U.S. Department of Education
November 2015

Evaluation of Response to Intervention Practices for Elementary School Reading

Rekha Balu
Pei Zhu
Fred Doolittle
Ellen Schiller
Joseph Jenkins
Russell Gersten

ies INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

This page has been left blank for double-sided copying.

Evaluation of Response to Intervention Practices for Elementary School Reading

November 2015

Rekha Balu
Pei Zhu
Fred Doolittle
MDRC

Ellen Schiller
SRI International

Joseph Jenkins
University of Washington

Russell Gersten
Instructional Research Group

Jonathan Jacobson
Project Officer
Institute of Education Sciences

NCEE 2016-4000
U.S. DEPARTMENT OF EDUCATION

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
Ruth Curran Neild
Deputy Director for Policy and Research Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance
Joy Lesnick
Acting Commissioner

November 2015

This report was prepared for the Institute of Education Sciences under Contract ED-04-CO-0111/0003 with MDRC. The project officer is Jonathan Jacobson in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the report.

This report is available on the Institute of Education Sciences website at http://ies.ed.gov/ncee.

Alternate Formats: Upon request, this report is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

Contents

List of Exhibits

Table

Table

Table

Table

Figure

Figure

Figure

Box

Acknowledgments

This study represents a collaborative effort among the authors, our colleagues at MDRC and SRI International, and Institute of Education Sciences (IES) staff. This study also benefits from collaboration between the study team and more than 1,200 schools included in the evaluation, plus many other schools that participated in our initial screening of sites. We appreciate the willingness of school staff to provide information on how they deliver reading instruction to students, especially those who are not yet meeting reading standards. Staff in participating schools and districts completed surveys, provided student records on reading assessments and instruction, and supported the study team as it fielded reading assessments. Without their help, the study would not be possible.

We also benefited from the advice of the members of our Technical Working Group who provided expert insights into reading assessments, instruction, the response to intervention (RtI) approach, and research methodology. Members of this group over the course of the project have included Donald Compton of Vanderbilt University, Carol Connor of Arizona State University, Judy Elliot formerly of Los Angeles Unified School District and now an independent consultant, David Francis of the University of Houston, Paul McDermott of the University of Pennsylvania, Rollanda O'Connor of the University of California at Riverside, Sean Reardon of Stanford University, Amy Sichel of the Abington Pennsylvania School District, Jeffrey Smith of the University of Michigan, Deborah Speece of the Virginia Commonwealth University, and Sharon Vaughn of the University of Texas. We also wish to extend our thanks to reviewers of the report, including Jose Blackorby, Gordon Berlin, Howard Bloom, William Corrin, John Hutchins, Michelle Maier, Shira Mattera, and Marie-Andrée Somers.

Many people at four organizations played a crucial role in conducting this study. Many of the MDRC staff members involved in processing the data, conducting the analysis, and preparing this report are named below in the listing of project staff. We especially want to acknowledge the contributions of Ivonne Garcia throughout the project for her role in skillfully organizing, checking quality control, and merging complex data from multiple sources; and of Amanda Ferrandino for serving as our report coordinator and producing multiple versions of numerous exhibits and chapters for the report as we sought the best way to convey the findings. In addition, Margaret Bald, Alison Black, Stephanie Cowell, Sonia Drohojowska, Anne Fenton, Kate Gualtieri, Kathy Jenkins, Ellie Leahy, Nathan Meyers, Seth Muzzy, Ushapriya Narasimhan, Elizabeth Nelson, Gina Price, Shelley Rappaport, Carolyn Thomas, and Robert Weber played important roles in site recruitment, survey design, project management, and report production. At SRI, project staff including Jake Cornett, Sara Kalb, Camille Marder, Christina Saldarriaga, and Katherine Nagle played important roles in early stages of the project. At the Instructional Research Group, Joseph Dimino, Mary Jo Taylor, and Becky Newman-Gonchar were

important vital contributors to the project. At SRM, Linda Kuhn, Tony Lavender, Julie Berry, Maegan McKee, Daniel Mackin, and Betsy Quicksall played crucial roles in data collection.

Following are the Data Processing and Analysis Team for the Report and Restricted Use File (in alphabetical order):

- Catherine Armstrong
- Colin Bottles
- Nick Commins
- Catherine Corbin
- Albert Fang
- Ivonne Garcia
- Zeest Haider
- Kateryna Lashko
- Sophia Litschwartz
- Heilyn Paulino
- Nick Sherefkin
- Sarah Spell
- Mark van Dok
- Edmond Wong

The Authors

Disclosure of Potential Conflicts of Interest

The research team for this evaluation consists of a prime contractor, MDRC, Inc., of New York City, NY, and three subcontractors: SRI International, of Menlo Park, CA; Instructional Research Group, of Los Alamitos, CA; and Survey Research Management (SRM) Corporation, of Boulder, CO. None of these organizations or their key staff has financial interests that could be affected by findings from the evaluation of the two supplemental literacy interventions considered in this report. No one on the eight-member Technical Working Group, convened by the research team to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

This page has been left blank for double-sided copying.

Executive Summary

The 2004 reauthorization of the Individuals with Disabilities in Education Act (IDEA) allows states and school districts to use a portion of federal special education funds to provide coordinated early intervening services to students at risk of reading failure or other academic or behavioral problems. One of the primary approaches that has emerged is called "Response to Intervention" (RtI). In the context of this report, RtI incorporates a range of assessment, instruction, and intervention principles, including (1) offering multiple tiers of support for students, depending on the level of reading difficulty they may be experiencing; (2) allocating staff to provide that tiered support to students; and (3) collecting and using data to make instructional and intervention decisions for students throughout the school year.

This study describes these RtI practices and compares their prevalence between two different samples: a *reference sample* of schools representative of elementary schools in the 13 states included in the evaluation and an *impact sample* of 146 elementary schools with three or more years of implementing RtI approaches in reading. In the impact sample, the study research team compared the intensity of services provided to reading groups at different reading levels to measure the extent to which support is more intense for students reading below grade level. For the impact analysis, the study research team estimated effects of assignment to reading interventions for students at the margins of eligibility for those services who read below grade level.

This report provides new information on the prevalence of RtI practices in elementary schools, illustrates the implementation of RtI practices for groups of students at different reading levels, and provides evidence on effects of one key element of RtI: assigning students to receive reading intervention services. The findings show, for the 2011-12 school year, that:

- A majority of schools in the 13-state reference sample (56 percent) reported full implementation of the RtI framework, while a higher proportion of impact sample schools (86 percent) in those states reported full implementation.
- Schools in the impact sample adjusted reading services to provide more support to students reading below grade-level standards than to those at or above the standards.
- For those students just below the school-determined eligibility cut point in Grade 1, assignment to receive reading interventions did not improve reading outcomes; it produced negative impacts.

The rest of the Executive Summary describes the evaluation's policy context and specific research questions, defines key terms and analytic approaches, and explains the findings.

Policy Context and Relation to Previous Research

For school year 2008-09, when this study began its planning and design, 70 percent of districts with elementary schools reported using RtI in reading/language arts.[1] The use of the RtI framework is an outgrowth of a change in approach related to special education policy and the process for identifying children with a Specific Learning Disability (SLD) — the disability category most associated with reading difficulties. The previous eligibility standard required educators to document an "educationally significant discrepancy" between achievement of specific skills (for example, reading performance) and general ability (that is, overall intellectual functioning as measured by an IQ test) that could not be explained by visual, hearing, or motor disabilities; emotional disturbances; mental retardation; or environmental, cultural, or economic disadvantage. The 2004 reauthorization of IDEA forbids states from requiring districts to identify SLD students using a discrepancy approach, and it permits districts to use an SLD identification process based on the child's response to scientific, research-based interventions. The law also allows districts to use up to 15 percent of their IDEA Part B special education funds to develop and implement coordinated early intervening services for students not yet identified as needing special education and related services but who need additional academic or behavioral support to be successful in general education classrooms. This funding change allows federal dollars to be used for RtI services.

Over the past 15 years, numerous studies have addressed the effect of interventions delivered to early readers in need of help within an RtI framework. A survey of the recent literature (since 1999) yields 27 studies that report the impact of providing certain types of interventions to students with reading difficulties on a range of reading skill measures. These recent studies support the conclusion of Gersten et al. that well-designed and closely monitored small-group reading interventions could be beneficial to early-grade readers in terms of improving their specific reading skills.[2] The evidence is stronger for first grade than for second or third grades. The effect of such intervention on students' more comprehensive reading skills is less clear. Also not clear is the impact of such interventions if they were to be implemented at a larger scale.

---

[1]M. C. Bradley, Tamara Daley, Marjorie Levin, Fran O'Reilly, Amanda Parsad, Anne Robertson, and Alan Werner, *IDEA National Assessment Implementation Study,* NCEE 2011-4027 (Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2011).

[2]Russell M. Gersten, Donald L. Compton, Carol M. Connor, Joseph Dimino, Lana Santoro, Sylvia Linan-Thompson, and W. David Tilly, "Assisting Students Struggling with Reading: Response to Intervention and Multi-Tier Intervention in the Primary Grades" (Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2009). Website: http://ies.ed.gov/ncee/wwc/PracticeGuide.aspx?sid=3.

This evaluation's analysis of RtI implementation and the impact of interventions on reading achievement expands the field's knowledge about RtI in three ways. First, this study describes implementation of RtI practices in multiple states at the school level, unlike previous studies that address RtI adoption at the district or state level. Second, this study describes practices in schools that had adopted RtI on their own and had implemented it for three or more years, rather than for a sample of schools that were monitored by researchers or that received special supports for first-year implementation. Third, while this study's school sample is broader than in earlier studies, the student sample is narrower. Unlike earlier studies, which address the overall effectiveness of RtI, this study's research design answers a question about effective targeting, by comparing the outcomes for students just below and just above the cut point of eligibility for intervention. This approach provides an estimate of the impact of interventions on the students *slightly below* grade-level reading standards, rather than for the full range of students served by interventions. This impact on the marginally eligible student served is important for assessing the effective targeting of intervention resources, but it does *not* assess whether the RtI framework as a whole is effective in improving student outcomes or whether reading interventions are effective for students *well below* grade-level standards.

Research Questions and Study Overview

This study answers three sets of major research questions:

1. **Comparison of practices between school samples.** How did the prevalence of RtI practices differ between a representative "reference" sample of schools and schools selected for the impact evaluation? To what extent were impact sample schools implementing more RtI practices than the reference sample schools? How do special education identification rates in the impact sample compare with rates for the states as a whole?

2. **Comparison of reading services between reading groups at different skill levels.** In impact sample schools (those with three or more years of implementing RtI): To what extent did schools place students in tiers as suggested by earlier RtI models? To what extent did schools adjust tier placement during the school year? To what extent is there variation in how schools organize reading services for specific reading levels? To what extent were services for students reading below grade level more intense than for students reading at or above grade level?

3. **Impacts on reading outcomes of students.** For students who fell just below school-determined standards for each grade on screening tests: What were the effects on reading achievement of actual assignment to receive reading intervention services (in addition to core instruction)? What is the extent of variation in estimat-

ed impacts across RtI schools? How is the estimated impact associated with certain school features or student characteristics?

Key Terms, Sample Selection, and Research Design

"Intervention" in this report generally refers to additional support for students who have difficulty reading. RtI schools may place students in reading groups and deliver services based, in part, on students' scores on *screening tests,* which are brief assessments of skills considered necessary for reading, such as word identification and letter sounds. In this way, how students score on screening assessments is related to the services they receive. Screening tests differ from the end-of-year comprehensive reading tests, which evaluate a wider variety of reading skills.

- **Tier 1.** "Tier 1" refers to the core instruction that all students receive. The National Reading Panel has recommended that reading instruction in the early grades focus on five reading components: phonemic awareness, phonics, fluency, reading comprehension, and vocabulary.[3] Tier 1 is intended to prevent the risk of reading failure for as many students as possible and to avoid inappropriate referrals to special education. Core instruction usually occurs during a period called the "core reading block." Students who receive only core instruction generally read at or above grade level.

- **Tiers 2 and 3.** Students placed in Tier 2 or Tier 3 receive intervention services in addition to Tier 1 core instruction services. Students in Tier 2 generally read at least *somewhat below* grade level based on screening tests. The typical mechanism that schools use to deliver services to students in Tier 2 is an adult-led small reading group — an approach that could be used to provide small-group instruction during the core reading block as well as additional intervention services. Students in Tier 3 generally read *far below* grade level or have not responded to Tier 2 interventions, and they may be assigned to more intensive interventions (characterized by smaller group size, additional intervention time, or both). To address the second research question of how services differ depending on students' reading skills, the descriptive analysis compares services received by reading groups at different skill levels: at or above grade level, somewhat below

---

[3]National Reading Panel, "Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction" (Washington, DC: U.S. Department of Health and Human Services, National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development, 2000).
Website: http://www.nichd.nih.gov/publications/pubs/nrp/pages/smallbook.aspx?renderforprint=1.

grade level, or far below grade level. This analysis compares features of groups receiving small-group reading instruction during the core reading block as well as features of groups receiving reading intervention services.

Schools purposively selected for inclusion in the impact study reported at least three years' experience with RtI implementation and are referred to as the "impact sample." The impact sample was selected to include schools implementing all of the following practices no later than 2009-10:

- Use of three or more tiers of increasing instructional intensity to deliver reading services to students
- Fielding of screening assessments of all students (universal screening) at least twice a year
- Use of data for placing students in Tier 2 or Tier 3
- Use of progress monitoring (beyond universal screening) for students reading below grade level to determine whether intervention is working for students placed in Tier 2 or Tier 3

Schools in the impact sample provided information about the score on a screening test that they used to determine a student's placement in Tier 2 or Tier 3. This score, referred to as a "cut point" (or "cut score"), allowed the study research team to determine whether schools followed a consistent quantitative decision rule for tier placement.

To address the third research question, which assesses the relationship between assignment of students to Tier 2 or 3 to receive intervention services and their reading outcomes, the study uses a *Regression Discontinuity (RD) design.* This quasi-experimental research design provides a causal impact estimate when random assignment is not possible. Schools participating in the impact evaluation used students' fall screening test scores to determine their assignment to intervention. Students whose scores are below the predefined cut point typically receive treatment (Tier 2 or 3 intervention services) in addition to core instruction, and those whose scores are at or above the cut point typically receive only core instruction (Tier 1). Students at or near either side of the cut score are expected to be comparable to each other, and they form the treatment and comparison groups for the impact analysis. Most but not all of the students with scores just below the cut point were placed in Tier 2, while most of the students with scores just above the cut score were placed in Tier 1.

Samples and Data

Different samples and data were used to answer each of the three main research questions. To study different RtI practices across schools, the *impact sample* of 146 unique schools across 13 states was compared with a random sample of 100 elementary schools in each of the same 13 states (referred to as the *reference sample*),[4] based on data collected through a school administrator survey.

To compare reading services provided to reading groups at different skill levels, in particular for students reading below grade level (or students receiving Tier 2 or 3 intervention services) and students reading at or above grade level (or students in Tier 1 only), the study research team collected survey data in spring 2012 from reading teachers and staff who provided reading intervention services. The survey data report information about reading services provided to reading groups of all reading levels by group, not by individual student.

Finally, to analyze the impacts of assignment to intervention on students' reading achievement, the study research team compared the difference in reading outcomes between students whose fall screening test scores were just above the cut point for Tier 2 intervention set by the schools and those whose scores were just below, based on the RD design described above. This design determines that the impact findings are applicable not to everyone receiving either Tier 2 or Tier 3 intervention, but only to students whose fall screening scores were close to the cut point. Students close to the cut point are largely Tier 2 students but also include a small portion of Tier 3 students.

To carry out this design, the study research team collected individual-level fall screening test scores and resulting tier placements for fall and winter of the 2011-12 school year for all students in grades 1-3 in the 146 impact sample schools.[5] The reading achievement outcomes used in the impact analysis vary by grade. The study research team administered the Early Childhood Longitudinal Study, Kindergarten Cohort, of 2011 (ECLS-K: 2011) Reading Assessment to first graders in the sample to measure their comprehensive reading skills; it also administered a Sight Word Efficiency test (the Test of Sight Word Reading Efficiency, 2nd edition, or TOWRE2) to measure students' decoding fluency skill in Grades 1 and 2. For third-graders, individual-level scores from the spring state reading achievement tests were used to measure students' comprehensive reading skills.

[4] Of the 1,300 schools randomly sampled for the reference sample, 1,105 (or 85 percent) completed the school administrator survey that principals of impact sample schools also received.

[5] Note that number of schools eligible for the impact analysis varies by grade, with 119 eligible schools for Grade 1 analysis, 127 eligible schools for Grade 2 analysis, and 112 eligible schools for Grade 3 analysis.

Summary of Findings

This study reports on services and impacts in the 2011-12 school year — the only year for which data were collected and analyzed. This section reports key findings related to the three types of analysis presented in the report.

Comparison of Practices Between Schools

- **More than half of the reference sample schools in the 13 study states adopted an RtI framework in Grade 1-3 reading for the 2011-12 school year. A higher proportion of impact sample schools than reference sample schools reported full implementation of an RtI framework for Grade 1-3 reading.**

Figure ES.1 shows that a majority of schools in both samples reported full implementation of an RtI framework for reading: 86 percent of impact sample schools, compared with 56 percent of reference sample schools. Because the impact schools were screened for experience with RtI implementation, this difference is to be expected. The study research team also examined the frequency of specific practices that correspond to three key aspects of an RtI framework, described below.

Multiple Tiers of Reading Instruction and Intervention

Although about two-thirds (68 percent to 70 percent) of both school samples reported offering more than 90 minutes per day of core reading instruction, the frequency of offering intervention differed between the two samples. Impact sample schools were more likely to report providing time for Tier 2 intervention at least three times a week than were reference sample schools (97 percent and 80 percent, respectively). Impact sample schools were also more likely to report providing time for Tier 3 intervention at least five times a week than were reference sample schools (68 percent and 47 percent).

Allocation of Staff

Impact sample schools were more likely than reference sample schools to allocate staff to assist teachers with using data (88 percent and 72 percent, respectively) and with reading instruction (69 percent and 56 percent).

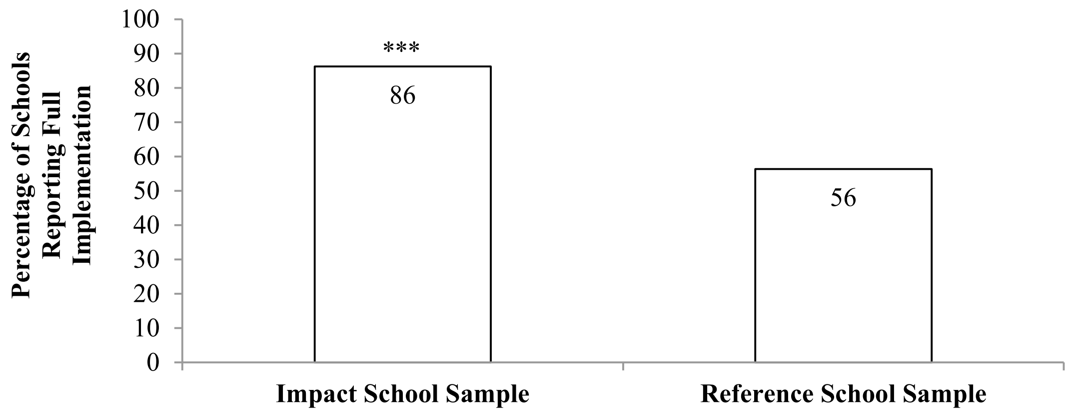
Use of Data to Inform Decisions

Among impact sample schools, 83 percent conducted universal screening assessments of students at least twice a year, compared with 59 percent of reference sample schools. Impact

The Response to Intervention (RtI) Evaluation

Figure ES.1

Full Implementation of RtI in Reading in Grades 1-3



SOURCE: School survey.

NOTES: The survey defined RtI as a "multistep approach to providing early and progressively intensive intervention and monitoring within the general education setting." Respondents could answer that RtI was "fully implemented," "partially implemented," or "not implemented" in reading for each grade. This exhibit reports the percentage of respondents reporting that RtI was "fully implemented" for each of Grades 1, 2, and 3 for which the school responded. Percentages reflect rounding. The statistical significance is indicated as follows: *** at the $p \leq 0.001$ level, ** at the $p \leq 0.01$ level, and * at the $p \leq 0.05$ level.

sample schools were also more likely to follow a prescribed sequence of steps to respond to students who read below grade-level benchmarks (95 percent, compared with 88 percent for reference sample schools). Impact sample and reference school samples were not significantly different in their use of data to monitor student progress following implementation of reading interventions for students suspected of having a Specific Learning Disability.

Comparison of Reading Services Between Reading Groups at Different Skill Levels

- **Impact sample schools followed RtI practices of adjusting student tier placement during the 2011-12 school year. In Grade 1, about three-fourths of students remained in the same reading tier, and one-fourth of students moved between tiers, from fall to winter.**

As shown in Figure ES.2, 59 percent of students in Grade 1 in impact sample schools were placed in Tier 1 as their highest tier in fall 2011. (Results are similar for other grades.) Fewer students were placed in Tier 2 or 3 as the highest tier in which they received services — 25 percent and 16 percent, respectively. This arrangement reflects that Tier 3 was typically reserved for students who had not responded to Tier 2 interventions, although some students were placed directly in Tier 3 in the fall. The majority of students placed in Tier 1 or Tier 3 remained there in the winter: 86 percent of students who began in Tier 1 remained in Tier 1, and 65 percent of students in Tier 3 in the fall remained in Tier 3 in the winter. In contrast, about half the students initially assigned to Tier 2 in the fall remained in Tier 2 in the winter, while the other half moved either to Tier 1 or Tier 3. Across all tiers in Grade 1, 74 percent of students remained in the same reading tier.

The stability of tier placement for the majority of students was coupled with movement to different tiers for other students. These patterns, as well as school reports of the types of data they used to make placement decisions, indicate that schools used screening data to adjust students' tier placement.

