- Elmore, R. (1996). Getting to scale with good educational practice. *Harvard educational review*, *66*(1), 1-27.
- Elmore, R. F. (1978). Organizational models of social program implementation. *Public Policy*, *26*(2), 185-228.
- Eisenhardt, K. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532-550.
- Emerson, R.M., Fretz, R.I., & Shaw, L.L. (2011). *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.
- Estice, R. M. (1997). Using patterns of responding to “follow the child”. *Running Record*, *9*(2), 1, 6-7.
- Ferguson, R. F. (2003). Teachers’ perceptions and expectations and the Black-White test score gap. *Urban Education*, *38*(4), 460-507.
- Fenstermacher, G., & Richardson, V. (2005). On making determinations of quality in teaching. *The Teachers College Record*, *107*(1), 186-213.
- Flynn, J. M., & Rahbar, M. H. (1994). Prevalence of reading failure in boys compared with girls. *Psychology in the Schools*, *31*, 66-71.
- Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. Levittown, PA: Falmer Press, Taylor & Francis.
- Gerstner, J. J., & Finney, S. J. (2013). Measuring the implementation fidelity of student affairs programs: A critical component of the outcomes assessment cycle. *Journal of Research & Practice in Assessment*, *8*, 15-29.

REFERENCES

- Glynn, T., Bethune, N., Crooks, T., Ballard, K., & Smith, J. (1992). Reading Recovery in context: Implementation and outcome. *Educational Psychology, 12*(3-4), 249-261.
- Gobet, F., & Chassy, P. (2009). Expertise and intuition: A tale of three theories. *Minds and Machines, 19*(2), 151-180.
- Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics, 89*(1), 134-150.
- Goldring, E., Porter, A., Murphy, J., Elliott, S. N., & Cravens, X. (2009). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Leadership and Policy in Schools, 8*(1), 1-36.
- Goldring, E. B., & Pasternack, R. (1994). Principals' coordinating strategies and school effectiveness. *School Effectiveness and School Improvement, 5*(3), 239-253.
- Gómez-Bellengé, F., Gibson, S., Tang, M., Doyle, M., & Kelly, P. (2007). *Assessment and identification of first-grade students at risk: Correlating the dynamic indicator of basic early literacy skills and an Observation Survey of early literacy achievement*. Retrieved from <https://www.idecweb.us/>
- Gómez-Bellengé, F., Rodgers, E., Wang, C., & Schulz, M. (2005). *Examination of the validity of the Observation Survey with a comparison to ITBS*. Retrieved from <https://www.idecweb.us/>
- Good, T. L. (1987). Two decades of research on teacher expectations: Findings and future directions. *Journal of Teacher Education, 38*(4), 32-47.
- Goodson, B., Price, C., & Darrow, C. (2015). Measuring Fidelity: The present and future. In C.V. Meyers and W.C. Brandt (Eds.) *Implementation fidelity in education research: Designer and evaluator considerations*, 8-22. Routledge: New York.
- Graven, M., & Lerman, S. (2003). Review of Wenger, E.(1998). Communities of practice: Learning, meaning and identity. *Journal of Mathematics Teacher Education, 6*(2), 185-194.
- Griffiths, M., & Tann, S. (1992). Using reflective practice to link personal and public theories. *British Journal of Teacher Education, 18*(1), 69-84.
- Greene, J.C. (1998). Qualitative program evaluation: Practice and promise. In N.K. Denzin and Y.S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 372-399). Thousand Oaks, CA: Sage Publications.
- Hall, B., & Howard, K. (2008). A synergistic approach conducting mixed methods research with typological and systemic design considerations. *Journal of Mixed Methods Research, 2*(3), 248-269.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education,*

33(3), 329-352.

- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, 9(2), 157-191.
- Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, 32(1), 5-44.
- Hallinger, P., & Heck, R. H. (2010). Leadership for learning: Does collaborative leadership make a difference in school improvement? *Educational Management Administration & Leadership*, 38(6), 654-678.
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (No. w11154). Cambridge, MA: National Bureau of Economic Research.
- Hassel, B. C., & Steiner, L. (2000). *Strategies for scale: Learning from two educational innovations*. Innovations in American Government, Taubman Center for State and Local Government, John F. Kennedy School of Government. Harvard University.
- Honig, M. (2006). Complexity and policy implementation: Challenges and opportunities for the field. In M. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 1-25). Albany, NY: State University of New York Press.
- 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.
- 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.
- Huang, F. L., & Moon, T. R. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing schools. *Educational Assessment, Evaluation and Accountability*, 21(3), 209-234.
- Huberman, A. M. & Miles, M.B. (1998). Data management and analysis methods. In N.K. Denzin and Y.S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp.179-210). Thousand Oaks, CA: Sage Publications.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2(1), 88-110.
- Ingvarson, L., & Rowe, K. (2007, February). *Conceptualising and evaluating teacher quality: Substantive and methodological issues*. Paper presented at Economics of Teacher Quality Conference, Australian National University, Canberra (ACT)

REFERENCES

Retrieved from http://works.bepress.com/lawrence_ingvarson1/5

- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. New York, NY: Pearson Education.
- Korthagen, F. A., & Wubbels, T. (1995). Characteristics of reflective practitioners: Towards an operationalization of the concept of reflection. *Teachers and Teaching: Theory and practice*, 1(1), 51-72.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology* 105(3), 805.
- Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. *The Elementary School Journal*, 109(5), 491-509.
- Laurentson, , M., Oh, Y.J., & LaBlanca, F. (2015). STEM21 Digital Academy fidelity of implementations: Valuation and assessment of program components and implementation. In C.V. Meyers and W.C. Brandt (Eds.) *Implementation fidelity in education research: Designer and evaluator considerations*, 8-22. Routledge: New York.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a conception of how school leadership influences student learning. *Educational Administration Quarterly*, 46(5), 671-706.
- Lin, A. C. (2002). *Reform in the making: The implementation of social policy in prison*. Princeton, NJ: Princeton University Press.
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., ... & Busick, M. D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. Washington DC: U.S. Department of Education, *National Center for Special Education Research*.
- Louden, W., Rohl, M., Barratt-Pugh, C., Brown, C., Cairney, T., Elderfield, J., ... & Rowe, K. (2005). Context for the study of effective literacy teaching practices in the early school years. *Australian Journal of Language and Literacy*, 28(3), 183-184,245-248.
- Lyons, C. Pinnell, G.S., & DeFord, D. (1993). *Partners in learning: Teachers and children in Reading Recovery*. New York: Teachers College Press.
- Mann, S., & Walsh, S. (2013). RP or 'RIP': A critical perspective on reflective practice. *Applied Linguistics Review*, 4(2), 291-315.
- May, H., Goldsworthy, H., Armijo, M., Gray, A. M., Sirinides, P. M., Blalock, T. J.,

- Anderson-Clark, H., Schiera, A.J., Blackman, H., Gillespie, J., & Sam, C. (2014). *Evaluation of the i3 scale-up of Reading Recovery: Year two report, 2012-13*. Philadelphia: Consortium for Policy Research in Education.
- May, H., Gray, A., Sirinides, P., Goldsworthy, H., Armijo, M., Sam, C., Gillespie, J.N., & Tognatta, N. (2015). *Year one results from the multisite randomized evaluation of the i3 scale-up of Reading Recovery*. *American Educational Research Journal*, 52(3), 547-581.
- May, H., Gray, A., Gillespie, J. N., Sirinides, P., Sam, C., Goldsworthy, H., Armijo, M. & Tognatta, N. (2013). *Evaluation of the i3 Scale-up of Reading Recovery: Year one report, 2011-12*. Philadelphia: Consortium for Policy Research in Education.
- McDonald, S. K., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-up exemplary interventions. *Educational Researcher*, 35(3), 15-24.
- McKown, C., & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology*, 32, 159-184.
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, 46(3), 235-261.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks, CA: Sage Publications.
- Millman, J. & Darling Hammond, L. (1990). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. London, UK: Sage Publications.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American journal of evaluation*, 24(3), 315-340.
- Mulford, B., & Silins, H. (2003). Leadership for organisational learning and improved student outcomes—What do we know? *Cambridge Journal of Education*, 33(2), 175-195.
- Mulford, B. & Silins, H. (2005). Developing leadership for organizational learning. In M. Coles & G. Southworth (Eds.). *Developing leadership: Creating the schools of tomorrow*. pp. 139-157. Maidenhead: McGraw-Hill and Open University Press. ISBN 0 335 21542 4
- Munn, P. & Ellis, S. (2005) Interactions between school systems and reading recovery programmes; Evidence from Northern Ireland. *The Curriculum Journal*, 16(3), 341-361.
- Naik, S. S., & Ball, D. L. (2014). Professional development in a laboratory setting