- **Impact sample schools varied in how they organized and delivered reading group services, in some ways differing from descriptions of RtI in prior literature.**
    - **In Grade 1, 45 percent of schools provided intervention services to some groups of students at *all* reading levels, rather than only for reading groups *below* grade level.**
    - **In Grade 1, 67 percent of schools provided at least some reading intervention *during* the core reading block, rather than only *in addition to* the core.**

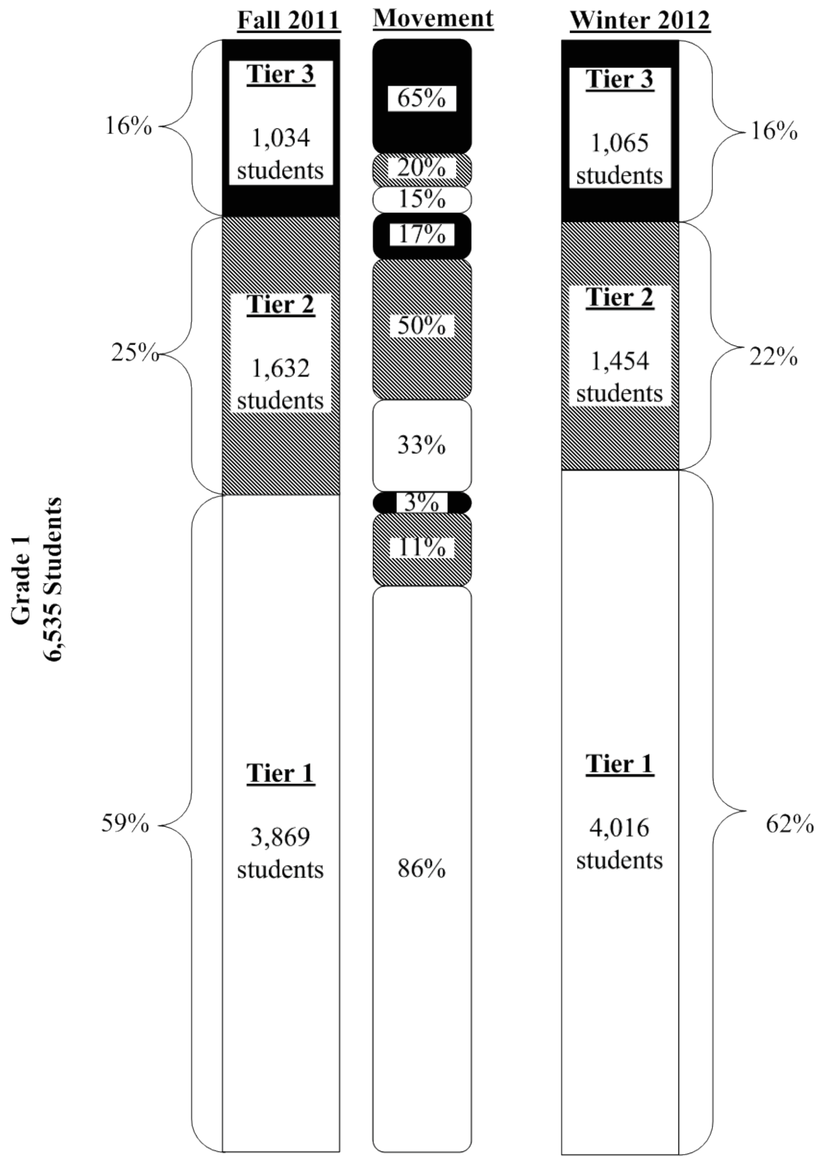
Although all impact sample schools complied with RtI implementation criteria, some schools showed variations on three aspects of RtI implementation described in prior literature. First, prior studies that designed or monitored the delivery of Tier 2 or 3 intervention services generally served only students reading *below* grade level. In contrast, 45 percent of schools in the impact sample offered reading intervention services to at least some students reading at or above grade level, as well as to those reading below grade level. However, these schools did not necessarily provide intervention services for *all* students at or above grade level. (Results are similar across grades; discussion here focuses on Grade 1.)

Second, previous studies of small-group intervention services often designed intervention as supplemental services that occurred *in addition to* the core reading block time. This study,

The Response to Intervention (RtI) Evaluation

Figure ES.2

Student Distribution, by Tier, and Highest Tier Movement



SOURCES: Fall 2011 and winter 2012 tier placement data.

NOTES: Students placed in Tier 1 typically receive only core reading instruction; those placed in Tiers 2 and 3 typically receive core reading instruction plus intervention services. Tier assignment occurs based on results from screening assessments conducted in the fall and winter. Each segment is shaded to represent the proportion of students who remain in that same tier between fall and winter or who move to a different tier (shown in different shading). The Grade 1 school sample size was restricted to 89 schools that had at least one student in each of Tier 1, Tier 2, and Tier 3 in both fall and winter.

in contrast, found that 69 percent of schools in the impact sample offered at least some intervention services *during* the core. In such schools, intervention may have displaced instruction time and replaced some small-group or other instruction services with intervention services. As a result, reading intervention services may have been different from, but not necessarily supplemental to, core reading instruction.

Third, in contrast to more controlled studies of RtI that have relied on non-classroom teaching staff to provide intervention services, the current study included intervention services provided by whoever was designated by schools to provide these services. This study found that, even in schools using the more traditional model of providing intervention services only to readers below grade level, classroom teachers played an additional role and provided intervention services to 37 percent of those groups in Grade 1. These results suggest that impact sample schools adapted time and staff resources to address student needs within an RtI framework.

- **Schools increased the intensity of both small-group instruction during the core and intervention services offered to reading groups below grade level relative to groups reading at or above grade level: group size was smaller, and instruction time was longer. A larger percentage of intervention groups that were below grade level than above it addressed phonics and phonemic awareness.**

The study research team examined whether schools provided more intense services to groups of students reading below grade level than to groups reading at or above grade level, by looking at differences in small-group instruction services during the core reading block (provided by teachers to all students in the class), as well as at differences in reading intervention services delivered either during or outside the core (provided by either teachers or interventionists for students in need of targeted reading support). Results are similar across grades; discussion here focuses on Grade 1.

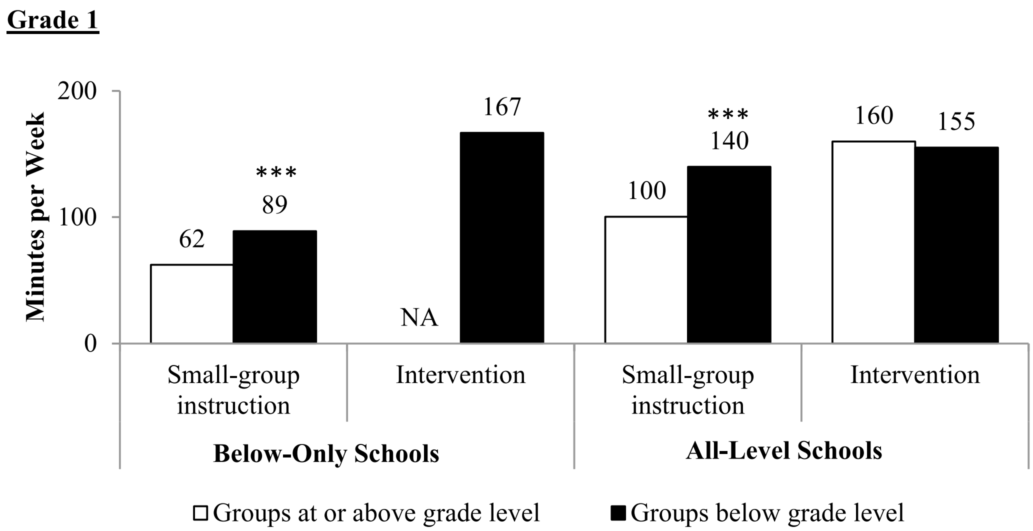
One way that schools provided more intense services was by reducing the size of groups receiving either instruction or intervention services. For small-group instruction during the core reading block in Grade 1, groups for readers below grade level served about one fewer student than groups reading at or above grade level. For reading intervention services in schools that intervened for groups at all reading levels in Grade 1, there were 1.5 fewer students in intervention groups below grade level than in intervention groups at or above grade level.

Weekly small-group instruction time during the core in Grade 1 was about 43 percent longer (27 minutes) for groups below grade level than for those at or above grade level, as shown in Figure ES.3. In schools that provided intervention only to readers below grade level in

The Response to Intervention (RtI) Evaluation

Figure ES.3

Service Contrast for Minutes per Week: Difference Between Groups At or Above Grade Level and Below Grade Level in Below-Only and All-Level Schools, for Grade 1



SOURCES: Teacher survey and interventionist survey.

NOTES: "Small-group instruction" refers to services provided by teachers during the core reading block to all students. Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. The Below-Only school sample represents schools that have at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that have at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services. No tests were performed between intervention groups in Below-Only schools, which do not provide intervention to At or Above grade-level groups. Means reflect rounding.

Statistical significance is indicated as follows: *** at the $p \leq 0.001$ level, ** at the $p \leq 0.01$ level, and * at the $p \leq 0.05$ level.

the corresponding grade, those groups received 89 minutes per week of small-group instruction time, compared with 62 minutes for groups at or above grade level. In schools that provided intervention services to all reading levels in the corresponding grade, weekly small-group instruction time during the core was 140 minutes per week for groups below grade level, compared with 100 minutes for groups at or above grade level. Unlike the differences in weekly small-group instruction time during the core, the difference in time provided to *intervention* groups serving students reading below grade level, compared with those reading at or above grade level, is not statistically significant in schools that provided services to all reading levels.

The reading skills that were addressed differed by the reading level of the group. (Results are similar across grades; discussion here focuses on Grade 1.) While 90 percent to 92 percent of groups below grade level for small-group instruction during the core reading block included content on phonics, about half (46 percent to 52 percent) of groups at or above grade level included that content. Among both small groups meeting during the core and reading intervention groups, 70 percent or more of groups both at or above and below grade level included content on fluency, reading comprehension, and vocabulary, regardless of whether the group served students reading below grade level or those reading at or above grade level. These findings suggest that small reading groups and intervention groups focused on multiple skills but that the more elemental skills of phonics were more likely to be addressed by small groups reading below grade level than by small groups at or above grade level.

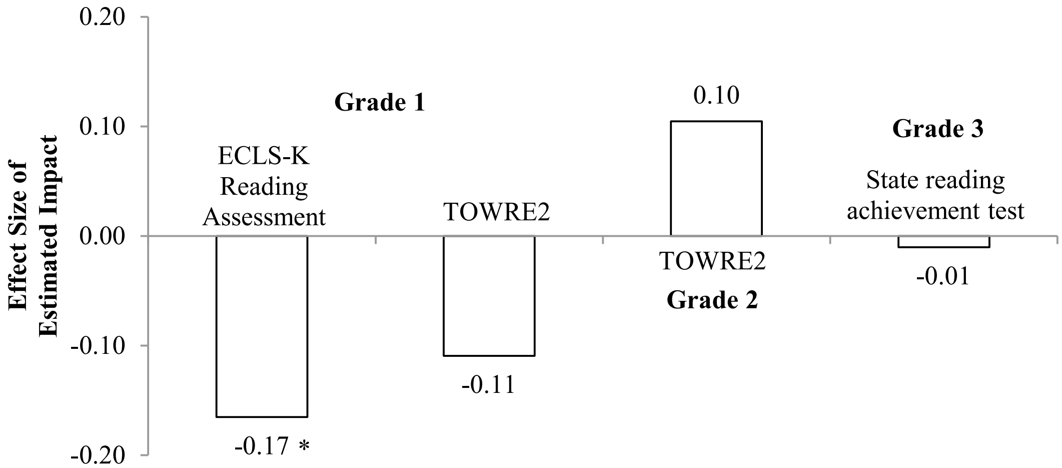
Impacts on Reading Outcomes of Students

- **Assignment to Tier 2 or Tier 3 intervention services in impact sample schools had a negative effect on performance on a comprehensive reading measure for first-graders just below the Tier 1 cut point on a screening test. The estimated effects on reading outcomes in Grades 2 and 3 are not statistically significant.**

Figure ES.4 presents the estimated effects across four outcomes and three grade levels. The height of each bar in the figure represents the magnitude of the estimated effect, and an asterisk indicates that an estimated effect is statistically significant at the 5 percent level. The study-administered tests were the ECLS-K:2011 comprehensive reading measure, used in Grade 1, and the TOWRE2 measure of decoding fluency, used in Grades 1 and 2. Data from state reading tests provided outcomes for Grade 3 students. Figure ES.4 shows that the estimate for the effect of assignment to Tier 2 or Tier 3 intervention on the ECLS-K Reading Assessment measure is –0.17 standard deviation and is statistically significant (p-value = 0.002). For students who were close to the cut point and were assigned to receive intervention, a negative effect of this magnitude is equivalent to approximately one-tenth of a year less learning than what they would have achieved had they not been assigned to intervention. The estimate for the effect of treatment assignment on the TOWRE2 Sight Word Efficiency test for first-graders close to the cut point is –0.11 standard deviation and is not statistically significant (p-value = 0.057); for second-graders close to the cut point, the estimated impact is +0.10 standard deviation and is not statistically significant (p-value = 0.084). The estimated impact on the state reading achievement test for third-graders in the vicinity of the cut point is –0.01 standard deviation and is not statistically significant (p-value = 0.823).

Figure ES.4

Estimated Impacts of Assignment to Tier 2 or Tier 3 Intervention Services for Students Within Optimal Bandwidth, by Grade and Outcome Measure



SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement test scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

Statistical significance at the $p \leq 0.05$ level is indicated as *.

ECLS-K Reading Assessment is a comprehensive reading measure; TOWRE2 is a decoding fluency exam; the state achievement test is a comprehensive reading measure.

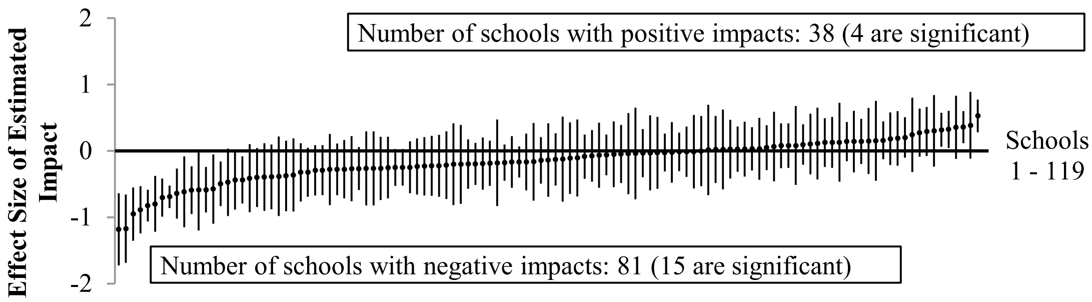
- **The estimated impacts of reading interventions on reading outcomes vary significantly across schools. This is true for all four outcomes across three grade levels.**

Figure ES.5 presents results for the Grade 1 ECLS-K Reading Assessment comprehensive reading measure to illustrate the extent and significance of impact variation across schools. The figure plots the estimated impact of assignment to intervention on Grade 1 students' ECLS-K Reading Assessment scores for every RtI school in the study sample. The estimates are ordered by their magnitude. A solid dot represents the impact estimate for each school, and a vertical line running through each solid dot represents the respective 95 percent confidence

Figure ES.5

Distribution of School-Level Impact Estimates of Actual Assignment to Tier 2 or Tier 3 Intervention Services for Grade 1

<u>ECLS-K Reading Assessment</u>



SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: The outcome was standardized to have a standard deviation of 1, so impact estimates are reported in effect-size units. A chi-squared test was used to test the statistical significance of the variation in the empirical Bayes impact estimates.

interval of the estimated impact. In this example, the estimated school-level impacts on the ECLS-K Reading Assessment score for Grade 1 range from –1.18 to +0.53 standard deviations in effect size. Of the 119 schools included in the impact analysis for Grade 1, there are 15 schools with significant negative findings and four schools with positive and significant findings. Similar patterns of variation were found for the estimated impacts on the other three reading outcomes. Statistical tests show significant variation in impact estimates across schools — for all four outcomes across three grade levels. This finding indicates that the estimated impact could be more negative or more positive in some schools than others, regardless of the overall average impact estimate.

- **The school-level features and student characteristics examined are not consistently associated with school impacts across grades and reading outcomes.**

Across grades or outcomes, there is no consistent association between the impact estimates and examined school features, which include measures of school-level RtI practices, school context, and composition of the student population. (See Box ES.1 for details.) Specifically, the analysis yielded no statistically significant associations between school features and the impact estimates for the two comprehensive reading measures: Grade 1 ECLS-K Reading Assessment scores and Grade 3 state achievement test scores. There are sporadic associations for the decoding-fluency measure for Grades 1 and 2.

Box ES.1

Exploratory Factors Examined in the RtI Evaluation

School-level RtI practices: Whether a school used single or multiple screening tests to assign students to tiers, the proportion assigned to Tier 2 or Tier 3 intervention services, whether the school provided intervention to at least one group at all reading levels, and the proportion of intervention groups served outside the core reading block

School context factors: Overall school reading performance in a baseline year, eligibility for Title I funds, and use of RtI practices for behavior-related interventions

Composition of the student population: Proportion of students who are male or who were English Language Learners, overage for grade, or low-income status or who had an Individualized Education Program (IEP) on account of a student disability

At the student level, for some outcomes and grades, students in specific learning circumstances (for example, those who were overage for grade or who had an Individualized Education Program [IEP]) appear to have been affected by the treatment more negatively. But this finding is not consistent across outcomes and grade levels, and it applies only to students in these circumstances who scored near the cut point on their fall screening test.

How to Interpret the Impact Findings and How This Study Differs from Prior Literature

The study uses a Regression Discontinuity (RD) design for its impact estimation. While this design demonstrates a causal relationship between assignment to receive intervention services and reading test outcomes in the impact sample, it also requires caution when interpreting the impact findings. In particular, the RD design estimates the impact of assignment to intervention by comparing outcomes of students just above or just below the cut point. Findings based on this design, therefore, cannot be generalized to all students receiving intervention services.

This is different from a randomized controlled trial (RCT), whereby similar eligible students are randomly assigned either to receive interventions or not to receive them. As a result, this design provides estimates of the average effect of intervention for students who would be added or dropped by marginally changing the eligibility criterion. In this sense, these results are relevant for decisions about expanding or reducing the scope of intervention but not, necessarily, for decisions about offering or not offering intervention. It would be misleading to conclude from these findings that providing increasing intensity of services to the students most at risk (for example, students whose screening test scores are far below the cut point) is inappropriate or ineffective.

In addition, this study is unique in the sense that it examines the RtI system as it operated in multiple states in a large sample of experienced schools that had implemented RtI on their own, without monitoring or support from researchers. This is different from most existing efficacy studies, in which the scale of the treatment is small (usually samples consist of fewer than 100 students and only a handful of schools) and the design and implementation of the RtI interventions are closely controlled by the researchers.

In order to understand the primary impact findings, the study explores the relationship between the impact estimates and school characteristics and RtI practices related to assignment to intervention. The key factors listed in Box ES.1 do not consistently explain the pattern of findings across grades. Unexplored but plausible factors that may be related to negative impacts of assignment to intervention on some Grade 1 students include (1) false or incorrect identification of students for intervention, (2) mismatch between reading intervention and the instructional needs of students near the cut point, and (3) poor alignment between reading intervention and core reading instruction.

This page has been left blank for double-sided copying.

Chapter 1

Introduction and Overview of Report

Under the 2004 reauthorization of the Individuals with Disabilities Education Act (IDEA),[1] states and districts are allowed to use a portion of federal special education funds to provide coordinated early intervening services to students at risk of reading difficulties or other academic or behavioral problems.[2] One of the primary early intervening approaches that has emerged is called "Response to Intervention" (RtI). In RtI frameworks focused on reading, "intervention" traditionally refers to additional support for students who have difficulty reading at or above grade-level standards. RtI incorporates a range of assessment, instruction, and intervention principles that schools may implement in various ways, rather than a prescribed set of tools and materials. As discussed in more detail below in this chapter, the RtI framework includes (1) offering multiple tiers of support for students, depending on the level of reading difficulty they may be experiencing; (2) allocating staff to provide that tiered support to students; and (3) collecting and using data to make instructional and intervention decisions corresponding to the tiered structure. Ideally, this framework facilitates the delivery of appropriate, evidence-based instruction to all students, including those at risk of reading difficulties. If students do not respond to core instruction, they are identified for intervention services. If students do not respond to the initial intervention, they receive additional intervention, which may include more intensive services and possibly evaluation for special education.[3]

The adoption of an RtI framework in districts, especially for reading, has gained momentum over the past 15 years.[4] As of the 2008-09 school year, 71 percent of districts and 61 percent of elementary schools were using RtI in at least one classroom in the district or school.[5] For school year 2008-09, when this study began its planning and design, 70 percent of districts with elementary schools reported using RtI in reading/language arts, and 36 percent of districts reported using RtI to address behavior issues.[6]

In 2008, the Institute of Education Sciences (IES) at the U. S. Department of Education awarded a contract to MDRC and its partners SRI International, Instructional Research Group,

---

[1]IDEA (Public Law 108-446) (2004).

[2]34 CFR 300.226(a) and 34 CFR 300.307(a)(2).

[3]The RtI framework is also applied to address student behavior, often under the label of "School-Wide Positive Behavioral Supports." See the discussion at http://www.pbis.org/ for the Technical Assistance Center on Positive Behavioral Interventions and Supports, established by the Office of Special Education Programs, U.S. Department of Education.

[4]Bradley et al. (2011).

[5]In Bradley et al. (2011), "implementing RtI in at least one classroom in the district or in a school constitutes 'using RtI' in the context of the report" (p. 50).

[6]Bradley et al. (2011).

and Survey Research Management for this evaluation to examine the implementation and impacts of RtI practices for elementary school reading.[7] This study has several components. First, it describes school-level RtI reading practices in 1,105 elementary schools for the 2011-12 school year in 13 states, and it compares these practices with those for a sample of 146 elementary schools that were experienced in implementing RtI in reading. This latter sample is used for the impact analysis and is called the "impact sample." Second, the study describes how the impact sample of schools offered reading support to students categorized as Somewhat Below or Far Below benchmark reading levels, and it examines the difference in services for groups of students reading below grade level and students reading at or above grade level.[8] Finally, for the impact sample schools, the study estimates the effects of receiving intervention services on reading outcomes, with a focus on a particular type of student on the margins of receiving those services. Through the use of a *Regression Discontinuity (RD) design,* this study assesses the impact on reading achievement of assignment to reading intervention services for students whose score on a fall reading screening test is near a predefined *cut point* that determines eligibility for intervention services.

This study was not designed as an impact analysis of IDEA services on outcomes for *all* children receiving reading interventions (in particular, those well below the cut point of eligibility) or as an evaluation of the *overall impact* of RtI on students and their schools. Consequently, the study does *not* provide evidence to inform schools' decisions to adopt RtI systems to improve *overall* student outcomes. However, the findings from the study can increase awareness of RtI implementation and can indicate areas for future investigation to understand the impacts of RtI practices.

Contributions of This Study

This evaluation's analysis of RtI implementation and the impact of interventions on reading achievement expands the field's knowledge about RtI in three ways. First, this study describes RtI adoption in a large sample of experienced schools in multiple states; previous studies described district- or state-level implementation only. Second, this study describes practices in schools that have adopted RtI on their own, rather than as a condition of participation in an impact study. As a result, this study captures implementation in a real-world context, rather than a controlled experimental setting. Third, this study's research design differs from earlier impact studies. Most prior impact studies compare the outcomes for intervention-eligible students who received intervention with the outcomes of similar intervention-eligible students who did *not*

---

[7]This research is one aspect of the IES response to the part of IDEA 2004 calling for a national assessment of the implementation progress of IDEA and the relative effectiveness of the law (Section, 664[b]).

[8]All student groups reading below grade level are combined to describe the service contrast with student groups reading at or above grade level and to best align with the contrast tested in the impact analysis.

receive intervention, often employing a random assignment design to create these two similar groups.[9] By contrast, this study includes sites with ongoing RtI programs that ranked students based on reading skills using a fall screening test and that used a cut point,[10] or a rating below which students generally received reading interventions. The impact estimation strategy employed in this study compares the outcomes for students just below and just above the cut point, providing an estimate of the impact of interventions on the students slightly below the cut point rather than for the full range of students served by interventions. This impact on the marginally eligible student served is important for assessing the effective targeting of intervention resources, but it does not assess whether the RtI framework as a whole is effective in improving student outcomes or even whether reading interventions are effective for all students receiving interventions in Grades 1 to 3 in the corresponding schools (including students well below the cut point for intervention).

Chapter 1 sets the context for the study by describing evolving public policy as it relates to RtI and the research in elementary reading that contributed to the development and adoption of RtI practices. It then outlines key RtI practices and presents the logic model and the research questions that motivate the study. The chapter concludes by providing a roadmap of the remaining chapters in the report.

Evolving Federal and State Policies

The increasing use of the RtI framework is an outgrowth of a debate related to special education policy. The 2004 reauthorization of IDEA changed procedures for identifying children with a Specific Learning Disability (SLD) — the disability category most associated with reading difficulties. These changes were motivated by dissatisfaction with the previous eligibility standard. This standard required educators to document an "educationally significant discrepancy" between achievement of specific skills (for example, reading performance) and general ability (that is, overall intellectual functioning as measured by an IQ test) that could not be explained by visual, hearing, or motor disabilities; emotional disturbances; mental retardation; or environmental, cultural, or economic disadvantage.[11] Critics asserted that waiting until students' achievement fell substantially below their ability as measured by IQ tests before providing them with intervention services was a "wait to fail" approach that deprived students of the benefits of early assistance.[12] Experts raised the concern that the discrepancy

[9]Gilbert et al. (2013); Vadasy, Jenkins, and Pool (2000); D. Fuchs et al. (2008); Mathes et al. (2005); McMaster, Fuchs, Fuchs, and Compton (2005).

[10]"Cut point," "cut score," and "cutoff" are used interchangeably in the RD design literature. See Jacob, Zhu, Somers, and Bloom (2012).

[11]34 CFR 300.8(c)(10).

[12]Carnine (1997); Bradley, Danielson, and Hallahan (2002); Fletcher, Coulter, Reschly, and Vaughn (2004); U.S. Department of Education, Office of Special Education and Rehabilitative Services (2002).

approach contributed to a disproportionate identification for special education by race, ethnicity, and cultural background, because children from disadvantaged backgrounds were less likely than other children to receive high-quality instruction and behavioral support and because the process of referring, evaluating, and assigning children for special education relied on subjective and idiosyncratic professional judgment instead of systematic screening procedures.[13]

To address the concerns, the December 2004 reauthorization of IDEA and the accompanying Part B regulations finalized in August 2006 include three broad and interconnected changes in federal policy. First, states cannot require districts to identify SLD students using a discrepancy between a student's ability and achievement.[14] Rather, districts are permitted to use an SLD identification process based on the student's response to scientific, research-based interventions.[15] Second, schools are told that they should not identify students for special education services without having first determined that the students have received appropriate and sufficient instruction in general education.[16] Third, the law allows districts to use up to 15 percent of their IDEA Part B special education funds to develop and implement coordinated early intervening services for students not yet identified as needing special education and related services but who need additional academic or behavioral support to be successful in general education classrooms.[17] This last change allows federal funds to be used for RtI services.

In response to these changes, multiple organizations began to offer support and guidance to states and districts on how to design and implement an RtI framework to address academic and behavioral problems. This multiplicity resulted in a variety of approaches to implementation rather than a single model against which fidelity is measured. The U.S. Department of Education, Office of Special Education Programs (OSEP),[18] funded four national technical assistance centers[19] to provide educators guidance in identifying valid, reliable, and diagnostically accurate tools for universal screening and progress monitoring and in identifying research-based language arts, math, and behavioral interventions. IES, through the What Works Clearinghouse (WWC), produced the practice guide *Response to Intervention in Elementary Reading*

[13]National Research Council (2002).

[14]34 CFR 300.307(a)(1); Bradley et al. (2011). In the same study, 37 states indicated that they allowed use of screening data as an alternative to using IQ-achievement discrepancy in determining SLD eligibility. Additionally, six states permitted a discrepancy model but also required inclusion of RtI systems; seven states used RtI systems or an alternative method and disallowed the use of a discrepancy model.

[15]34 CFR 300.307(a)(2).

[16]34 CFR 300.306(b)(1).

[17]34 CFR 300.226(a).

[18]U.S. Department of Education (2015).

[19]The Center on Positive Behavioral Interventions and Supports, National Center on Progress Monitoring, National Center for Response to Intervention, and National Center on Intensive Intervention.

(hereafter referred to as the "IES Practice Guide on RtI for Elementary Reading"), derived from research evidence and experts' recommendations.[20] Additionally, professional organizations supported adoption and implementation of best practices.[21] Likewise, states have undertaken activities to support RtI implementation.[22]

From these multiple efforts to support RtI designs and implementation, a general framework emerged that is discussed in more detail later in this chapter. There is variation within this general framework, as states, districts, and schools adopting RtI brought their own perspectives and needs to the effort. One of the goals of this evaluation is to describe the differing ways that RtI was implemented in a variety of schools, rather than to assess the fidelity of implementation to a single RtI design.

RtI in Elementary Reading

Based on a comprehensive review of reading research, the National Reading Panel in 2000 identified five essential components of reading instruction: phonemic awareness, word attack (decoding or pronouncing unfamiliar words), fluency, vocabulary, and comprehension.[23] These skills relate to a National Research Council report that summarized the "accomplishments that the successful learner is likely to exhibit during the early school years,"[24] highlighting the sequence of skills that early readers develop as they move from letter-sound understanding to more complex and irregular words to growing fluency in reading and understanding of text. Tiered interventions, described below, target one or more of the five components of reading instruction.

A burgeoning research literature on the uses and advantages of universal screening of students to determine risk of reading difficulties also supported the foundation of the RtI

[20]Gersten et al. (2009).

[21]For example, the National Center for Learning Disabilities created an RtI Action Network, and the International Reading Association published a framework to help reading educators understand the critical features of RtI in the elementary grades and to assist them in coordinating RtI services to improve reading outcomes (Fuchs, Fuchs, and Vaughn, 2008; International Reading Association, 2009). The National Association of State Directors of Special Education (NASDSE) produced an RtI Implementation Blueprint to provide states a step-by-step approach to move implementation to the local level (National Association of State Directors of Special Education, 2007), followed by blueprints for the district and school-building levels that illustrated core RtI features, implementation stages, and resources (National Association of State Directors of Special Education, 2008a, 2008b).

[22]The IDEA National Assessment of Implementation Study (also known as the NAIS study or as Bradley et al., 2011) found that, as of the 2008-09 school year, all but two states reported having a state-level RtI task force, commission, or internal working group.

[23]National Reading Panel (2000).

[24]Snow, Burns, and Griffin (1998), pp. 79-84.

framework, which involves screening all students at least twice a year.[25] With the advent of comprehensive and accessible electronic data systems for screening and progress monitoring, schools were positioned to collect and analyze data for making decisions about student performance and progress more frequently during the year. Further, interest was growing in school accountability systems combined with data-based instruction.[26]

Research on early reading intervention laid the groundwork for implementing interventions in an RtI framework.[27] This research provided information on implementation measures (such as frequency, duration, and group composition and size) and the targeting of instruction on specific reading skills (for example, phonics, fluency, and comprehension).[28] The resources described in this section emphasize strong reading practices and standards. These standards, and the growing use of data systems to identify students in need of support, contributed to adoption of the RtI framework to provide reading services in a tiered model of support.

Key RtI Practices

This study aims to describe the practices and processes schools use to implement the RtI framework in elementary reading. Model developers and researchers vary in their emphasis on and approach to specific practices,[29] but versions of the RtI framework generally incorporate the following three sets of core practices.

1. Provide Multiple Tiers of Support Differing in Intensity[30]

The concept of multiple tiers of increasing support has often been depicted as a triangle or pyramid.[31] Figure 1.1 uses a pyramid to show a hypothetical distribution and movement of students across three tiers. *Tier 1* is the core reading instruction provided to all students. Students who score below grade-level benchmarks (to the right of the dashed line) also are placed in Tier 2 in addition to Tier 1. In *Tier 2,* schools provide targeted interventions to those students who appear to be at risk for experiencing reading problems, based on scores on valid and reliable screening measures.

---

[25]Compton, Fuchs, Fuchs, and Bryant (2006); Fuchs, Fuchs, and Compton (2004); Jenkins, Hudson, and Johnson (2007); Jenkins and O'Connor (2002); McCardle, Scarborough, and Catts (2001); O'Connor and Jenkins (1999); Scarborough (1998); Speece, Mills, Ritchey, and Hillman (2003).

[26]Black and Wiliam (2009); Fuhrman and Elmore (2004); Gallagher, Means, and Padilla (2008).

[27]Blachman, Ball, Black, and Tangel (1994); O'Connor, Fulmer, Harty, and Bell (2005); Torgesen et al. (2001); Vadasy, Jenkins, and Pool (2000); Vellutino et al. (1996).

[28]Vaughn (2008).

[29]Burns and Ysseldyke (2005); Callender (2007); Fuchs and Fuchs (2002); Gersten et al. (2009); Haager, Klingner, and Vaughn (2007); Jimerson, Burns, and VanDerHeyden (2007); Johnson, Mellard, Fuchs, and McKnight (2006); Shapiro, Zigmond, Wallace, and Marston (2011).

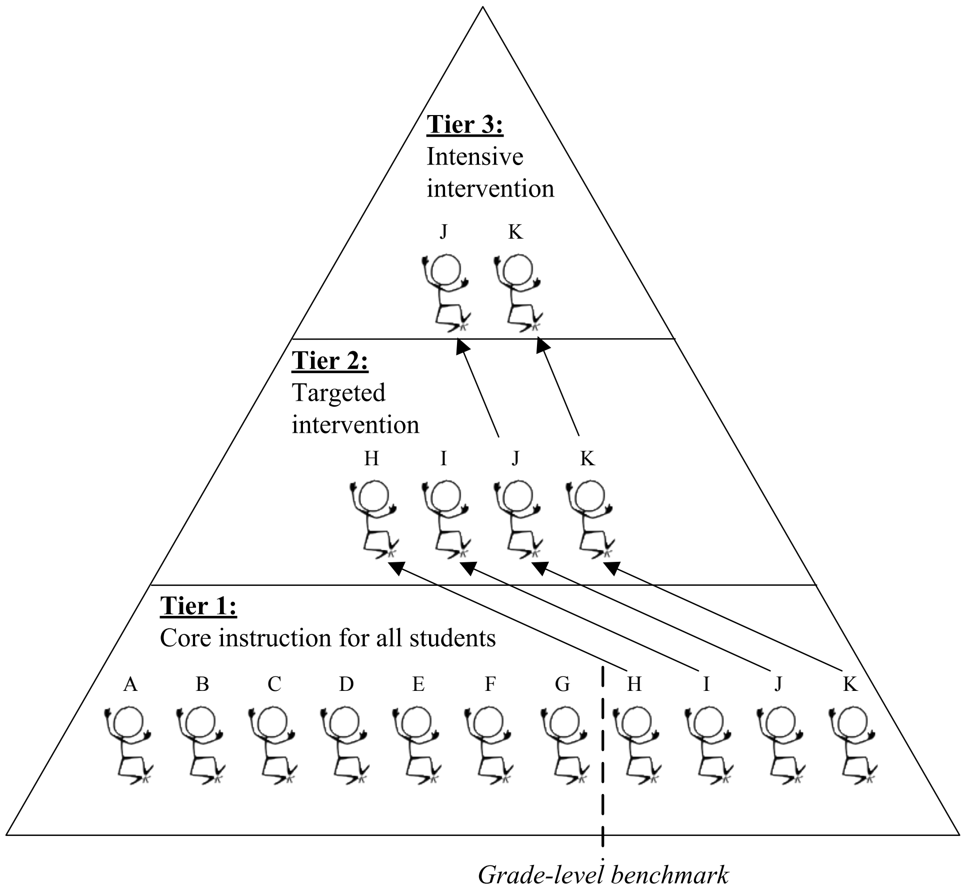
[30]Gersten et al. (2009); Haager, Klingner, and Vaughn (2007).

[31]RTI Action Network (2015).

The Response to Intervention (RtI) Evaluation

Figure 1.1

Distribution of Students Across Tiers in a Typical RtI Model



NOTES: This illustration of a school's placement of students is based on screening test scores. Numbers are hypothetical and are not based on actual data from this study.

Students are arrayed from low to high risk, with student K being at highest risk.

Typically, students who do not respond to Tier 2 intervention may then receive Tier 3 intervention. *Tier 3* interventions are reserved for students who are performing Far Below grade-level benchmarks, who are at high risk of reading failure, and/or whose progress is unsatisfactory after having received a Tier 2 intervention for a reasonable time.[32] In some cases, Tier

[32]Gersten et al. (2009); National Center on Response to Intervention (2010); Wanzek and Vaughn (2008). Students who are at some risk of reading failure as evidenced during universal screening should receive target-

3 services are in addition to Tier 2 services; in some cases, they may replace Tier 2 services. Tier 3 services generally are the most intensive, involving more time per week and smaller groups of students than Tier 2 services. Fewer students receive the more intense Tier 2 and 3 services.

Figure 1.2 illustrates the service delivery model for RtI that is typically discussed in earlier reading intervention studies. In theory, students identified by screening tests as reading "At or Above" grade level have Tier 1 as their highest tier placement; those identified as reading "Somewhat Below" grade level have Tier 2 as their highest tier placement; and those identified as reading "Far Below" grade level have Tier 3 as their highest tier placement.[33] In practice, however, tier placement may not correspond precisely with reading level as measured by screening tests.[34] Students may cycle in and out of placement in intervention tiers.

All students receive core reading instruction during a designated core reading time block, shown in the top segment of each bar in Figure 1.2. This time includes whole-class and small-group instruction, for which students may be grouped with peers at similar reading levels. In addition, students who are identified for Tier 2 or 3 receive additional intervention services outside, or in addition to, the core reading block.[35] This means that students who are reading Somewhat Below grade level or who are at moderate risk of reading difficulties receive Tier 2 interventions plus core instruction; students who are reading Far Below grade level or who are at higher risk and do not respond to Tier 2 interventions receive Tier 3 interventions plus core instruction. After a student responds to intervention services, a school may decide to conclude those services for that child.

This evaluation explores the extent to which these RtI approaches are present in a large sample of schools. Schools in the impact evaluation sample described later in the report placed some students immediately in Tier 3 in the fall, without waiting to see whether they responded to Tier 2 services. Some schools made this assignment based on whether students scored below

---

ed evidence-based support services of moderate intensity. While sometimes these services involve individualizing supports to address a student's reading deficits, most interventions occur in small groups rather than one-on-one.

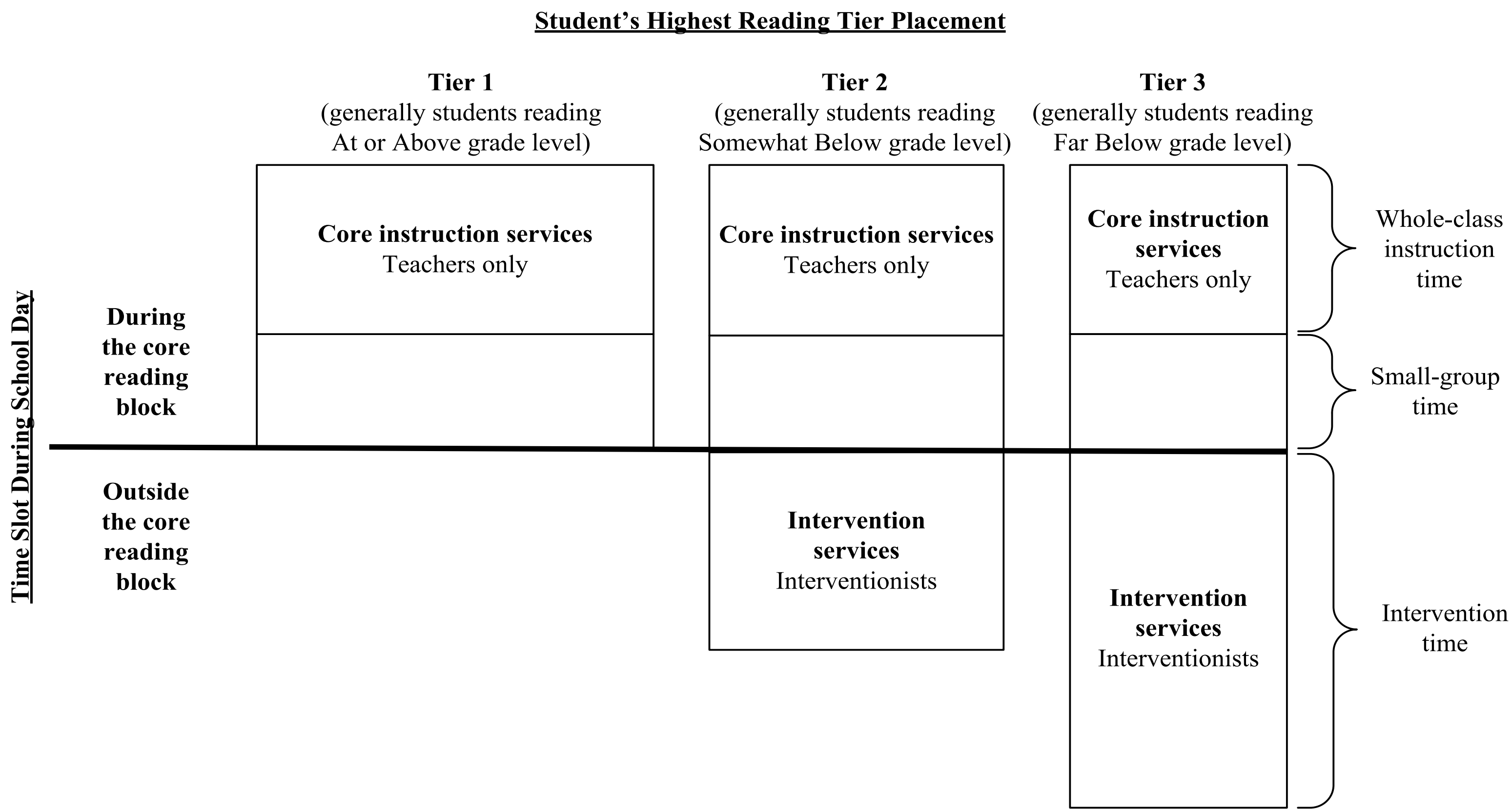
[33]Screening tests use grade-level benchmarks that refer to a student's likelihood of reading at that level by the end of the school year.

[34]While it is generally true that placement in Tier 2 corresponds with reading "Somewhat Below" grade level and placement in Tier 3 corresponds with reading "Far Below" grade level, this is not always the case in practice. Either limitations on service availability, or the professional judgment of school staff, may result in some students who read "Far Below" grade level having Tier 2 as their highest tier placement, and other students who read "Somewhat Below" grade level having Tier 3 as their highest tier placement.

[35]If services are provided outside the reading block, students may receive these intervention services during subjects other than reading.

Figure 1.2

Allocation of Time and Services in RtI Schools, by Student Reading Level, as Described in the Literature



NOTE: The correspondence between tiers and reading levels was described in this study's surveys as "students Somewhat Below grade-level benchmarks (sometimes called 'Tier 2 students')" and "students Far Below grade-level benchmarks (sometimes called 'Tier 3 students')."

a prespecified point on the fall screening tests, different from the cut point for Tier 2 intervention; others made the determination based on other criteria such as teacher judgment.