REFERENCES

- examining evolution in teachers' questioning and participation. *Journal of Mathematics Education*, 7(2), 40-54.
- National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, DC: National Institute of Child Health and Human Development, National Institutes of Health.
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, 39(4), 374-396.
- Nelson, B. C., Ketelhut, D. J., Clarke, J., Dieterle, E., Dede, C., & Erlandson, B. (2007). Robust Design Strategies for Scaling Educational Innovations. *The Design and Use of Simulation Computer Games in Education*, 209.
- North American Trainers Group (2015). *Rationales and guidelines for selecting the lowest-achieving first-grade students for reading recovery*. Worthington, OH: Reading Recovery Council of North America.
- Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes. *The Journal of Behavioral Health Services & Research*, 39(4), 417-433.
- Patton, M.Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Piaget, J. (1971). *Biology and knowledge*. Chicago, IL: University of Chicago Press. (First French ed., 1967).
- Pinnell, G. S. (1989). Reading Recovery: Helping at-risk children learn to read. *The Elementary School Journal*, 90, 161-183.
- Pinnell, G. S., Fried, M. D., & Estice, R. M. (1990). Reading Recovery: Learning how to make a difference. *The Reading Teacher*, 43, 282-295.
- 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.
- Pressman, J., & Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland*. Berkeley: University of California Press.
- Prestine, N. A., & Nelson, B. S. (2005). How can educational leaders support and promote teaching and learning? New conceptions of learning and leading in schools. In W.A. Firestone, and C. Riehl, C. (Eds.), *A new agenda for research in educational*

- leadership*, (pp. 46-60). New York, NY: Teachers College Press.
- 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.
- Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal*, 45(1), 206-230.
- Ravitch S.M., & Mittenfelner Carl, N.M. (2016). *Qualitative research: bridging the conceptual, theoretical, and methodological*. Thousand Oaks, CA: Sage Publications.
- Reading Recovery Council of North America. (2012). *Standards and Guidelines of Reading Recovery in the United States* (6th ed.). Columbus, OH: Author.
- Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.
- Richardson, J. T. (2005). Students' approaches to learning and teachers' approaches to teaching in higher education. *Educational Psychology*, 25(6), 673-680.
- Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal*, 28(3), 559-586.
- Richert, A. E. (1992). Writing cases: A vehicle for inquiry into the teaching process. *Case methods in teacher education*, 155-174.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 10(2), 417-458.
- Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, 48(2), 268-302.
- Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*. doi: 10.1177/0013161X08321509
- Rogoff, B., Turkkanis, C. G., & Bartlett, L. (2001). *Learning together: Children and adults in a school community*. New York, NY: Oxford Press.
- Rodgers, A., & Fried, M. (2009). A design for action: Analyzing problems of implementation. *Journal of Reading Recovery*, 8(2), 51-60.
- Rodgers, E. (2016). Scaling and sustaining an intervention: The case of Reading Recovery. *Journal of Education for Students Placed at Risk*, 21(1): 10-28.

REFERENCES

- Rodgers, E.M., D'Agostino, J.V., Harmey, S.J., Kelly, R.H. & Brownfield, K. (In Press). Examining the nature of scaffolding in an early literacy intervention. *Reading Research Quarterly*, 51, (3).
- 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.
- Sabatier, P., & Mazmanian, D. (1979). The conditions of effective implementation: A guide to accomplishing policy objectives. *Policy Analysis*, 5(4), 481-504.
- Schmitt, M. C., Askew, B. J., Fountas, I. C., Lyons, C. A., & Pinnell, G. S. (2005). *Changing futures: The influence of Reading Recovery in the United States.* Columbus, OH: Reading Recovery Council of North America.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action* (Vol. 5126). New York, NY: Basic Books.
- Schön, D. A. (1987). *Educating the reflective practitioner.* San Francisco, CA: Jossey-Bass Inc.
- Schön, D. A. (1991). *The reflective turn: Case studies in and on educational practice.* New York, NY: Teachers Press, Columbia University.
- Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J.R., Porter, J., Smith, J. (2010). *Standards for Regression Discontinuity Designs.* Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf
- Seashore Louis, K., Leithwood, K., Wahlstrom, K.L., & Anderson, S. E. (2010). *Learning from leadership: Investigating the links to improved student learning.* St. Paul, MN: University of Minnesota, Center for Applied Research and Educational Improvement.
- Shanahan, T. & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly*, 20(4), 958-995.
- Smith-Burke, T. (2010). Implementation: A necessary concept for success. *Journal of Reading Recovery*, Spring 2010, 63-65.
- Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, 18(2), 141-179.

- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*(3), 387-431.
- Stein, M. K., & Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis, 25*(4), 423-448.
- Styhre, A. (2011). *Knowledge sharing in professions: roles and identity in expert communities*. Aldershot, UK: Gower Publishing, Ltd.
- Summerfelt, W. T. (2003). Program strength and fidelity in evaluation. *Applied Developmental Science, 7*(2), 55-61.
- Sun, J., & van Es, E. A. (2015). An exploratory study of the influence that analyzing teaching has on preservice teachers' classroom practice. *Journal of Teacher Education, 0022487115574103*.
- Supovitz, J., Sirinides, P., & May, H. (2009). How principals and peers influence teaching and learning. *Educational Administration Quarterly, 46*(1), 31-56.
- Tang, M. & Gómez-Bellengé, F. X. (2007, April). *Dimensionality and concurrent validity of the observation survey of early literacy achievement*. Paper presented at the annual meeting of the American Educational Research Association in Chicago, Illinois.
- Teddlie, C., & Tashakkori, A. (Eds.). (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publications Inc.
- Thurmond, V. A. (2001). The point of triangulation. *Journal of Nursing Scholarship, 33*(3), 253-258.
- Tremmel, R. (1993). Zen and the art of reflective practice in teacher education. *Harvard Educational Review, 63*(4), 434-459.
- Trochim, William (1984). *Research Design for Program Evaluation: The Regression Discontinuity Approach*. Beverly Hills: Sage.
- Van Manen, M. (1991). *The tact of teaching: The meaning of pedagogical thoughtfulness*. Albany, NY: Suny Press.
- Vandevoort, L. G., & Berliner, D. C. (2004). National Board Certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46).
- Wahlstrom, K. L., Louis, K. S., Leithwood, K., & Anderson, S. E. (2010). Learning from leadership: Investigating the links to improved student learning. The Informed Educator Series. *Educational Research Service*.
- Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 Years of research tells us about the effect of leadership on student achievement. A Working*

REFERENCES

- Paper*. Aurora, CO: Mid-continent Research for Education and Learning.
- Weatherley, R. & Lipsky, M. (1977) Street-Level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Educational Review*, 47(2), 171-197.
- Weiss, M. J., Bloom, H. S., & Brock, T. (2013). A conceptual framework for studying the sources of variation in program effects. Available at SSRN 2393091. <http://ssrn.com/abstract=2393091>
- Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems thinker*, 9(5), 2-3.
- Wenger, E., Trayner, B., & de Laat, M. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework*. Retrieved from <http://wenger-trayner.com/resources/publications/evaluation-framework/>
- What Works Clearinghouse (2014). *What Works Clearinghouse procedures and standards handbook version 3.0*. Washington, DC: US Department of Education, Institute of Education Sciences. Retrieved from <http://whatworks.ed.gov>.
- Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 398-425.
- Yanow, D. (1996). *How does a policy mean?: Interpreting policy and organizational actions*. Washington, DC: Georgetown University Press.
- Yanow, D., & Tsoukas, H. (2009). What is reflection-in-action? A phenomenological account. *Journal of Management Studies*, 46(8), 1339-1364.
- Yin, R.K. (2012). Case study methods. In H. Cooper (Ed-in-chief), *APA handbook of research methods in Psychology, Volume 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 141-156). Washington, DC: American Psychological Association.
- Zeichner, K. M. (1994). Research on teacher thinking and different views of reflective practice in teaching and teacher education. In I. Carlgren, G. Handal, and S. Vaage, (Eds.), *Teachers' minds and actions: Research on teachers' thinking and practice* (pp. 9-27). Bristol, PA: The Falmer Press, Taylor & Francis, Inc.