2. Allocate School Staff to Perform RtI Practices

In order to implement RtI, schools allocate staff to administer and analyze tests that determine tier placement, make determinations about which students need more intense reading support, identify appropriate interventions, and deliver these interventions in small groups or individually. Staff who are involved in data use may include school administrators, reading coaches, and school psychologists, while a variety of staff may provide interventions.[36]

3. Use Data to Make Instructional and Intervention Decisions

Reliance on data for decision-making runs throughout RtI, generally in three main stages: begin with universal screening to decide which students are at risk of reading failure; monitor student progress, and make decisions about tier placement; and make decisions on identification of students for special education. Chapter 3 discusses these steps in detail.

Study's Logic Model, Research Goals and Questions, and Organization of the Report

Because a central goal of this evaluation is to estimate the effects of reading interventions on student achievement, the study research team sought schools with more fully developed RtI practices rather than schools in earlier stages of RtI implementation. When the study began screening schools for eligibility, many districts were reporting implementation of RtI at various stages of completeness in some of their schools.[37] This evaluation examines RtI practices and impacts in a sample of schools with at least three years' experience with RtI, referred to as the "impact sample" schools.

This study was designed to answer three sets of major research questions:

- **Comparison of practices between school samples.** How did the prevalence of RtI practices differ between a representative "reference" sample of schools and schools selected for the impact evaluation?[38] To what extent were impact sample schools implementing more RtI practices than the reference sample

[36]Fuchs, Fuchs, Hamlett, and Stecker (1991).

[37]Bradley et al. (2011).

[38]As discussed in Chapter 2, the random sample of schools was drawn from the same states as the experienced schools because RtI policy is heavily influenced by state policy. Thus, this analysis is relevant to the states in the study but not necessarily to other states.

schools? How do special education identification rates in the impact sample compare with the rates for the states as a whole? (Findings are in Chapter 3.)

- **Comparison of reading services between reading groups at different skill levels.** In impact sample schools (those with three or more years of implementing RtI): To what extent did schools place students in tiers as suggested by earlier RtI models? To what extent did schools adjust tier placement during the school year? To what extent is there variation in how schools organize reading services for specific reading levels? To what extent were services for students reading below grade level more intense than for students reading at or above grade level? (Findings are in Chapter 4.)

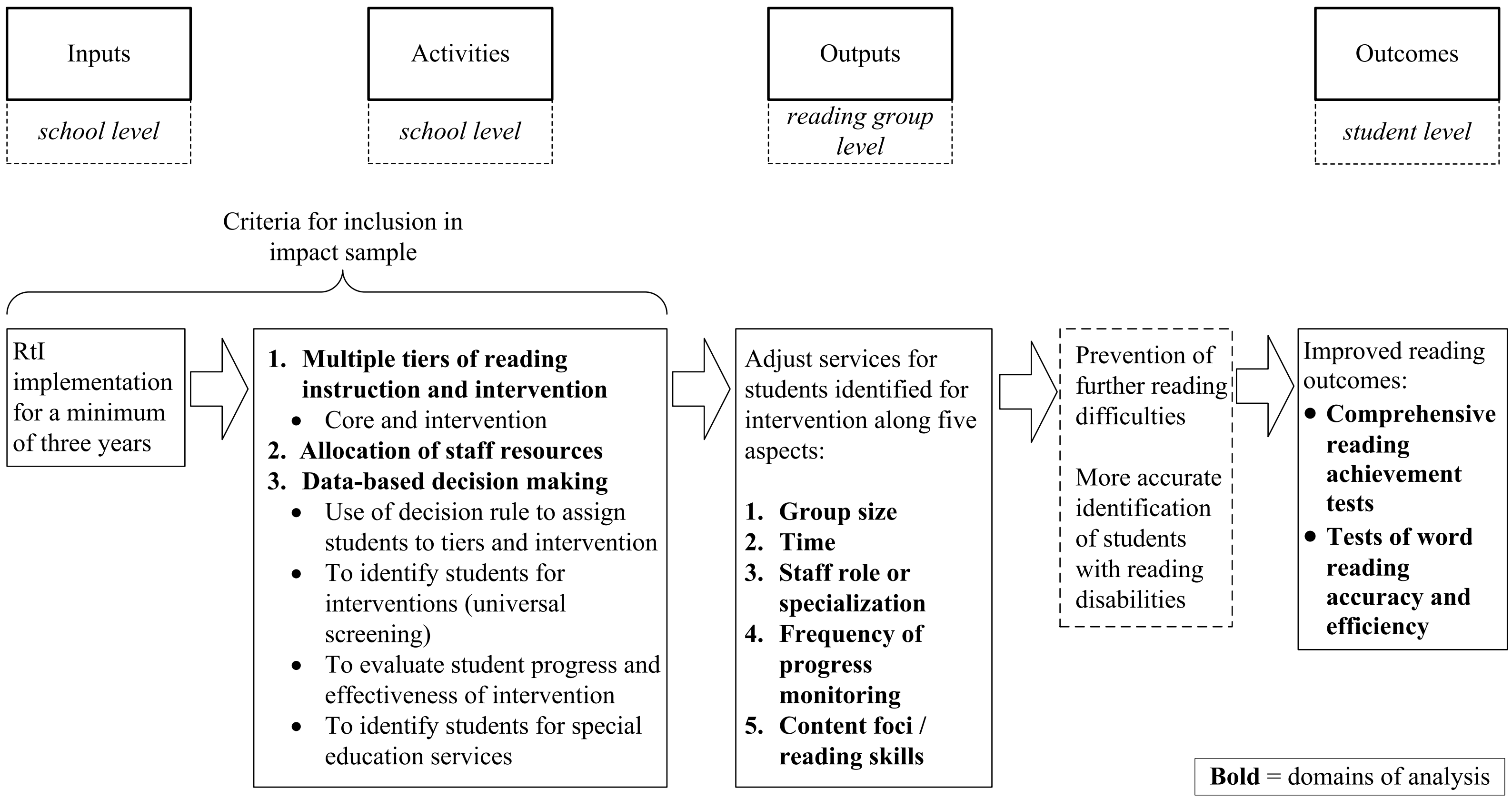
- **Impacts on reading outcomes of students.** For students who fell just below school-determined standards for each grade on screening tests: What were the effects on reading achievement of actual assignment to receive reading intervention services (in addition to core instruction)? What is the extent of variation in estimated impacts across RtI schools? How is the estimated impact associated with certain school features or student characteristics? (Findings are in Chapters 5 and 6.)

Figure 1.3 illustrates the logic model used to frame this analysis of RtI programs in the impact sample. On the left, it shows that schools that were selected for the impact analysis had been implementing RtI for early-grade reading for a minimum of three years (that is, starting no later than the 2009-10 school year). The second column shows that, to be included in the impact sample, schools needed to have in place key RtI practices related to tiered instruction, allocation of staff resources, and use of data for tier placement and progress monitoring. The third column describes ways in which schools implementing these practices can adjust reading services based on the needs of students with a goal of preventing further reading difficulties and more accurate identification of students for special education. These practices are intended to result in improved reading outcomes for students with potential reading problems, as shown in the rightmost column.

The research questions for the evaluation focus on the implementation of RtI practices in elementary school reading on the school level and within impact study schools, as well as the impact of reading interventions on students screened as just below grade-level benchmarks determined by each school. To describe differences in reading instruction and intervention within impact study schools, the five dimensions described in the third column of Figure 1.3 are measured for student groups reading At or Above grade level (who would in theory be expected to receive core instruction only) and then are compared with groups reading below grade level

Figure 1.3

Logic Model for the RtI Evaluation



(who in theory would receive core instruction plus intervention).[39] The Somewhat Below and Far Below grade-level groups are combined in the Chapter 4 analysis to describe the corresponding service contrast with groups reading At or Above grade level.

The report is organized as follows.

- Chapter 2 explains the process of school recruitment for the study and provides an initial overview of the data collected for the study as well as the analytic approaches used to address the three sets of research questions listed above. Chapters 3 through 5 then answer the research questions in sequence.

- Chapter 3 begins the story by outlining how RtI is organized and how prevalent specific practices are in the two samples of schools. The description is anchored by the three key RtI practices listed above in this chapter.

- Chapter 4 focuses on one of those key elements — multiple tiers of increasing intensity of support — and explores to what extent the impact sample schools provided increasingly intense reading instruction and intervention for students who read below grade level. The chapter focuses on such aspects as instructional time provided in small groups, size and composition of small groups, the qualifications of those who provided the services, the frequency of progress monitoring, and the content of reading-group sessions. It also describes the extent of movement between tiers of instruction from the fall to the winter screening assessments.

- With this background on how reading support was adjusted depending on students' needs, Chapter 5 then uses a Regression Discontinuity (RD) design to examine one aspect of the RtI framework: whether actual assignment to receive reading intervention had an impact on reading performance of early-grade students identified as needing help.

- Finally, Chapter 6 explores the relationships between estimated impacts and the characteristics of schools and students.

The findings from this evaluation are relevant to those interested in RtI and those seeking information on how data-based instruction and reading interventions are implemented in elementary schools.

[39] See Chapter 2 for a discussion of the screening tests used to determine student tier placement.

This page has been left blank for double-sided copying.

Chapter 2

School Samples for the Analyses

Chapter 2 introduces the samples of schools used for the analyses in the subsequent chapters of this report on the Response to Intervention (RtI) evaluation. It documents the criteria and procedure used for identifying and confirming schools' suitability for the study, and it describes the resulting school samples and their characteristics. Appendixes provide further detail about the different data sources, samples, and analytic approaches used to address the three sets of key research questions outlined in Chapter 1.

The study relies on two samples of schools for analysis: (1) the *impact sample* of schools that were purposively selected for their experience in implementing the RtI approach for reading in elementary school and for their feasibility of being included in the impact analysis and (2) the *reference sample* of schools from each of the same states as the experienced schools.

Each sample includes schools from 13 states. Experts familiar with RtI systems nominated schools or districts as candidates for the impact sample. The study research team sought to obtain geographic diversity in the impact sample, as well as particular practice and experience criteria, as discussed below. In each of the 13 states in the impact sample, elementary schools were then randomly sampled to receive the school administrator survey in order to form the reference sample.

Impact Sample Selection

Defining the Impact Sample

Public elementary schools that were experienced in the implementation of RtI practices form the impact sample of schools for this evaluation. The study research team defined a school's experience with RtI in several ways. First, schools had to adopt the framework by at least the 2009-10 school year, in order to have at least three years of RtI implementation at their schools at the time of site selection. Second, by at least 2009-10, schools had to begin implementing key practices recommended in the IES Practice Guide on RtI for Elementary Reading[40] that correspond to the key RtI practices described in Chapter 1. Those conditions informed the following initial eligibility criteria, all of which had to be met:

- Use of three or more tiers of increasing instructional intensity to deliver reading services to students

[40]Gersten et al. (2009).

- Assessment of all students (universal screening) at least twice a year
- Use of data for placing students in Tiers 2 or 3
- Use of progress monitoring (beyond universal screening) for students reading below grade level to determine whether intervention is working for students placed in Tier 2 or 3

In addition, for inclusion in the impact analysis, schools needed to meet the following requirements: (1) use a quantitative data-based rating system in fall 2011-12 to identify students in need of more intense reading assistance, (2) meet thresholds for the number of students in each grade and in each tier of instruction,[41] and (3) agree to meet data collection requirements for the evaluation. Schools also had to be located in states that administered Grade 3 state testing of students in spring 2012.[42] As the selection process was conducted, further screening occurred, as described below, to arrive at the final sample of experienced schools for the impact analysis.

The study schools that are included in the impact sample constitute a purposive rather than a representative sample. Nevertheless, the way that these schools implemented RtI features varied, allowing this sample to represent various types of RtI implementation models actually used by schools. For example, although all eligible schools in the sample used screening test scores to determine students' tier placement, the measures employed and the extent to which schools abided by the rating system (or decision rule) could have differed.

Process for Selecting the Impact Sample

All schools in the impact sample were selected through a multistage procedure over the course of two years. Figure 2.1 illustrates the selection process based on the eligibility requirements stated above, and the details of it are summarized below, by stage, with Stage 1 corresponding to the top of the figure.[43]

- **Stage 1: Nomination by Experts.** Beginning in fall 2010, the study research team sought school or school district nominations by contacting RtI researchers; relevant professional organizations (including the National Association of State Directors of Special Education); informants from national, regional,

[41]Schools with fewer than 30 students in any one of Grades 1, 2, and 3 and/or with fewer than 8 students in either Tier 1 or Tiers 2 and 3 combined were excluded from the core sample because there were an insufficient number of students to conduct the impact analysis.

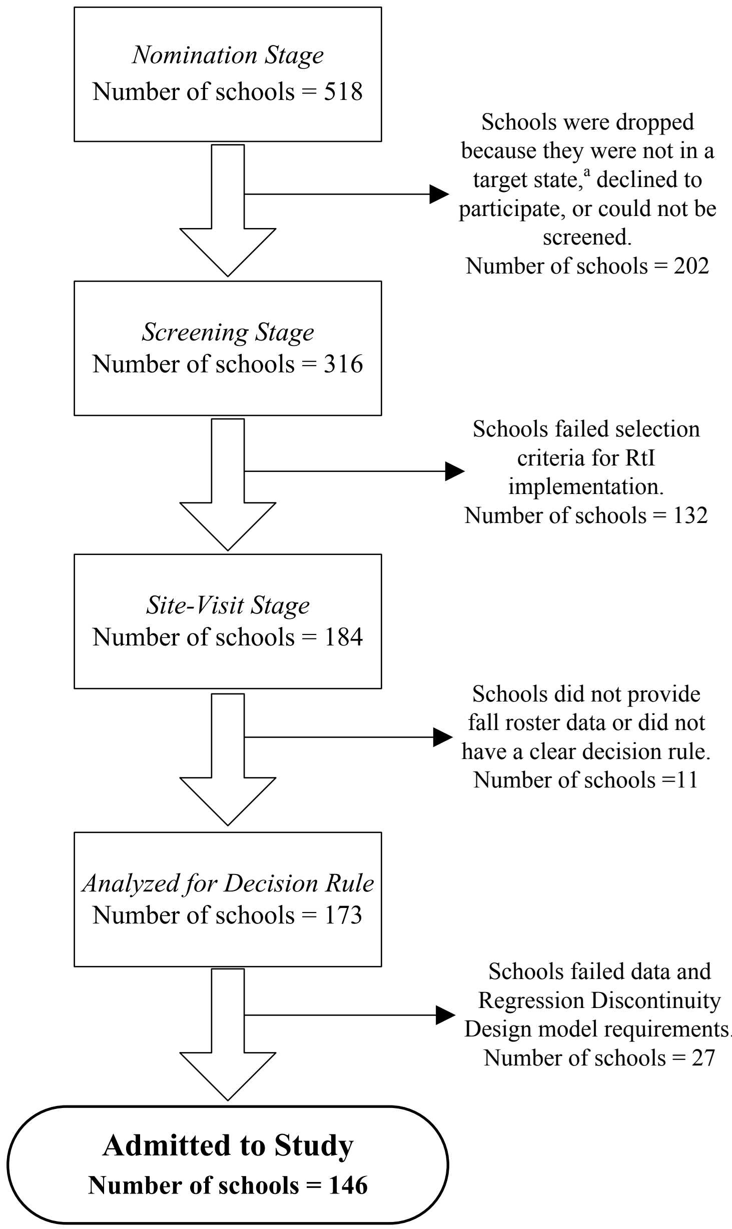
[42]These requirements ensured that the study research team had data to construct both rating and outcome variables for the impact analysis, described at the end of this chapter and in more detail in Chapter 5.

[43]Appendix B provides detailed information about the school selection process.

The Response to Intervention (RtI) Evaluation

Figure 2.1

School Selection Process for the Impact Sample



SOURCE: School selection process data.

NOTE: [a]The study target states are Arizona, California, Florida, Illinois, Massachusetts, Minnesota, Missouri, Montana, Pennsylvania, Texas, Utah, Washington, and Wyoming.

state, and local education agencies; and other RtI experts (for example, training providers). Through this effort, approximately 160 individuals nominated a total of 518 schools from 25 states.

- **Stage 2: Screening for Key Features of RtI.** To conserve project resources, from this stage on, the study research team focused on recruitment in 13 target states that met initial state-level criteria:[44] the state administered its third-grade reading test in the spring, which provided an outcome measure; there was more than one nominated school in a state, to protect anonymity of each school in the data; and the state provided geographic diversity to the sample. Among schools nominated in these states, the study research team conducted a series of screening activities to identify schools that met the school-level eligibility criteria for inclusion in the study. (See "Defining the Impact Sample," above.) When the study research team contacted schools, it asked about the four key RtI practices listed and the year in which the schools began implementing RtI in Grade 1. After this screening, 316 schools met the initial eligibility criteria for RtI implementation. Subsequently, these schools were asked whether they used screening test scores to identify students who required a reading intervention, the names of the assessments used, and whether they used a specific cut point or threshold score to assign students to receive an intervention.[45]

- **Stage 3: Site Visits.** The screening and site visit process described above resulted in 184 schools meeting the eligibility criteria and demonstrating interest in the study, including participating in a site visit from members of the study research team. The goals of the site visits were to confirm a school's use of each of the practices included in the initial eligibility requirements mentioned above and to gain a clearer understanding of the school's process to determine which students were assigned to receive intervention services. In addition, the study research team arranged a process to obtain fall screening test scores and tier placements for all students in Grades 1 to 3.

- **Stage 4: Analysis for Decision Rule.** Upon receiving the screening test scores and tier placements for each student in a visited school (as well as the cut point for assigning students in that grade to receive intervention services), the study research team used these data to verify whether the school truly

[44]The states are Arizona, California, Florida, Illinois, Massachusetts, Minnesota, Missouri, Montana, Pennsylvania, Texas, Utah, Washington, and Wyoming.

[45]In some cases, schools had a different cut point or threshold score for assigning students to intervention than what the assessment system recommended.

used one or more continuous measures (such as test scores) and some predetermined criteria based on these measures to assign a large proportion of individual students into different tier placements. This analysis was carried out separately for Grades 1 to 3 in these candidate schools, and schools with at least one grade that satisfied the above requirement were included in the impact sample for analysis.

Out of the 184 schools that the study research team visited, 173 schools were able to submit their fall screening test score results. After the verification process, the study research team found that 27 schools did not consistently use a quantitative decision rule that could be identified. The remaining 146 schools demonstrated a quantifiable system to determine whether or not a student was in need of more intense reading instructions for at least one grade.[46] These 146 elementary schools form the impact sample of schools for the study.

Reference Sample Selection

The evaluation also includes a reference sample of schools in each of the 13 study states. A random sample of 100 schools in each state was selected to represent all public elementary schools, charter schools, and magnet schools serving students in Grades 1 to 3 (other than the schools in the impact sample).[47] The study research team then fielded the same school-level survey as was used for the impact sample in these schools. Of the 1,300 schools sampled, 1,105 completed the survey, for a response rate of 85 percent. These schools serve as the reference sample for the impact sample of experienced RtI schools.

Description of Study Design and Key Data Sources

This section is organized by the three sets of research questions listed in Chapter 1. It presents an overview of the analytic approach used to address each set of questions, including the sample, the analytical method, and data sources.[48] Chapters 3 to 5 focus on these sets of questions, respectively, and provide details about the analytic approaches. The section below describes the design, methods, and data sources used in each chapter. Table 2.1 summarizes this information for the 13 states in the study, and appendixes elaborate on the data and methods.

---

[46]The specific procedure used for the selection is described in Appendix B. Note that this exercise means that the impact sample was selected based partially on the degree to which the schools adhered to a quantifiable tier assignment process and the enrollment in each grade. As a result, this sample should not be considered a random sample of all the nominated and/or screened schools.

[47]A simple random sampling procedure was used for each state. See Appendix B.

[48]Appendix Table B.1 presents details about the collection, response rates, and timing of collection for each data item, by level of data.

The Response to Intervention (RtI) Evaluation

Table 2.1

Research Questions, Samples, Study Design, and Data Sources for School Year 2011-12 in the 13 Study States[a]

Research Question	Sample	Study Design and Methods	Data Sources
Comparison of practices between school samples: How did the prevalence of RtI practices differ between a representative "reference" sample of schools and schools selected for the impact evaluation? To what extent were impact sample schools implementing more RtI practices than the reference sample schools? *Comparison with the states:* How do special education identification rates in the impact sample compare with the rates for the states as a whole?	• Impact evaluation sample of 145 experienced RtI schools that have Grades 1-3. • Reference sample of 100 schools randomly selected in each of the same states.	• Describe, compare key characteristics of the two samples. • Describe, compare the two samples on adoption and implementation of key RtI practices. • Compare special education counts in the impact sample with their respective state averages. • *Method:* Logistic regression analysis using survey weights. • *Outcomes:* Survey items as practice indicators.	• School surveys for fall 2011 • School-level 2010-11 reading proficiency data for Grade 3 from states • School characteristics data from Common Core of Data, 2010-11 • Grade-level special education identification data from impact sample schools and Office of Special Education Programs for average state values
Comparison of reading services between reading groups at different skill levels: To what extent did schools place students in tiers as suggested by earlier RtI models? To what extent did schools adjust tier placement during the school year? To what extent is there variation in how schools organize reading services for specific reading levels? To what extent were services for students reading below grade level more intense than for students reading at or above grade level?	• Impact evaluation sample of up to 131 experienced RtI schools with relevant data. • School sample varies by data source and grade.	• Describe distribution of students by reading level and movement between levels/tiers. • Categorize schools by organization of reading intervention services. • Describe, compare instruction and intervention services for reading groups by level/tier. • *Method:* Linear regression with school fixed effects and clustering of reading groups within schools, by grade. • *Outcomes, by grade:* Survey items as intensity factors; tier placements.	• Student tier placement data from schools for fall 2011 and winter 2012 • Teacher surveys, including description of reading groups, spring 2012 • Interventionist surveys, including description of reading groups, spring 2012 (continued)

Table 2.1 (continued)

Research Question	Sample	Study Design and Methods	Data Sources
***Impacts on reading outcomes of students*:** For students who fell just below school-determined standards for each grade on screening tests: What were the effects on reading achievement of actual assignment to receive reading intervention services (in addition to core instruction)? What is the extent of variation in estimated impacts across RtI schools? How is the estimated impact associated with certain school features or student characteristics?	• Grade 1-3 students in the impact evaluation sample schools with necessary data for the analysis. • School and student sample varies by grade.	• Define treatment group as those assigned to Tier 2 or Tier 3 in fall 2011 to receive intervention services (in addition to Tier 1 core instruction). Comparison group is defined as those assigned to Tier 1 to receive core instruction only. • Compare outcomes for students just below the cut point (the treatment group) with those for students just above the cut point (the comparison group). • *Design:* Regression Discontinuity (RD) design; tier placements for much of the sample in selected schools are based on a predetermined, quantifiable combination of fall benchmark test scores. • *Method:* Two-stage least squares estimation of Local Average Treatment Effect (LATE), by grade. • *Outcomes:* Spring 2012 test scores.	• Fall 2011 student screening test scores and tier placements • ECLS-K[b] Reading Assessment score and TOWRE2[c] reading test score for Grade 1 students (study-administered in spring 2012) • TOWRE2 test score for Grade 2 students (study-administered in spring 2012) • State reading achievement test score for Grade 3 students (school-administered in spring 2012) • Student demographic characteristics from schools or districts for fall 2011

NOTES: Analyses for Chapters 4 and 5 do not distinguish Tier 2 and Tier 3 students (or groups serving students reading Somewhat Below or Far Below grade level). This is due in part to the small sample size of Tier 3 students and groups, and to maximize contrast and to align with the research questions.

[a]The study states are Arizona, California, Florida, Illinois, Massachusetts, Minnesota, Missouri, Montana, Pennsylvania, Texas, Utah, Washington, and Wyoming.

[b]ECLS-K Reading Assessment: Early Childhood Longitudinal Survey, Kindergarten Cohort, of 2011.

[c]TOWRE2: Test of Sight Word Reading Efficiency - Second Edition.

Comparison of Practices Between School Samples

Research questions: *How did the prevalence of RtI practices differ between a representative "reference" sample of schools and schools selected for the impact evaluation? To what extent were impact sample schools implementing more RtI practices than the reference sample schools? How do special education identification rates in the impact sample compare with the rates for the states as a whole? (Chapter 3)*

To answer this set of questions, the analysis compares the impact sample schools and the reference sample schools using two main data sources. First, school demographic characteristics for both samples were collected from the National Center for Education Statistics' Common Core of Data Public School Universe file for 2010-11.[49] The school characteristics are summarized and contrasted to provide contextual information about these two samples.

Second, a school-level survey was administered in all schools to collect information on the whole school's reading practice, focusing on instruction, progress monitoring, and use of data for instructional decision-making for students with different levels of reading needs. The survey was administered to schools in the impact sample in spring 2012 and to schools in the reference sample in fall 2012.[50] Both samples were asked to describe practices during the 2011-12 school year. Survey results were used to describe the extent to which both samples of schools adopt and implement key RtI practices.

The analysis uses survey items as the school-level practice outcomes. Schools in the reference sample were weighted with a sampling weight, to represent that they were randomly sampled from each state. Regression models were used to estimate differences in a school-level practice between the two school samples.

Third, this chapter compares the impact sample schools' rates of identifying students for special education with their respective state averages for age groups relevant to Grades 1 to 3. School-by-age counts for fall 2011 were collected from schools or districts. State-level data were downloaded from an Office of Special Education Programs website.[51] Count data are described in Appendix B.

[49]Data are available from http://nces.ed.gov/ccd/search.asp. Data for 2011-12 were not available at the time of analysis.

[50]Of the 146 impact sample schools, 145 responded. Of the 1,300 reference sample schools, 1,105 responded.

[51]Collection of child count data is authorized under IDEA Section 618, Part B. These are available by year, as of May 2015, at: http://www2.ed.gov/programs/osepidea/618-data/state-level-data-files/index.html.

Comparison of Reading Services Between Reading Groups at Different Skill Levels

Research questions: *In impact sample schools (those with three or more years of implementing RtI): To what extent did schools place students in tiers as suggested by earlier RtI models? To what extent did schools adjust tier placement during the school year? To what extent is there variation in how schools organize reading services for specific reading levels? To what extent were services for students reading below grade level more intense than for students reading at or above grade level? (Chapter 4)*

The placement and movement of students between tiers permitted schools to adjust service intensity and characteristics for students at different reading levels. This student-level analysis consists of simple averages of winter tier placement conditional on a student's fall tier placement. These show the proportion of students who remained assigned to the same tier or who moved to a more intense or a less intense tier. The school sample varies by grade and is discussed in Appendix B.

Analysis for the reading-services questions focuses on detailed reading-group data. During data collection, the study research team requested that all classroom teachers and interventionists serving students in Grades 1, 2, and 3 complete surveys in spring 2012.[52] (In this study, an adult who was identified by the school as providing intervention services is called an "interventionist.")[53] Respondents provided information on aspects of reading services that were provided to student groups at different reading levels in spring 2012. Teacher survey responses describe small-group instruction, and interventionist survey responses describe the group services that interventionists deliver. Each respondent described services to one or more small reading groups, limited to six groups for teachers and ten groups for interventionists.

For specific reading-group services, the school sample varies by grade and is described in more detail in Appendix C. The sample for describing reading-group services is based on whether respondents completed the survey variable concerning the relevant dimension, as well as having completed the question regarding the reading level and primary grade served.

The study research team examined the difference in services between groups reading at or above and below grade level for two types of reading supports: small-group instruction ser-

[52]Appendix B describes data collection procedures and response rates in detail. Of the 146 schools in the impact sample, all schools completed teacher surveys, and 141 (96.6 percent) completed relevant items on the interventionist survey.

[53]Throughout this report, the term "interventionist" refers to an instructor providing reading interventions to students in need of them, regardless of staff title. In cases where classroom teachers also provided reading interventions, they were asked to complete both the teacher survey and the interventionist survey. Chapter 4 describes the various staff roles that serve as interventionists.

vices during the core reading period, as reported by teachers, and intervention-group services, as reported by interventionists. Regression models were estimated to analyze differences in practices between groups serving students reading at or above grade level and those reading below grade level in the same school and grade. Appendix C describes additional details about the analysis methods.

By documenting whether schools were moving students between tiers based on student performance on screening assessments and showing differences in the intensity of services between reading groups at different levels, these results build on evidence presented in Chapter 3 about how schools are implementing RtI.[54]

Impacts on Reading Outcomes of Students

> ***Research question:*** *For students who fell just below school-determined standards for each grade on screening tests: What were the effects on reading achievement of actual assignment to receive reading intervention services (in addition to core instruction)? (Chapter 5)*

The way that schools in the impact sample identified at-risk students for reading intervention provided a unique opportunity to use the Regression Discontinuity (RD) design. These experienced RtI schools used a universal screening test at the beginning of the fall semester to identify students for additional reading support. Students whose screening test scores fell below a predetermined cutoff point were deemed at risk of reading failure and were usually assigned to receive reading interventions (the "treatment group").[55] Those students whose screening test scores were above the cutoff usually received only core reading instruction (the "comparison group"). The Chapter 5 impact analysis defines assignment to receive Tier 2 or Tier 3 intervention services as the treatment condition. This way of identifying at-risk students allowed the use of the RD design to compare reading achievement outcomes between two sets of students: (1) students who, based on their screening test scores, just missed reading benchmarks and qualified to receive reading intervention support and (2) students in the same school who just met reading benchmarks and thus were not identified for intervention in reading. Essentially, by statistically controlling for the value of the screening test score in a regression model, one can (under appropriate conditions) account for any unobserved differences between the treatment group and the comparison group and, thereby, can obtain internally valid impact estimates for receiving more intense, additional reading support.[56] Chapter 5 discusses the conditions needed for

[54]While impact sample schools were screened and selected on the basis of implementing key school-level practices, they were not selected based on the level of differentiation of services for students reading at different levels.

[55]More information is provided in Chapter 5 on the percentage of students who were served based on the cut point and the extent to which this analysis constitutes a "fuzzy" rather than a sharp RD design.

[56]See Lee and Lemieux (2010) and Bloom (2012) for reviews of the literature on the RD design.

this analysis, describes that the conditions were met, and provides guidance on how to interpret the resulting impact estimates.

The study research team conducted the impact analysis at the individual student level, separately for each reading outcome in Grades 1 to 3. The recruitment process identified schools that met the selection criteria for RD analysis: 119 schools for Grade 1; 127 schools for Grade 2; and 112 schools for Grade 3. The analysis sample, therefore, consists of all the students who had required data in the relevant grades in the 146 schools in the impact sample.

The RD analysis for this study required several key data components for each student: the fall screening test score, which is the basis of students' assignment to reading interventions;[57] the determination of whether a student was assigned to receive reading intervention services in fall 2011; the spring 2012 reading outcome measure; and information on the characteristics of the student.

Different reading outcome measures were used for each of the three targeted grade levels. For Grade 1, the 2011 Early Childhood Longitudinal Survey-Kindergarten Cohort of 2011 (ECLS-K:2011)[58] first-grade reading assessment provided a comprehensive measure of students' overall reading achievement. For both Grade 1 and Grade 2, the Test of Sight Word Reading Efficiency-Second Edition (TOWRE2) was used to gauge students' decoding fluency.[59] Reading performance for third-graders was measured by their scores on the state reading achievement tests collected from extant district administrative data.

In addition to the primary impact question discussed above, the study research team examined two types of exploratory questions (below). Both questions seek to explain the pattern and generate hypotheses about possible mechanisms underlying the results of the primary impact analysis. The methods used are described in more detail in Chapter 6 and Appendix H, and they build on methods described in Chapter 5:

- *What is the extent of variation in estimated impacts across RtI schools?* This analysis assesses the variation in estimated impacts by school, for each grade, and tests whether variation in estimated impacts differs significantly across sites.

- *How is the estimated impact associated with certain school features or student characteristics?* School characteristics and practices from the descriptive analyses in Chapters 3 and 4 serve as explanatory variables of the variation in

[57]In the RD design literature, this "assignment variable" is a continuous variable measured before treatment, the value of which determines whether or not a group or individual is assigned to the treatment. It is also called the "rating variable," "forcing variable," or "running variable" (Jacob, Zhu, Somers, and Bloom, 2012).
[58]Website: http://nces.ed.gov/ecls/kinderinstruments.asp.
[59]Torgesen, Wagner, and Rashotte (1999).

estimated impacts across schools. In addition, the chapter explores the relationship between student characteristics and the estimated impact for each grade.

Chapter 3

Comparison of Schools in the Impact and Reference Samples

Prior research that is cited in Chapter 1 documents the adoption of the Response to Intervention (RtI) framework across school districts. This chapter describes the prevalence of specific RtI practices in the 13 states included in this study and describes these practices at the school level for both the reference sample and the impact sample described in Chapter 2.

Specifically, the chapter addresses two aspects of the "prevalence" question presented in Chapter 1:

1. How did the prevalence of RtI practices differ between a representative "reference" sample of schools and schools selected for the impact evaluation (the impact sample)? To what extent were impact sample schools implementing more RtI practices than the reference sample?

2. How do the special education identification rates in the impact sample compare with the rates for the states as a whole?

The hypotheses examined in this chapter are that at least some RtI practices were present in a majority of elementary schools in the states in the study and that a larger percentage of the impact sample schools — screened to have three years' experience — than of reference sample schools implemented key RtI practices, as described in Chapter 1.

Major findings include:

- The majority of reference sample schools adopted an RtI framework, including implementing multiple reading tiers and using data-based decision-making practices.

- A larger percentage of impact sample schools than of reference sample schools implemented an RtI framework for reading.

- The majority of both samples of schools used a combination of eight key RtI practices, and the prevalence of six of these practices was greater in the impact sample.

The next section of this chapter compares the characteristics of schools in the reference sample with the corresponding characteristics of schools in the impact sample. The chapter then describes and compares the implementation of key RtI practices in the two samples, focusing on those differences that are statistically significant.[60] Then the chapter discusses patterns of special education identification for the schools in the impact sample and how these rates compare with rates for all schools in the same states. Practice data came from the responses of school administrators to a survey about practices in the 2011-12 school year.[61] School characteristics and special education identification rates came from extant data.

Description of School Characteristics in the Impact and Reference Samples

This section compares demographic and achievement characteristics of the impact and reference samples, which include public, charter, and magnet schools serving students in Grades 1 to 3. Table 3.1 shows the average characteristics of reference sample schools and of impact sample schools, based on data from the U.S. Department of Education's Common Core of Data.[62] The two rightmost columns present the differences between these samples and give their associated significance levels.[63]

The impact sample was *not* selected to be a representative sample of the state. As such, the impact and reference samples differ significantly on several characteristics. The percentage of white students in the impact sample is 13 percentage points higher (61 percent, compared with 48 percent), while its percentage of Hispanic students is 12 percentage points lower (22 percent and 34 percent). The percentage of rural schools is lower in the impact sample than in the reference sample (15 percent and 26 percent). Compared with the reference sample, the impact sample has a lower percentage of students who were low-income[64] (42 percent and 52 percent), a lower proportion of Title I-eligible schools (70 percent and 81 percent), and a higher percentage of students identified for special education with an Individualized Education Program (13 percent and 12 percent). The percentage of charter and magnet schools in the reference sample is higher (7 percent and zero percent), because such schools were included in the sampling frame for the reference sample but were not selected for the impact sample.

---

[60]These comparisons are of the two samples as a whole, rather than comparing reference sample schools with impact sample schools in each state separately. Percentages discussed in the text may round up from the tables in some cases.

[61]Administrators could include the school principal, school psychologist, or other school leaders familiar with the RtI program. Multiple individuals could complete a survey for a school, but there is only one survey data point per school.

[62]The impact sample used in this chapter's analysis is 145 — one school fewer than the full impact sample — because one school did not complete the school survey.

[63]The comparison analysis uses survey sampling weights, to reflect that some states in the study are larger than others. Details about the analysis are discussed in Appendix C.

[64]"Low income" is measured as the proportion of students eligible for free or reduced-price lunch.

The Response to Intervention (RtI) Evaluation

Table 3.1

Characteristics of the Impact Analysis School Sample and the Reference School Sample Serving Grades 1-3

Characteristic	Impact Sample Mean	Reference Sample Mean	Mean Difference Between Samples	P-Value
Race/ethnicity[a] (%)				
Asian	5.5	5.3	0.2	0.768
Black	7.4	9.7	-2.2	0.065
White	60.9	47.8	13.1	0.000
Hispanic	21.8	33.7	-11.9	0.000
Other	4.3	3.6	0.8	0.046
Sex[a] (%)				
Male	51.2	51.2	0.0	0.988
Locale (%)				
Urban	31.9	29.7	2.2	0.612
Suburban	40.3	35.2	5.1	0.258
Town	13.2	9.6	3.6	0.239
Rural	14.6	25.5	-10.9	0.001
Low-income status[b] (%)	42.0	51.8	-9.8	0.000
Title I eligible schools[c] (%)	69.4	81.1	-11.6	0.005
Charter and magnet schools[c] (%)	0.0	6.7	-6.7	0.000
Average school size (number of students in all grades)	509	502	7	0.652
Average school size (number of students in Grades 1-3)	240	233	6	0.412
Number of full-time-equivalent staff	29	30	-1	0.529
English Language Learners[d] (%)	7.8	8.1	-0.3	0.690
Individualized Education Program[d] (%)	13.4	12.2	1.1	0.000
Deviation from state mean of percentage proficient on Grade 3 standardized state reading test	2.8	1.0	1.8	0.194
Number of schools	144	1,105		

(continued)

Table 3.1 (continued)

SOURCES: Common Core of Data, including The U.S. Department of Education, National Center for Education Statistics, Public Elementary/Secondary School Universe Survey School Years 2010-11 and 2009-10, and the Local Education Agency (School District) Universe Survey 2010-11. State achievement data downloaded from 13 state websites.

NOTES: The number of schools in the impact sample is 144 rather than 146, which reflects that one school did not complete the survey and one school was combined in the 2010-11 CCD data but subsequently split into two buildings during the 2011-12 year, when survey data were collected. Some schools did not have 2010 data available. In these cases, 2009 data are used, where available, for variables missing in the 2010 data.

[a]Race/ethnicity and sex calculations are based on school-level student populations in Grades 1 through 3.

[b]Low-income status indicates school data on proportion of students receiving free and reduced-price lunch.

[c]Title I and charter and magnet status variables take on values of zero or 100. The means represent the percentage of schools of each type in each sample.

[d]English Language Learners (ELL) and Individualized Education Program (IEP) data come from district-level data and thus are based on district-level student populations.

The reference and impact samples do not differ significantly on the proportion of students who are male or were English Language Learners. Nor do the samples differ in terms of school size, the number of full-time-equivalent staff, or deviation from the state's mean percentage of students proficient in Grade 3 reading during 2010-11.[65]

Findings on the Prevalence of Key RtI Practices

School administrators were asked to what extent they were implementing RtI, which the survey defined as a "multi-step approach to providing early and progressively intensive intervention and monitoring within the general education setting."[66] Respondents could answer that RtI was "fully implemented," "partially implemented," or "not implemented" in reading for each grade between Kindergarten and Grade 5. For Grade 1, respondents also could have indicated RtI implementation in other subjects.[67]

[65]Reading proficiency is based on state achievement reading test scores for Grade 3 for spring 2011, measured as the percentage of students in each school who were at or above proficiency and expressed as deviations from the state's mean percentage.

[66]The online appendix includes the survey instrument.

[67]Survey response rates are high for both samples of schools: 99 percent of impact sample schools and 85 percent of reference sample schools across 13 states. When looking at response rates for individual survey items, there were no clear patterns of skipping certain items among either the reference or the impact sample schools, so it is unlikely that results reported in this chapter are biased by response patterns.

- **The majority of both impact sample and reference sample schools reported fully implementing an RtI framework for reading in Grade 1.**

As would be expected, Figure 3.1 shows that a higher percentage of schools in the impact sample than in the reference sample reported fully implementing RtI in Grade 1 reading (91 percent and 71 percent, respectively).[68] Consistent with previous research,[69] schools focused their RtI implementation efforts in the area of early reading, with relatively fewer of them — about or less than one-third — reporting fully implementing an RtI framework in other subject areas in Grade 1. For both samples, a similar percentage of schools implemented RtI in math, writing, and behavior/social skills in Grade 1.

- **Across Grades 1 to 3, a larger percentage of impact sample schools than of reference sample schools reported full implementation of an RtI framework for reading.**

The pattern described above for Grade 1 is more pronounced for Grades 1 to 3: 86 percent of impact sample schools reported full implementation of RtI in reading, compared with 56 percent of reference sample schools, as shown in Figure 3.2.

To describe the RtI framework, the study research team defined key RtI practices for elementary school reading, based on the following rationale:

- Practice guidelines that experts recommend,[70] such as administering universal screening at least twice a year, delivering Tier 2 intervention at least three times a week, and providing Tier 3 intervention four to five times a week

- Practices critical to all RtI models, such as the use of data to make instruction and intervention decisions, to target interventions, to monitor progress, and to contribute to special education eligibility decisions[71]

- Features that support RtI implementation, such as staff supports[72]

- School's perception of the extent of RtI implementation in Grades 1 to 3

[68]This survey question followed from a skip question, which asked "Is RtI currently used in at least one grade at your school, either partially or fully implemented?" Among those eligible to respond to the Grade 1 reading implementation question (those who answered "yes" to the previous question), reference sample schools had a 90 percent response rate and impact sample schools had a 97 percent response rate. The difference in missing rates between the samples is significant, with a p-value of 0.007.

[69]Bradley et al. (2011).

[70]Gersten et al. (2009).

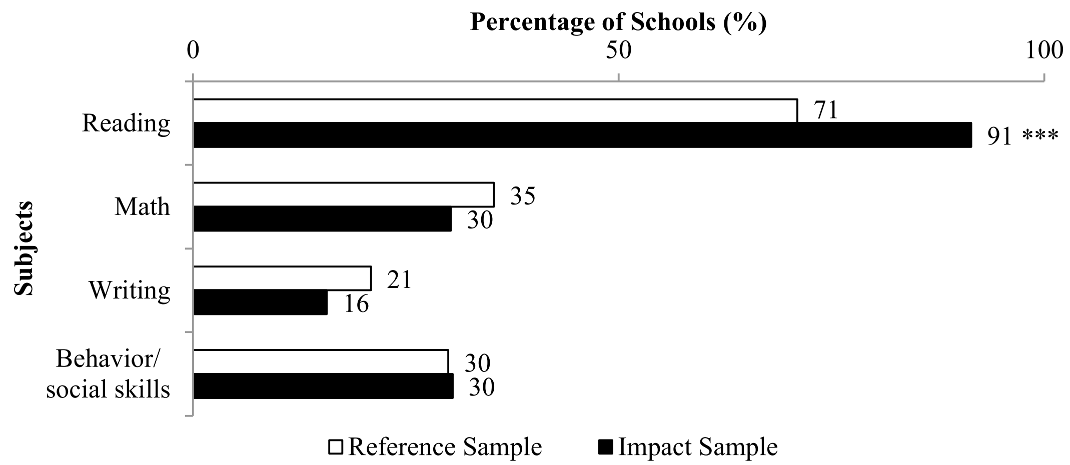
[71]Burns and Ysseldyke (2005); Fletcher, Coulter, Reschly, and Vaughn (2004); Fuchs and Fuchs (2002); Gersten et al. (2009); VanDerHeyden, Witt, and Barnett (2005).

[72]Borman, Hewes, Overman, and Brown (2003); Glennan, Bodilly, Galegher, and Kerr (2004).

The Response to Intervention (RtI) Evaluation

Figure 3.1

Implementation: Percentage of Schools That Have Fully Implemented RtI Practices in Grade 1, by Subject



SOURCE: School survey.

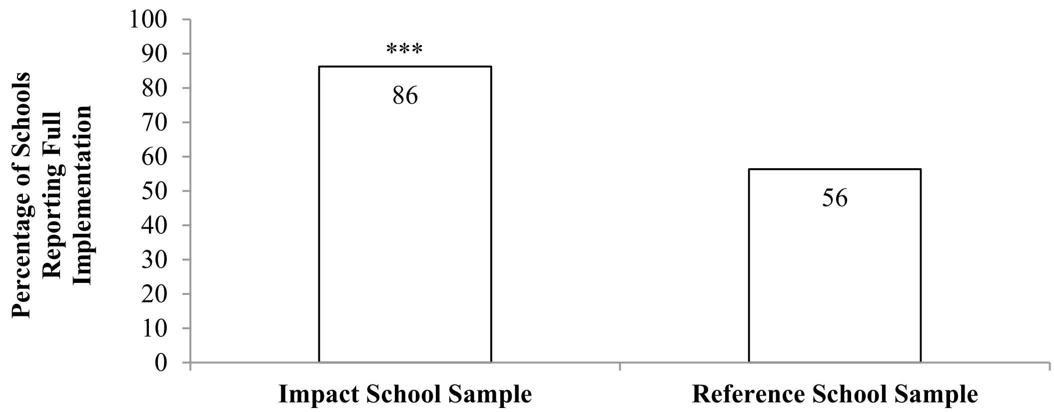
NOTES: Means and percentages are conditional on having responded to the item presented and also conditional on having responded “Yes” to “Is RtI currently used in at least one grade at your school, either partially or fully implemented?” 986 schools of the reference sample answered yes, while 144 schools of the impact sample answered yes.

Among eligible respondents, the missing rate for the item about Reading among the reference sample is 10.4 percent and among the impact sample is 2.8 percent, and the difference in missing rates between the samples has a p-value of 0.0001. The missing rate for the item about Math among the reference sample is 5.0 percent and among the impact sample is 1.4 percent, and the difference in missing rates between the samples has a p-value of 0.09. The missing rate for the item about Writing among the reference sample is 6.1 percent and among the impact sample is 2.8 percent, and the difference in missing rates between the samples has a p-value of 0.2057. The missing rate for the item about Behavior/Social Skills is 5.5 among the reference sample and 2.1 among the impact sample, and the difference in missing rates between the samples has a p-value of 0.127. The percentage of missing respondents for the items for Math, Writing and Behavior is less than 3 percent each and not significantly different between samples. Means reflect rounding.

Statistical significance is indicated as follows: *** at the $p \leq 0.001$ level, ** at the $p \leq 0.01$ level, and * at the $p \leq 0.05$ level.

Figure 3.2

Full Implementation of RtI in Reading in Grades 1-3



SOURCE: School survey.

NOTES: The numbers of responding schools are 145 for the impact sample and 1,105 for the reference sample. The survey defined RtI as a "multistep approach to providing early and progressively intensive intervention and monitoring within the general education setting." Respondent schools could answer that RtI was "fully implemented," "partially implemented," or "not implemented" in reading for each grade between Kindergarten and Grade 5.

Schools were considered "fully implemented" if they completed at least one of the three items (one each for Grades 1, 2, and 3) and responded "fully implemented" to all of the items that they did complete.

Percentages reflect rounding. The statistical significance is indicated as follows: *** at the $p \leq 0.001$ level, ** at the $p \leq 0.01$ level, and * at the $p \leq 0.05$ level.

These definitions resulted in eight key practices that could be measured in the school survey and that correspond to the three components of the RtI framework shown in the logic model (Chapter 1, Figure 1.3):

1. For use of **multiple tiers of reading instruction and intervention,** the practices are (a) provided more than 90 minutes of instruction in the daily core reading block in Grades 1 to 3; (b) for each child receiving Tier 2 reading intervention services, provided Tier 2 intervention at least three times a week; (c) for each child receiving Tier 3 reading intervention services, provided Tier 3 intervention at least five times a week.

2. For **allocation of staff,** the practices are (d) allocated staff to assist teachers with using data and (e) allocated staff to assist teachers with reading instruction.

3. For **use of data to inform decisions,** the practices are (f) conducted universal screening at least twice a year in Grades 1 and 3; (g) followed a prescribed sequence of steps to make data-based decisions; and (h) used data to monitor student progress following reading intervention and to inform the determination of special education eligibility.

Table 3.2 reports the percentage of schools in the two samples that implemented each of the practices as well as an average total number of practices in place.[73] In Grades 1 to 3, impact sample schools implemented 6.5 of these eight practices, on average, and reference sample schools implemented 5.4 of these eight practices, on average; the differences are not statistically significant. However, as shown in the rest of the table, the prevalence of individual practices is significantly different for six practices.

Multiple Tiers of Reading Instruction and Intervention. Although about two-thirds (68 percent to 70 percent) of both school samples reported offering 90 minutes per day of core reading instruction, the frequency of offering intervention differed between the two samples. Impact sample schools were more likely to report providing time for Tier 2 intervention at least three times a week than were reference sample schools (97 percent and 80 percent). Impact sample schools were also more likely to report providing time for Tier 3 intervention at least five times a week than were reference sample schools (68 percent and 47 percent).

Allocation of Staff. Compared with reference sample schools, impact sample schools were more likely to allocate staff to assist teachers with using data (88 percent and 72 percent) and with reading instruction (69 percent and 56 percent).

Use of Data to Inform Decisions. Eighty-three percent of impact sample schools conducted universal screening assessments of Grade 1 and Grade 3 students at least twice a year, compared with 59 percent of reference sample schools. Impact sample schools were also more likely than reference sample schools to follow a prescribed sequence of steps to place students in Tier 2 or Tier 3 interventions (95 percent and 88 percent). Impact sample and reference sample schools were not significantly different in their use of data to monitor student progress following implementation of reading interventions to determine whether intervention was working.

---

[73]The variables are binary indicators of whether a practice was in place or not. When the survey items asked about practices for different grades, the measure draws on responses for Grades 1 and 3, to align with the grades included in the impact analysis.

The Response to Intervention (RtI) Evaluation

Table 3.2

**Summary of Key RtI Practices:
Percentage of Schools Reporting Implementation of Each Practice**

School Administrators Reported	Percentage of Schools: Impact	Percentage of Schools: Reference	Mean Difference Between Samples	P-Value
Summary				
Average number of 8 key practices in place (a - h)	6.5	5.4	1.1	0.058
Multiple Tiers				
a. Provided more than 90 minutes in daily core reading block (Grades 1-3)	69.7	68.2	1.5	0.822
b. Provided Tier 2 intervention at least 3 times a week	96.6	80.0	16.5	0.000
c. Provided Tier 3 intervention at least 5 times a week	67.6	47.1	20.5	0.000
Staffing and Resources				
d. Allocated staff to assist teachers with using data	88.3	72.1	16.2	0.000
e. Allocated staff to assist teachers with reading instruction	69.0	55.5	13.5	0.005
Data-Based Decision Making				
f. Conducted universal screening at least 2 times a year in Grades 1 and 3	83.4	58.9	24.5	0.000
g. Followed a prescribed sequence of steps for responding to students who are below benchmark in reading	94.5	87.9	6.6	0.042
h. Used data to progress-monitor students following implementation of reading interventions, as part of determining special education eligibility	80.0	72.6	7.4	0.092

SOURCE: School survey; wording of items is listed below.

NOTES: For all items, the number of responding schools is 145 for the impact sample and 1,105 for the reference sample. Four reference schools and zero impact schools reported implementing none of the eight RtI practices. The percentages reported in this table reflect the number out of the sample total that affirmatively responded that they implemented the defined practice.

(continued)

Table 3.2 (continued)

For the purposes of summing observations, missing or skipped practices are interpreted as zero values. As a result, the means reported in this table may differ from those reported for individual corresponding items in other tables, because schools that did not answer certain items have a value of zero for each practice but are excluded from other tables. Sampling weights were applied to each reference school that responded to the survey.

P-values for the average number of practices in place were estimated from a linear regression with number as the outcome.

P-values for individual practices were estimated from a logistic regression with treatment status as the outcome.

The eight practices are defined as follows:

a: "How many total minutes during the school day are allocated to the core reading block (for example, phonemic awareness, phonics, fluency, vocabulary, and reading comprehension, but excluding spelling, grammar, and writing) for students in Grades K-5?" Responses were given in time ranges, and the average of the high end of the time ranges for each of Grades 1, 2, and 3 must have been greater than 90.

b: "How many days per week do most students receive Tier 2 intervention(s)?" Response must have been 3, 4, or 5.

c: "How many days per week do most students receive Tier 3 intervention(s)?" Response must have been 5.

d: "In the 2011-12 school year, is there someone in the building whose role is to assist teachers in using and interpreting assessment data on reading?" Response must have been "yes."

e: "Is there someone in the school who provides coaching to classroom teachers on teaching reading?" Response must have been "yes."

f: "In what months are screening or benchmark measures administered to students in each grade?" Response must have indicated at least two nonconsecutive months of assessment in each of Grades 1 and 3.

g: "Does your school follow a prescribed sequence of steps for responding to students who are below benchmark in reading?" Response must have been "yes."

h: "In your school, which of the following kinds of data are used for informing special education eligibility determinations for students suspected of having a specific learning disability?" For "data and other information from systematic monitoring of student progress following implementation of reading interventions," response must have been "always used."

The next two sections of the chapter discuss these practices in additional detail. The RtI practices that were implemented by the impact sample set the context for a detailed examination of reading-group services in Chapter 4 and for the impact findings in Chapters 5 and 6.

Multiple Tiers of Reading Instruction and Intervention

As described in Chapter 1, a basic premise of RtI is serving students in multiple tiers of increasing intensity. *Tier 1* provides core reading instruction to all students, including differentiated instruction for all students based on assessments of their current reading levels. Recommended practice for *Tier 2* is a small reading group that meets three to five times a week for 20 to 40 minutes.[74] *Tier 3* is typically more individualized than Tier 2; recommended practice for students in Tier 3 is more intense daily intervention that promotes the development of the various components of reading proficiency.

The survey of school administrators asked them to report the amount of time that their schools allocated for core reading instruction provided to all students (Tier 1) and for reading interventions provided to students who needed more help (Tiers 2 and 3).

As shown in Table 3.3, among schools that implemented all three tiers and reported either full or partial implementation of RtI, differences are not statistically significant for any tier. For Tier 1 instruction for all students, impact sample schools reported allocating, on average, 510 minutes per week, while reference sample schools reported 493 minutes per week. This is approximately 90 to 100 minutes per day, which is consistent with scheduling a 90-minute reading block. Impact sample schools reported 196 minutes for Tier 2, compared with reference sample schools reporting 160 minutes per week. Time allocated for Tier 3 intervention was 243 minutes for the impact sample and 205 minutes for the reference sample, on average. These time allocations are consistent with the recommended standards described above.[75]

Within each sample, the amount of Tier 3 time per week was greater than the amount of Tier 2 time per week. However, the difference in minutes allocated for Tier 3 and Tier 2 for the impact sample does not differ significantly from the difference for the reference sample (Table 3.3, last row).

[74] Gersten et al. (2009); National Center on Response to Intervention (2010).

[75] As discussed in Chapter 4, some schools implementing RtI provided a portion of reading interventions inside the core reading time, so these Tier 2 and 3 averages may not always be addition to time in core reading instruction (Tier 1).

Table 3.3

Multiple Tiers: Average Minutes per Week of Allocated Instruction and Intervention, by Student Tier, for Grades 1-3 Among Schools That Implemented an RtI Model and Offered Intervention Tiers

Minutes, by Student Tier	Impact Sample Mean	Reference Sample Mean	Mean Difference Between Samples	P-Value
Average minutes per week[a]				
Instruction time allocated for all students in core	510	493	18	0.676
Intervention time allocated for students in Tier 2	196	160	36	0.148
Intervention time allocated for students in Tier 3	243	205	38	0.295
Difference in minutes allocated for Tier 3 and Tier 2	46	44	2	0.884

SOURCE: School survey.

NOTES: The maximum number of schools that responded to the average minutes per week item is 132 for the impact sample and 770 for the reference sample. These first were filtered to ensure that they either partially or fully implemented RtI reading in Grades 1-3. The number of schools for Tier 2 and Tier 3 represents the schools that responded that they have fully or partially implemented an RtI model and offered both Tiers 2 and 3. Among eligible respondents, the overall missing rate for each item was less than 3 percent.

[a]Minutes per week were calculated by multiplying days per week with minutes per day. For respondents who did not answer the days per week question but did answer the minutes per day question, data were coded assuming those schools met five days per week.

The minutes-per-day values were calculated by averaging minutes per day in Grades 1 through 3. If a school skipped minutes per day for Grade 1 and Grade 2 but answered for Grade 3, the school's response for Grade 3 was taken as the overall minutes per day.

Because of rounding, difference in minutes allocated may not exactly calculate from the values in the table.

Data-Based Decision-Making and Involvement of Staff

Within an RtI framework, implementing multiple tiers of reading instruction and intervention involves schools using assessments to (1) screen all students and target intervention support to those identified below grade-level benchmarks, (2) monitor struggling students' responsiveness to interventions, and (3), in some cases, inform decisions about determining eligibility for special education services. This section describes these practices in detail and then provides results.

- **Screen all students, and target intervention support.** Early identification of students at risk for long-term reading difficulties begins with systematic screening near the beginning of the school year and at least once again in the middle of the year.[76] Elements of a screening battery include standardization of screening procedures; grade-level benchmarks or expectations; designated risk levels; ease and efficiency of administration; and documented reliability, validity, and diagnostic accuracy of the screening measures. Such measures can be used to identify individual students or can be aggregated to examine the adequacy of the core curriculum as well as the effectiveness of different instructional strategies used in a school.[77]

- **Monitor struggling students' responsiveness to interventions.** A signature feature of RtI is frequent assessment of student performance on a valid and reliable progress monitoring measure (for example, oral reading fluency). Examined across time, these measures depict students' growth and signal whether their response to intervention is on track to reach a learning goal within a specified period of time.[78] Such progress monitoring in addition to periodic universal screening is seen as a way to more accurately identify students requiring intervention.[79] Recommended practice is to monitor progress of students in Tier 2 at least monthly to determine their response to intervention.[80] At least weekly progress monitoring is recommended for students in Tier 3.[81] Schools can choose from a range of commercial and free progress monitoring measures.

- **Inform decisions regarding eligibility for special education services.** As part of an RtI framework for academic instruction, when progress measures indicate insufficient response to interventions, schools are encouraged to intensify intervention and/or consider eligibility for special education services (for example, under the category of Specific Learning Disability).[82]

As reported in Chapter 2, schools in the impact sample needed to use universal screening data and progress monitoring data to be eligible for inclusion in that portion of the study. Nearly 94 percent of reference sample schools used these data as well.

---

[76]Gersten et al. (2009); National Center on Response to Intervention (2010).
[77]National Center on Response to Intervention (2010).
[78]Fuchs, Deno, and Mirkin (1984); Fuchs, Fuchs, Hamlett, and Stecker (1991).
[79]Compton et al. (2010); Fuchs and Fuchs (2002).
[80]Gersten et al. (2009); National Center on Response to Intervention (2010).
[81]National Center on Response to Intervention (2010).
[82]Fletcher, Coulter, Reschly, and Vaughn (2004); Torgesen (2009); VanDerHeyden, Witt, and Barnett (2005).

Among schools that administered universal screening and progress monitoring tests, the survey asked which staff members *analyzed* these data. In both samples, responsibility for analyzing screening and progress monitoring data was distributed across multiple personnel. As shown in Figure 3.3, more than 80 percent of schools in both samples reported that classroom teachers played a primary role in analyzing these data; the difference between the samples is not statistically significant.

- **Impact sample schools used a variety of staff — including classroom teachers — to analyze data.**

In general, the impact sample of schools reported using a wider variety of staff types to analyze data than the reference sample of schools (Figure 3.3). More schools in the impact sample than in the reference sample reported that specialists analyze universal screening data (78 percent and 60 percent) and progress monitoring data (75 percent and 58 percent). A significantly larger percentage of impact sample schools than of reference sample schools reported using the school psychologist to analyze universal screening data (31 percent and 12 percent) and progress monitoring data (24 percent and 11 percent). Similarly, more impact sample schools than reference sample schools reported using a coach to analyze universal screening data (49 percent and 32 percent) and progress monitoring data (40 percent and 28 percent).[83]

- **The majority of schools in both samples used universal screening and progress monitoring data to assess student progress and placement.**

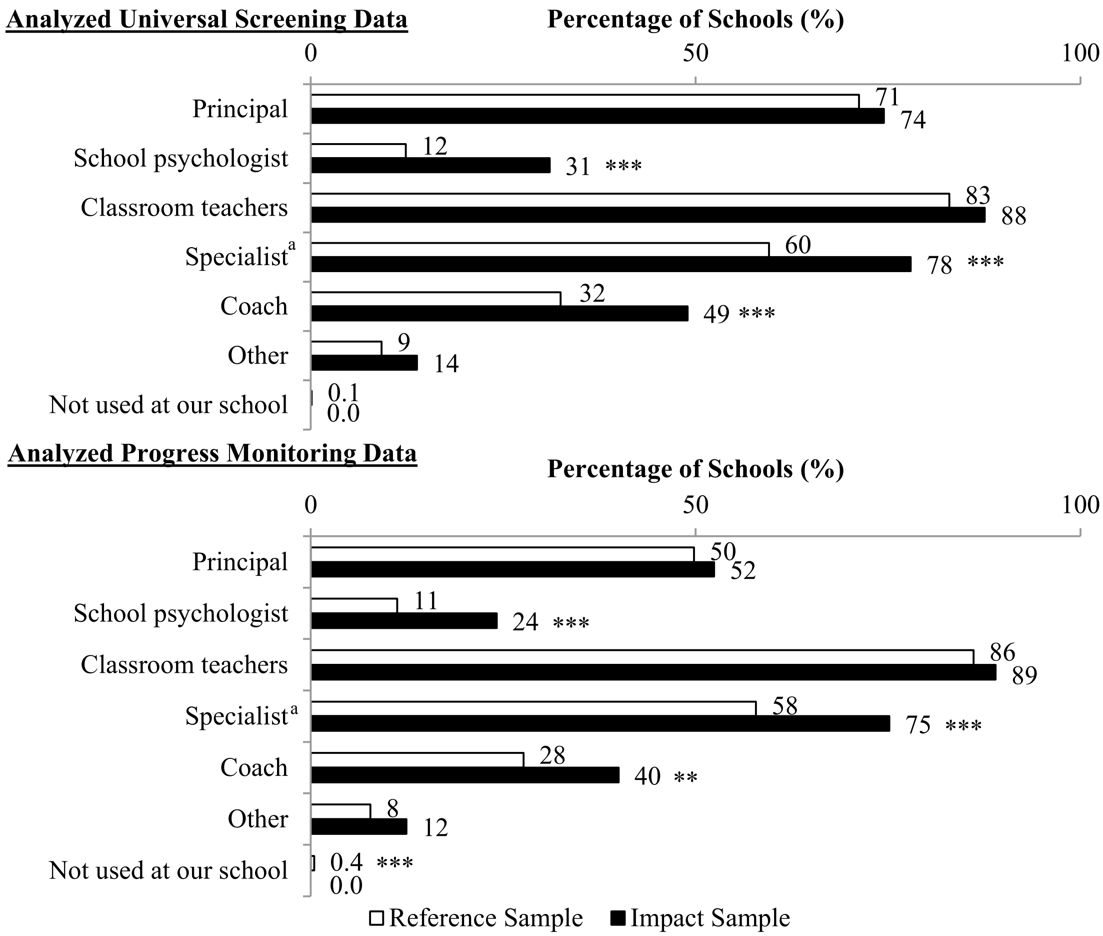
As shown in Table 3.4, schools in both the impact and reference samples favor progress monitoring measures in determining whether a child will likely reach grade-level reading benchmarks: 90 percent of impact sample schools and 81 percent of reference sample schools reported progress monitoring as "very important." Fewer impact sample schools than reference sample schools rated teacher observations as "very important" (69 percent versus 75 percent of schools, respectively) and curriculum-embedded tests (that is, tests integrated into classroom tasks) as "very important" (32 percent versus 48 percent). By contrast, the use of publishers' recommended screening scores to determine whether a child will likely reach grade-level reading benchmarks was considered "very important" by 56 percent of impact sample schools, compared with 40 percent of reference sample schools.

In addition to universal screening and progress monitoring data, schools could have tapped a variety of other data sources to help them make decisions about determining student

[83]The proportion of impact schools that use specialists in this role does not differ significantly from the proportion of reference sample schools that use them, as measured by the school's Title I eligibility status. Title I may be a relevant measure of school resources.

Figure 3.3

Types of Staff Who Analyzed Universal Screening and Progress Monitoring Data, Conditional on the School's Having Administered the Tests



SOURCE: School survey.

NOTES: [a]Specialists included reading interventionists, special education specialists, and English Language Learner specialists, as listed on the school survey.

Means and percentages were conditional on the school's having responded to the item presented. The first panel was conditional on the school having indicated someone administered universal screening assessments, and the second panel was conditional on having indicated the school administered progress monitoring assessments. Means reflect rounding. Statistical significance is indicated as follows: *** at the $p \leq 0.001$ level, ** at the $p \leq 0.01$ level, and * at the $p \leq 0.05$ level. The percentage of missing respondents for the items presented in this table is less than 3 percent.

The Response to Intervention (RtI) Evaluation

Table 3.4

Data-Based Decision-Making

	Percentage of Schools		Mean Difference	
	Impact	Reference	Between Samples	P-Value
Data Considered "Very Important" for Whether Students Will Reach Benchmarks				
Progress monitoring measures	90.2	80.8	9.4	0.078
Teacher observation	69.2	75.4	-6.2	0.035
Reading diagnostic tests	66.0	64.8	1.1	0.183
Standardized reading tests	53.8	59.8	-5.9	0.532
Curriculum embedded tests	32.4	47.9	-15.5	0.006
Used publisher's recommendations for universal screening or benchmark scores assessments	56.4	39.5	17.0	0.003
Data Considered "Always Used" to Inform Determinations of Eligibility for Special Education				
Universal screening or a benchmark in reading[a]	84.3	80.5	3.8	0.644
Information from systematic monitoring of student progress	86.6	77.8	8.8	0.108
Cognitive and reading assessments	77.6	76.5	1.1	0.953
Standardized reading tests	62.7	63.9	-1.2	0.940
Data from other procedures[b]	84.2	81.8	2.4	0.826

SOURCE: School survey.

NOTES: The maximum number of schools that responded to the item about data considered "very important" for whether students will reach benchmarks is 143 for the impact sample and 1,094 for the reference sample. The maximum number of schools that responded to the item about data "always used" to inform special education eligibility determinations is 134 for the impact sample and 1,028 for the reference sample.

Percentages and sample sizes were conditional on having responded to the specific item stem presented. Among eligible respondents, the overall missing rate for the item about Reading Benchmarks was less than 3 percent. Among eligible respondents, the missing rate for the item about Eligibility for Special Education among the reference sample was between 7.0 and 7.7 percent for the various items, and among the impact sample was between 7.6 and 8.3 percent for the various items. For each item about Eligibility for Special Education, the difference in missing rates between the samples had a p-value of greater than 0.5.

[a]Responses in this table are not grade-specific. As context, when schools were asked if they use universal screening of all students to identify those who may need support in reading, more than 90 percent of responding reference schools and all of the responding impact schools reported using universal screening data in Grade 1, and similar proportions did so in Grade 3; this survey item did not ask about Grade 2.

[b]As listed on the survey, “other procedures” include teacher observations, student work products, and parent reports.

eligibility for special education referral or placement. More than 60 percent of schools in each sample reported that they "always use" specific kinds of data in deciding a student's eligibility determination for special education; however, the samples did not differ significantly in the use of any of the data sources. Approximately 80 percent of the schools in both samples reported using universal screening data and benchmark testing scores, suggesting that these data are critical in making a decision about special education eligibility, which often occurs after a student has received intervention services for some time (Table 3.4). A large percentage of both types of schools (87 percent of impact sample schools and 78 percent of reference sample schools) used progress monitoring data for special education determination; the difference between the two samples is not statistically significant.

Findings on Special Education Identification

To provide a context for practices relating to special education determination, the study research team collected data on the percentage of students identified with any Individualized Education Program (IEP) and with an IEP for a Specific Learning Disability (SLD). The latter is the disability category most associated with using the student's nonresponsiveness to an intervention to inform determination of the student's eligibility for special education. The analysis compares the average SLD identification rates in fall 2011 between the impact sample schools and the states in the study, for each age group related to Grades 1 to 3. Figure 3.4 plots the averages for the state sample and the impact sample. (Data for the reference sample were not available.)

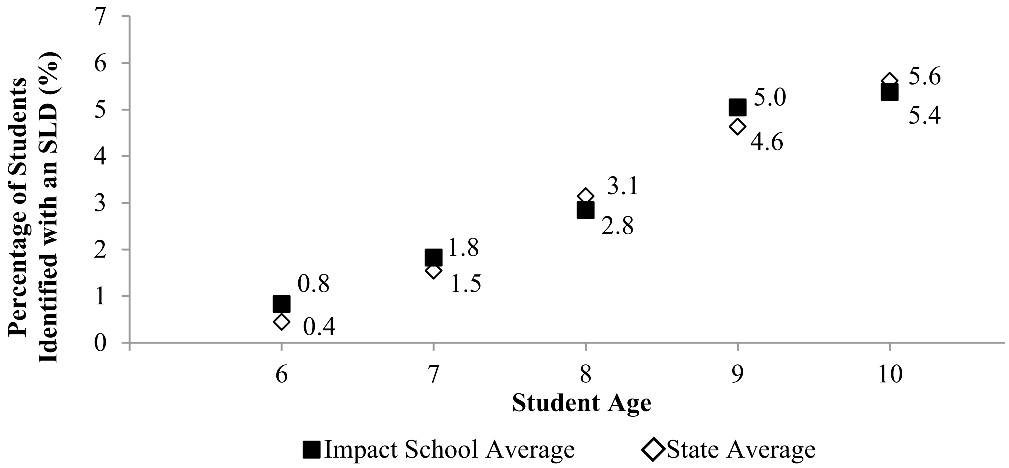
- **Special education identification rates are comparable between the impact sample and the 13 states as a whole.**

In both the impact sample and the states as a whole in fall 2011, less than 1 percent of 6-year-olds had been determined eligible for services under the SLD category (0.8 percent, compared with 0.4 percent, respectively). More students were determined eligible in older age groups. While 5.6 percent of 10-year-old students across all schools in the 13 states were categorized as having an SLD, 5.4 percent of impact sample students in the same age group were given such a diagnosis. This pattern of increase in identification rates by age is consistent with prior reports[84] as well as with the state averages for each age group. In fall 2011, the differences between the impact sample's rate of special education identification and the average rate in the 13 states were less than 1 percentage point for each age group. Note that these averages do not represent an estimate of the effect of RtI on the rate of identification for special education services and do not indicate that the type of student identified for these services is equivalent between the impact sample and the states.

[84] Blackorby et al. (2010).

Figure 3.4

Average Percentage of Students Identified with a Specific Learning Disability (SLD) for the State Sample and the Impact Sample, by Student Age in Fall 2011



SOURCES: Fall 2011 enrollment and counts data provided by a majority of districts in 13 states. Fall 2011 statewide data for children with disabilities served under the Individuals with Disabilities Education Act (IDEA)-Part B were downloaded from the Office of Special Education Programs (OSEP), Data Accountability Center. Enrollment data for 16 schools and for the state average were obtained from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), Public School Universe Data. The maximum number of schools in the impact school sample is 132.

NOTES: Percentages were calculated using the number of students identified in the numerator and the total enrollment in the denominator, for each age group. The numerator for "Specific Learning Disability" is those students identified with an Individualized Education Program (IEP) just in that category; the denominator is still total enrollment for that age.

The figure presents two averages for each age group. The state average reflects the average percentage of students identified across the 13 study states (the identification rate), by age group. The impact school average represents an average identification rate for each state among the impact schools in that state; each state's average rate was then weighted by the number of impact schools in that state, to reflect that each state has a different number of schools in the impact sample. The figure plots the mean of this weighted average for the impact school sample across the 13 study states.

CCD enrollment data for 2010 were used in the denominator for calculations of state proportions, since 2011 data were not yet released at the time of the analysis. Fall 2011 enrollment data that were provided by districts were used as the denominator to calculate proportions for impact schools. For 16 schools that did not provide fall 2011 enrollment data, fall 2010 enrollment data from the CCD were used as the denominator.

Most schools reported counts by age. For those that reported counts by grade, the study team assigned Grade 1 students to age 6, Grade 2 to age 7, Grade 3 to age 8, Grade 4 to age 9, and Grade 5 to age 10.

Conclusion

This chapter shows that RtI practices were widespread in the 13 study states. Across a variety of key components, the RtI framework was in place in a majority of reference sample schools, suggesting high adoption of the framework. In addition, a larger percentage of impact sample schools that had at least three years' experience with the framework reported having implemented key RtI features at a level in line with full implementation of an RtI model. Compared with the reference sample schools, the impact sample schools offered interventions more frequently during the week, conducted universal screening more often during the year, allocated staff to assist teachers with using data, and deployed a wider variety of staff to analyze data.

The characteristics of the impact sample are comparable to those of the reference sample in some respects but differ in others; therefore, the impact sample is not strictly representative of the average schools in these states. The analysis of practice differences between the two school samples is not causal, but it describes the context of RtI implementation in the study states. The findings about the impact sample demonstrate that those schools followed the RtI process, and they inform the focus on those schools for an in-depth description in Chapter 4 of the placement of students in different tiers and the intensity of small-group instruction and intervention services offered to students at different reading levels.

This page has been left blank for double-sided copying.

Chapter 4

Comparison of Reading Group Services in the Impact Sample

Chapter 3 illustrates that elementary schools throughout the 13 states that are included in the Response to Intervention (RtI) evaluation adopted RtI practices and that implementation was more pronounced in the impact sample schools than in the reference sample schools. Chapter 4 now focuses on a key aspect of these RtI practices: the use of multiple tiers of reading support. As discussed in Chapter 1 and depicted in Figure 1.1, RtI is often portrayed as a pyramid: all students receive core instruction (Tier 1) as the foundation, and students who read below grade level receive more support (Tier 2 or Tier 3, in addition to Tier 1), as needed. This framework of layered services allows schools to supplement core reading instruction with intervention for some students. "Intervention" does not necessarily refer to a different reading program but, rather, to support for struggling students that can be added or removed as needed, based on students' reading progress. This model informs the hypothesis for the chapter: students who need more assistance are placed in higher tiers, and students in groups for below-grade-level readers receive more intense reading services.

The chapter addresses three research questions regarding the tiered structure of RtI for elementary reading.

In impact sample schools (those with three or more years of implementing RtI):

1. To what extent did schools place students in tiers as suggested by earlier RtI models? To what extent did schools adjust tier placement during the school year?

2. To what extent is there variation in how schools organize reading services for specific reading levels?

3. To what extent were services for students reading below grade level more intense than for students reading at or above grade level?

The findings in this chapter confirm that the impact sample schools implemented key features of the RtI model, and they support the hypothesis that students reading below grade level received more intense and more varied reading-group services, as stated in the logic model (Chapter 1, Figure 1.3). Major findings include the following:

- The majority of students in Grades 1 to 3 in fall 2011 were initially placed in Tier 1 only. Schools adjusted tier placement of some students, although most students remained in the same highest tier between fall 2011 and winter 2012.

- Schools offered more intense reading support to students reading below grade level than to those reading at or above grade level. Reading groups for students reading below grade level included fewer students, and time offered to groups reading below grade level was greater than for groups reading at or above grade level. The content of instruction and qualifications of personnel providing reading interventions also differed for students reading at different levels.

The chapter begins by reporting findings about tier placement and movement between fall and winter of the 2011-12 school year. To the extent that impact sample schools were following the practice of placing fewer students at higher tiers, it allowed them to focus resources on students placed in those tiers who were reading below grade level and to intensify services for them. Next, the chapter describes the extent to which schools varied in how they organized services within the RtI framework. This includes a description of core reading instruction (the base that all students receive) and interventions (which only some students receive, to supplement core reading instruction). Finally, the chapter describes average service intensity for reading groups: small-group instruction within the core reading block (which all students receive to varying degrees) and intervention services (which only some students receive to varying degrees).[85] It reports differences between groups serving students identified as at or above versus below grade level, using descriptions of services offered in spring 2012. These estimated differences help inform the impact analysis in Chapter 5, which compares outcomes for students assigned to receive intervention with outcomes for those who were not. Discussion focuses on differences that are statistically significant. Results discussed in text may round up from the tables in some cases, for clarity of presentation in text.

Findings on Student Movement Between Tiers

As Chapter 3 discusses, several aspects of data-based decision-making support the use of multiple tiers in the RtI framework. As shown in Chapter 3, impact sample schools used universal screening to place students in tiers initially and then used progress monitoring data to update tier placements periodically, based on student progress in response to services.

This section describes the initial placement of students based on fall screening tests and the extent of subsequent movement to a different placement based on winter screening tests.[86]

[85]By the time of the survey, students had been tested and identified as reading at a particular proficiency or skill level. They were placed in a corresponding small group for intervention and for small-group instruction. The rest of the chapter uses "reading level" as a shorthand term.

[86]Data regarding the placement of students in reading tiers and the number of trimesters spent at each reading tier come from fall and winter tier placement data provided by schools. The sample of schools varies by

Although this descriptive analysis did not address the reasons or mechanisms behind the movement — whether it is due to student progress and responsiveness to services, or lack thereof, or to a result of inappropriate initial placement by schools — it illustrates the extent to which schools used data and adjusted student placements. It also illustrates the extent to which impact sample schools followed the tiered aspect of the RtI pyramid and placed fewer students in higher tiers. (See Chapter 1, Figure 1.1.) In prior studies of student placements, the percentage of students placed in Tiers 2 and 3 was 20 percent or greater.[87] The results of fall tier placement are consistent with prior studies of RtI.

For the Grade 1 sample, Figure 4.1 shows the initial placement of students in the fall in the leftmost column, the movement of students in the middle column, and the final winter placements in the rightmost column.[88] Each column shows the distribution of students, by tier. The majority of students are placed in Tier 1 in the fall (59 percent, shown in the bottom left segment), while a smaller percentage are in Tier 2 (25 percent) and an even smaller percentage are in Tier 3 (16 percent, shown in the top left segment). The distribution of students across tiers is in line with the idea of having fewer students in higher tiers. The percentages in the middle column indicate the percentage of students in each tier (or segment) who either remain in that tier or shift to a different tier in the winter. The shading indicates the direction of the movement: the black shading indicates movement to the most intense tier (Tier 3); the gray shading indicates movement to Tier 2; and the unshaded segments indicate movement to Tier 1.

Two findings are evident. First, the majority of students in Tiers 1 and 3 remained within the same tiers between trimesters, while about half of all students in Tier 2 changed tiers (to either Tier 1 or Tier 3). Within Tier 1 — which had the largest number and percentage of students — 86 percent of those initially placed in the tier remained in Tier 1 between fall and winter, meaning that they received just core instruction (and likely no intervention services) for two trimesters. Of students initially placed in Tier 3, 65 percent of students remained in that tier, meaning that they received Tier 3 intervention services in the same tier for at least two trimesters.

Second, schools did not assign students to receive the same level of instruction or intervention services permanently, but instead adjusted student placements. For example, the largest percentage of students who moved were in Tier 2; one-third (33 percent) of students initially placed in this tier moved to Tier 1, and 17 percent moved to Tier 3. This suggests that schools

---

grade: 89 schools in Grade 1, 102 in Grade 2, and 93 in Grade 3. These schools placed at least one student in all three reading tiers and had students with tier placement data for both trimesters.

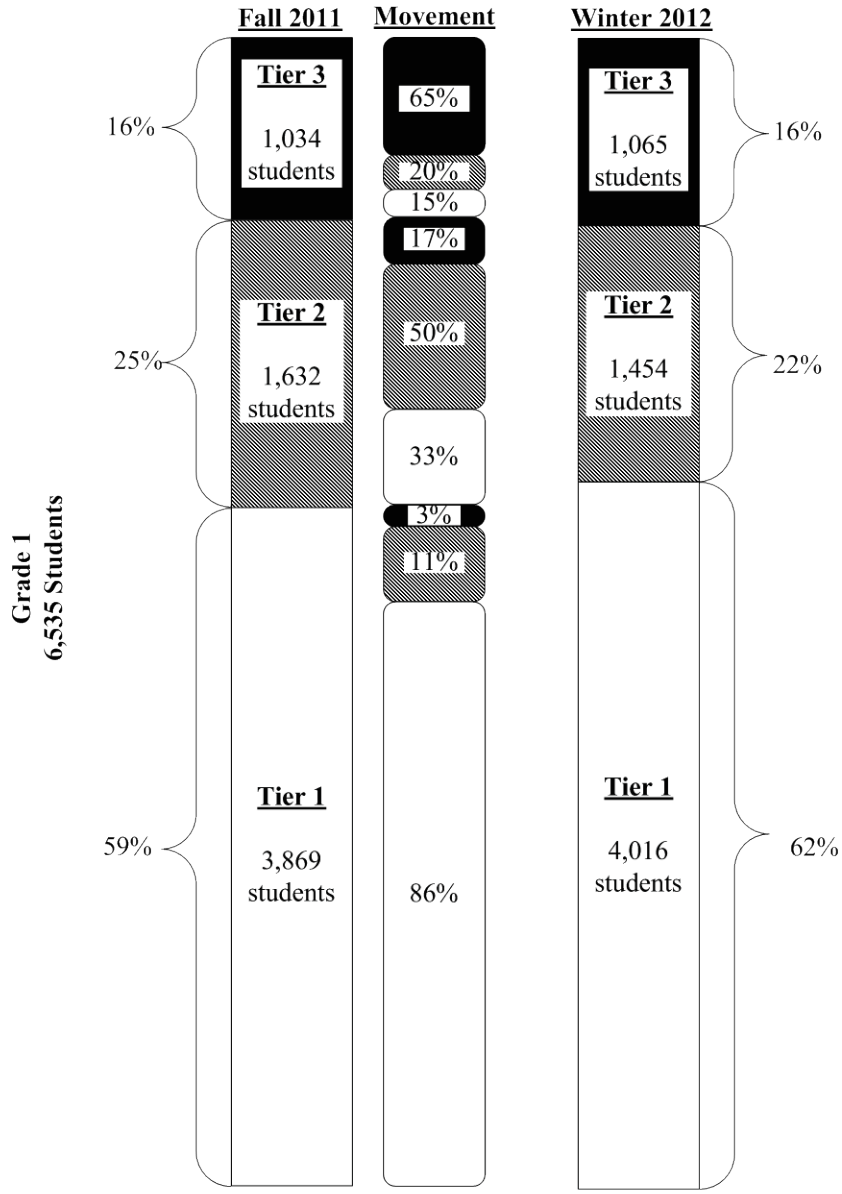
[87]Mellard, McKnight, and Jordan (2010).

[88]In each grade, the analysis pools students across schools to describe placement and movement in the sample of students who had data for both fall and winter.

The Response to Intervention (RtI) Evaluation

Figure 4.1

Student Distribution, by Tier, and Highest Tier Movement



SOURCES: Fall 2011 and winter 2012 tier placement data.

NOTES: Students placed in Tier 1 typically receive only core reading instruction; those placed in Tiers 2 and 3 typically receive core reading instruction plus intervention services. Tier assignment occurs based on results from screening assessments conducted in the fall and winter. The Grade 1 school sample size was restricted to 89 schools that had at least one student in each of Tier 1, Tier 2, and Tier 3 in both fall and winter.

were reassessing these students to determine whether they needed to remain in that tier or needed to receive more intense or less intense services in the second trimester. Despite some movement, the distribution of students overall remained similar in the winter and the fall, with the majority of students placed in Tier 1 and fewer students placed in Tiers 2 and 3.

Figure 4.2 details the stability and movement of students for all three grades, with separate bars representing each tier in each grade. The majority of students in all three grades remained “stable” — shown in the dark-gray segments as the students who remained in the same tier in fall and winter. A smaller number and proportion of students were placed in Tiers 2 and 3 in the Grade 3 sample.

- **Impact sample schools followed RtI practices of adjusting students’ tier placement over time. About three-fourths of Grade 1 students in these schools during 2011-12 remained in the same reading tier between fall and winter, and about one-fourth of students moved between tiers from fall to winter in Grade 1.**

Table 4.1 summarizes whether students moved to a less intense or a more intense tier. The majority of students in all grades remained in the same tier (74 percent in Grade 1 and 83 percent in Grades 2 and 3). In Grade 1, 13 percent of students moved to a more intense tier, while 14 percent moved to a less intense tier; in total, 27 percent moved. The percentages of movement are smaller for Grades 2 and 3. Across grades, stability of tier placement was coupled with movement of some students to different tiers, suggesting that schools used screening data to adjust placement between two points in time.

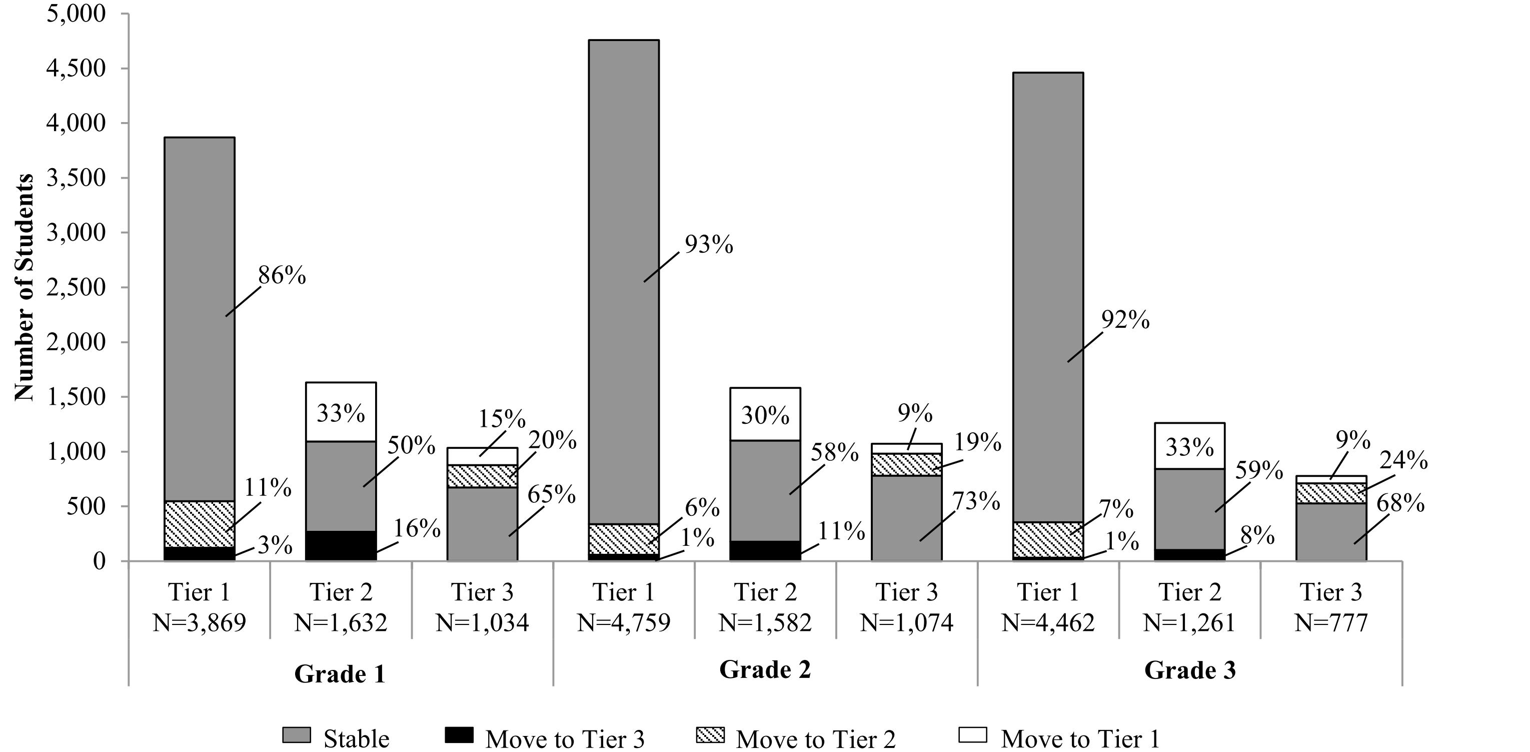
Findings on Variation in Schools’ Organization of Reading Services

Given the tier placements described in the preceding section, the impact sample schools faced choices about how to organize services for students at different reading levels, based on the staff and other resources available. In some cases, this means that the reality of how schools offered and delivered services differs from the ideal distinctions suggested by the three-tiered model described in Chapter 1. Although there is a rough equivalence between Tier 2 and reading Somewhat Below grade level, in some impact sample schools, the threshold score that individual schools set for placement in Tier 2 does not always correspond to standards of reading at grade level (discussed further in Chapter 5). As a result, the rest of this chapter describes services in terms of students’ reading-group level, rather than in terms of tiers, in order to describe services designed to boost the reading skills of those in need of support. Service providers

The Response to Intervention (RtI) Evaluation

Figure 4.2

Detail of Student Tier Movement Between Fall and Winter, by Grade



SOURCES: Fall 2011 and winter 2012 tier placement data.

NOTES: The school sample sizes are: 89 for Grade 1, 102 for Grade 2, and 93 for Grade 3. These schools had at least one student in each of the three tiers in both fall and winter. Only students with fall *and* winter tier placements are included in the figure.

Stacked columns do not consistently add to 100 percent due to rounding. "N" represents the number of students in that tier.

Table 4.1

Summary of Student Tier Movement Between Fall and Winter, by Grade

	Moved to More Intense Tier (%)	Remained in the Same Tier (%)	Moved to Less Intense Tier (%)	Number of Students
Grade 1	13	74	14	6,535
Grade 2	7	83	10	7,415
Grade 3	7	83	10	6,500

SOURCES: Fall 2011 and winter 2012 tier placement data.

NOTES: The school sample sizes are 89 for Grade 1, 102 for Grade 2, and 93 for Grade 3. These schools had at least one student in each of the three tiers in both fall and winter. Only students with fall and winter tier placements are included in the table.

Row percentages sum to 100 percent, but rounding makes the Grade 1 value appear slightly greater than 100.

completed surveys that asked them to describe services "provided to students somewhat below grade-level benchmarks (sometimes called Tier 2 students) or far below grade-level benchmarks (sometimes called Tier 3 students)."

This section describes three key aspects of how schools delivered reading services: (1) the reading levels into which student reading groups were divided, (2) the time when core reading instruction and intervention occur, and (3) how support services were offered. In each of these aspects, the analysis finds differences from the prior literature. These data were collected for reading groups, not for individual students.

Reading Groups, by Reading Level and Type

In order to test the hypothesis of increasing intensity, the analysis needs to compare differences in services between groups serving students at *distinct,* mutually exclusive reading proficiency levels. As a result, the analysis focuses on reading groups in which each group serves only students At or Above *or* Somewhat Below *or* Far Below grade level in reading. As shown in the first three rows of Table 4.2, a total of 94 percent of groups for small-group instruction and a total of 87 percent of intervention groups across schools, respectively, were sorted into such homogeneous groups. Groups with students from more than one reading level (for

Table 4.2

Distribution of Reading Groups, by Reading Level and Grade

	Small-Group Instruction (%)	Intervention (%)
Grade 1		
At or Above grade level	57	16
Somewhat Below grade level	23	40
Far Below grade level	14	31
At or Above and Somewhat Below grade level	3	5
Somewhat Below and Far Below grade level	4	8
Grade 2		
At or Above grade level	53	13
Somewhat Below grade level	23	43
Far Below grade level	17	29
At or Above and Somewhat Below grade level	3	5
Somewhat Below and Far Below grade level	3	9
Grade 3		
At or Above grade level	51	12
Somewhat Below grade level	29	47
Far Below grade level	15	27
At or Above and Somewhat Below grade level	3	3
Somewhat Below and Far Below grade level	3	10

SOURCES: Teacher and interventionist surveys.

NOTES: Reading levels were reported by respondents. Groups that served multiple reading levels are classified as serving either At or Above and Somewhat Below students or as serving Somewhat Below and Far Below students. Otherwise, groups are homogeneous by reading level.
In Grade 1, there are 1,590 instruction groups in 119 schools and 1,425 intervention groups in 131 schools. In Grade 2, there are 1,380 instruction groups in 118 schools and 1,096 intervention groups in 126 schools. In Grade 3, there are 1,265 instruction groups in 111 schools and 969 intervention groups in 124 schools.

example, students At or Above and Somewhat Below grade level in the same group) were excluded from the analysis, because the service intensity for those groups could not be attributed to a single reading level.

The rest of the chapter refers to reading groups as follows: groups comprising students identified as reading At or Above grade level are in "AA groups"; groups comprising students Somewhat Below grade level are in "SB groups"; and groups comprising students Far Below grade level are in "FB groups." These reading groups roughly correspond with students whose highest tier placement was Tier 1, Tier 2, or Tier 3, respectively, with the understanding that students in Tier 2 or Tier 3 also received core Tier 1 reading instruction.

Small-Group Instruction by the Classroom Teacher

The percentage of small instruction groups for each reading level roughly reflects the percentage of students reading at each of these levels. In Grade 1, the majority of teacher-led small reading groups were AA groups (57 percent), SB groups made up 23 percent of all groups, and FB groups made up a smaller percentage, at 14 percent of all groups. A similar pattern held in other grades.

The study research team did not have exact percentages of students at each reading level, because reading levels were measured only at the group level and not the student level, and the data collected could not reliably link individual- and group-level data. Based on the correspondence between tiers and reading levels described in Chapter 1 and previously in this chapter, the distribution of small instruction groups by reading level is similar to the distribution of students by tier reported in Figure 4.1; the percentage of students whose highest tier of instruction is Tier 2 or Tier 3 can be viewed as a rough approximation of the percentage of students Somewhat Below or Far Below reading level, respectively.

Intervention Groups

Intervention groups that served At or Above grade-level students were not discussed in the existing RtI literature, because intervention was associated with services for children placed in Tier 2 or Tier 3. In the impact sample, however, interventionists reported that 16 percent of intervention groups in Grade 1 served exclusively AA students (the first row of data in Table 4.2). The majority, but not all, of intervention groups did not serve these students; 40 percent are SB groups, and 31 percent are FB groups. This result suggests that, in some schools, a wider variety of students received intervention services.[89] A similar pattern held in other grades.

Allocation of Time

As discussed in Chapter 1, the school day in the RtI evaluation consisted of the core reading block and time outside the core. The "core reading block" is defined as the time that

---

[89]Some intervention groups include one or more students with an Individualized Education Program (IEP). This includes students with behavioral, physical, or reading-related IEPs. As the literature on reading needs of students in special education might suggest, the highest percentage of groups with IEP students are for those reading Far Below grade level (56 percent). For SB student groups, the percentage with IEP students is smaller than the percentage without IEP students (26 percent, compared with 68 percent). These results suggest that IEP status and FB reading level are correlated, but some IEP students are not in FB groups, and not all FB groups include students with an IEP.

schools dedicated to teaching reading to all students.[90] Reading groups that met during the core reading block are referred to as "during the core." Groups that met outside, or in addition to, the core reading block are referred to as "outside the core."

During the core reading block, instruction time was divided as follows:

- Whole-class instruction services provided to all students
- Small-group instruction provided to all student reading levels
- Partner or peer work
- Independent reading or other enrichment activity
- One-on-one tutoring or other services focused on students reading below grade level

Teachers in the impact sample schools described how they allocated time to these modes during the core reading block. As shown in Figure 4.3, the largest percentage of time was spent on whole-class and small-group instruction — 33 percent and 25 percent of the time, respectively. The other modes of reading instruction together made up approximately 42 percent of the reading block.

Small-group instruction, which represented one-fourth of the core reading block, was the time when different instruction could be provided to students, based on their current reading levels.[91] As a result, small-group instruction is a focal point for the comparison of services between reading levels.

A schematic of how schools might allocate time and services during the school day, based on prior descriptions of RtI, is presented in Chapter 1 (Figure 1.2).[92] That schematic illustrates that schools operating according to a model used in earlier randomized controlled trials would provide small-group instruction services during the core reading block to all reading

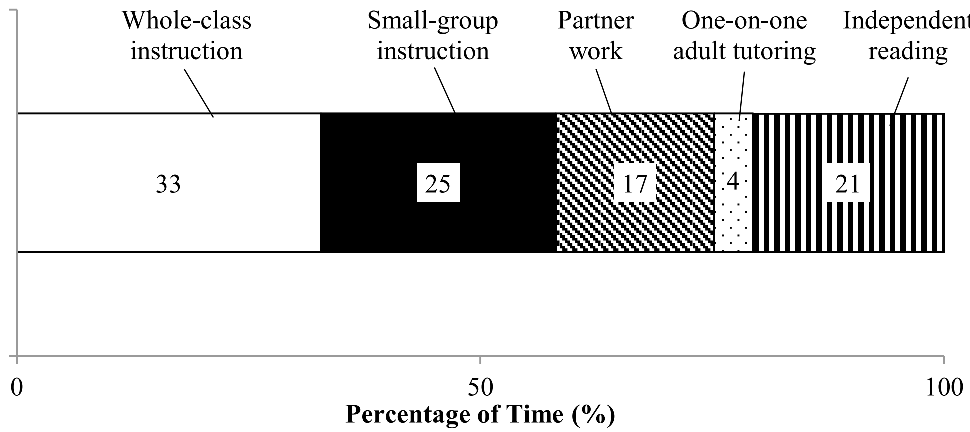
[90]On average, classroom teachers in the impact sample reported devoting about 97 minutes per day to the core reading block. Results reported from the school survey in Chapter 3 are similar, with 102 minutes per day, on average. The National Reading Panel report, policies such as Reading First, and programs such as Success for All recommend a reading block length of about 90 minutes.

[91]The use of small groups is discussed in the IES Practice Guide on RtI for Elementary Reading (Gersten et al., 2009) as a way to differentiate instruction. It is usually provided by the teacher, which likely means that a single adult is moving among multiple reading groups during the same time period (unlike intervention groups, whereby at least one adult works with one group during a given time period).

[92]Gilbert et al. (2013); Mathes et al. (2005); Vadasy et al. (1997); Fletcher and Vaughn (2009); Vaughn (2008).

Figure 4.3

Allocation of Time in the Core Reading Block, by Mode of Instruction



SOURCE: Teacher survey.

NOTES: The school sample is 137, and the teacher sample is 1,128.
Respondents reported an average of 97 minutes for the core reading block.

levels, and they would target intervention services to students reading below grade level, primarily outside the core reading block.

However, in a portion of the impact sample schools, the organization of services differed from prior literature, or Figure 1.2, in three ways:

- Intervention groups included students reading at or above grade level, rather than just students reading below grade level;
- Interventionists provided services *during* and outside the core reading block, rather than just outside the core reading block. As a result, not all students were receiving intervention time *in addition to* the full core reading time; and
- Schools designated a variety of staff types to provide intervention services, including classroom teachers.

The next section of the chapter discusses the differences between the study findings and Figure 1.2 in more detail.

How Services Are Delivered

To provide a context for understanding the impact analysis in Chapter 5, the remainder of this chapter describes service delivery. Because the impact analysis relies on the contrast between students who receive only core instruction and those who receive core instruction plus intervention services, this chapter presents service contrast similarly. As a result, groups serving readers Somewhat Below and Far Below grade level are combined to represent those students who are most likely to be in the "treatment" group, as defined in Chapter 5, and then are compared with groups serving readers At or Above grade level, which constitutes the "comparison" group in the impact analysis.[93]

While the RtI structure described in earlier impact studies and conceptual frameworks presumes that all students receive the *same* amount of core instruction and that *some* students below grade level receive additional supports through intervention, the impact sample schools organized services in ways that made use of additional time and provided services for more students.

- **Impact sample schools varied in how they organized and delivered intervention services for reading groups, in ways that differ from earlier RtI studies.**

For *each grade,* the study research team sorted schools into two categories:

- **"Below-Only" Schools.** These schools provided intervention services only for groups with students identified as reading below grade level. In a given grade, these schools did not offer intervention services to groups made up of AA students.
- **"All-Level" Schools.** These schools provided intervention services to some groups at *all* reading levels. In a given grade, they served at least one intervention group at each reading level.

This second category is in contrast to prior RtI small-scale randomized controlled trials that had focused intervention resources on readers below grade level. Schools that reported serving AA groups did not necessarily provide intervention to *all* At or Above readers but, rather, had at least one AA intervention group in a grade. Such schools tended to serve these

[93]It is important to keep in mind that, given the specific design of the study, the impact findings are only applicable to a specific subset of the treatment group (discussed in more detail in Chapter 5 and Appendix E), while the service contrast findings presented in this chapter reflect the difference between the full treatment and comparison groups. Therefore these service contrast findings can provide contextual, but not directly corresponding, information for interpreting the impact findings.

groups both during and outside the core, and these groups tended to be served by a variety of interventionist staff types.

- **In Grade 1, 45 percent of impact sample schools provided intervention services to some groups at *all* reading levels, not just to reading groups *below* grade level.**

Table 4.3 details the division of schools into these categories for each grade. Of the 131 schools in the Grade 1 sample, 55 percent followed the more traditional model of interventionists serving only below-grade-level reading groups; however, 45 percent provide intervention services to at least one group from each reading level, which means that at least one group of At or Above students received intervention services that may have been designed for students reading below grade level. For Grades 2 and 3, the percentages of All-Level schools are 36 percent and 31 percent, respectively, suggesting that more schools in those grades than in the Grade 1 sample followed the traditional model of delivering intervention services only to readers below grade level.

The delivery of reading-group services could have differed between these two categories of schools. Therefore, the study research team used this categorization of schools to describe the intensity of support provided to students at different reading levels and in the impact analysis in Chapter 5.

- **In Grade 1, 67 percent of impact sample schools provided some intervention *during* the core reading block, not just *outside* the core.**

Earlier studies of RtI that used randomized controlled trials designed intervention to occur *in addition to* the core reading block time. Of the 131 schools in the Grade 1 sample, 88 (or 67 percent) provided some intervention during the core, and 43 provided intervention outside the core only. Across reading levels and grades, the fraction of schools that served intervention groups only outside the core is similar: about one-third, with the remaining two-thirds of schools serving some intervention groups during the core.

Table 4.4 shows the average percentage of intervention groups, by reading level, that met outside the core reading block, with a separate row for each school category. For Grade 1, in All-Level schools, only 36 percent of groups at or above grade level and 42 percent of groups below grade level were served outside the core reading block.[94] In the Below-Only schools, 60

[94]Percentages do not add to 100 due to missing data on the survey variable indicating whether the group meets during or outside the core.

Table 4.3

Impact Sample Schools Categorized by Intervention Model, by Grade

	Below-Only Schools (%)	All-Level Schools (%)	Number of Schools
Grade 1	55	45	131
Grade 2	64	36	126
Grade 3	69	31	124

SOURCE: Interventionist survey.

NOTES: The Below-Only school sample represents schools that had at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. Row percentages sum to 100.

percent of groups below grade level met outside the core.[95] The pattern of Below-Only schools serving a majority of groups outside the core held in Grades 2 and 3 as well. This finding suggests that, in these schools, intervention time for some students was supplemental, as the RtI literature suggests.

The organization of services described in earlier impact studies is to offer supplemental time (intervention services only outside the core) focused on students reading below grade level. Across grades, only one-fifth of schools conformed to this approach. This section provides evidence that, instead, (1) intervention groups included students reading At or Above grade level and (2) interventionists provided services *during* and outside the core reading block. In some schools, both characteristics occurred together; in other schools, just one of these differences may have manifested.

The implication is that intervention may have *replaced* rather than supplemented some instruction services during the core. In All-Level schools, where more than half of all intervention groups met *during* the core, intervention may have displaced instruction for more groups of students. The reasons why this practice occurred — whether it was because time for reading

[95]Results for during the core are not shown, because they are simply the difference between 100 percent and the percentage for outside the core.

Table 4.4

Percentage of Intervention Groups That Meet Outside the Core Reading Block, by Grade

	Grade 1 Groups		Grade 2 Groups		Grade 3 Groups	
	At or Above Grade Level	Below Grade Level	At or Above Grade Level	Below Grade Level	At or Above Grade Level	Below Grade Level
All-Level schools (%)	36	42	40	46	35	43
Below-Only schools (%)	NA	60	NA	61	NA	60
Average percentage across all groups (%)	36	49	40	53	35	51
Total groups	222	1,014	140	795	115	721

SOURCE: Interventionist survey.

NOTES: The Below-Only school sample represents schools that had at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services. A group is defined as meeting outside the core reading block if the respondent answered, "at a time other than the core reading block but within the school day."

The school samples for Grade 1 are 72 Below-Only schools and 59 All-Level schools. The school samples for Grade 2 are 81 Below-Only schools and 45 All-Level schools. The school samples for Grade 3 are 85 Below-Only schools and 39 All-Level schools. Percentages reflect rounding.

services outside the core was more constrained or less flexible than time during the core, or for other reasons — could not be answered with data collected by the study.[96]

Findings on Service Contrast Between Reading Groups

The analysis in this section describes the reading services along five dimensions or mechanisms that schools can manipulate to adjust services for weaker readers. The study research team selected mechanisms that are included in the IES Practice Guide on RtI for Elementary Reading

[96]Prior studies using randomized controlled trials to demonstrate the efficacy of RtI designed Tier 2 time to be in addition to Tier 1. In one such trial by Vadasy, Sanders, and Tudor (2007), some students received tutoring interventions during the core reading instruction period, despite the specification that Tier 2 intervention time should supplement the core and monitoring from research staff. Mellard, McKnight, and Jordan (2010) found that total reading time was fixed and that schools made choices about how to allocate instruction and intervention time within that fixed block, rather than adhering to prespecified allocations.

and that researchers in reading instruction addressed in prior impact studies.[97] That literature suggests that small-group reading instruction can be adjusted to student needs and that intervention is more intense if reading groups:

- Serve fewer students (that is, in smaller groups)
- Meet for more time

In addition, findings from RtI impact studies that focused on reading suggest that schools can vary the services provided to reading groups if they:

- Target reading skills to student needs
- Use specialized or targeted staff to provide interventions
- Monitor student progress more often

The study research team created variables for each of these five mechanisms, based on responses to the teacher and interventionist surveys, with separate measures from each survey. In each subsection below, the chapter describes the variable and compares whether the average below-grade-level reading group received more intense services than the average at-or-above-grade-level group.

The rest of this section of the chapter presents the service contrast between reading groups for the five mechanisms listed above. The results show differences between reading levels for small-group instruction separately from differences for intervention groups. (Hereafter, the chapter uses "instruction" to refer to small-group instruction by teachers during the core reading block.) Unless otherwise noted, the discussion focuses on statistically significant estimated differences between reading levels. For simplicity of presentation and to aid in interpretation, the analysis is presented separately for Below-Only schools and All-Level schools. (To maximize sample sizes, results are not additionally broken down by whether services were provided during or outside the core.) For each mechanism, the discussion starts with results for Grade 1 and then briefly summarizes results for Grades 2 and 3.

Mechanisms to Intensify Support Services

Prior RtI impact studies that have focused on Tier 2 interventions generally have not compared Tier 2 services with Tier 3 services. The current study does not distinguish between Tier 2 interventions and Tier 3 interventions (or small-group reading instruction and interventions for students reading Somewhat Below in contrast to Far Below grade level) either. The

[97]Connor, Alberto, Compton, and O'Connor (2014); Fuchs, Fuchs, and Compton (2004); Mellard, McKnight, and Jordan (2010); Speece, Palombo, and Burho (2013); Wanzek and Vaughn (2008, 2010).

relevant comparison for the study is between services for student groups reading at or above grade level versus student groups reading below grade level. As a result, tables present results for groups with students Somewhat Below and Far Below grade level combined.[98]

Group Size

The IES Practice Guide on RtI for Elementary Reading suggests small groups for Tier 2 students, but it does not recommend a specific group size. Figure 4.4 shows group size by types of schools and groups in the impact sample for Grade 1. The left-hand side of the figure shows findings for Below-Only schools, and the right-hand side shows findings for All-Level schools. For each type of school, findings for small-group instruction are shown on the left, and intervention groups are shown on the right. For each of these, the exhibit compares the size of AA groups (white bars) with the size of groups below grade level (black bars).

- **Across the impact sample schools, the group sizes for small-group instruction and for intervention were smaller for readers below grade level than for readers at or above grade level, by one student, on average.**

In Grade 1, small instruction groups during the core generally served one fewer student in groups below grade level than in AA groups (Figure 4.4): 4.3 students, compared with 5.3 students in Below-Only schools (a difference of –0.9 student), and 4.5 students, compared with 5.6 students in All-Level schools (a difference of –1.1 students). Intervention groups in All-Level schools had 1.5 fewer students in groups below grade level than in AA groups (4.1 students and 5.6 students, respectively).[99]

Table 4.5 shows that differences in group size by group reading levels are also statistically significant in Grades 2 and 3 both for small-group instruction during the core and for intervention groups. Small instruction groups serving students below grade level in Grade 2 were 1.2 students smaller than AA groups, on average, in Below-Only schools and were 1.5 students smaller in All-Level schools. For Grade 3, instruction groups serving students below grade level were 1.1 students smaller than AA groups in Below-Only schools and 1.5 students smaller in All-Level schools. In All-Level schools, intervention groups below grade level had 1.9 fewer students than AA groups in Grade 2 and 1.8 fewer students than AA groups in Grade 3.

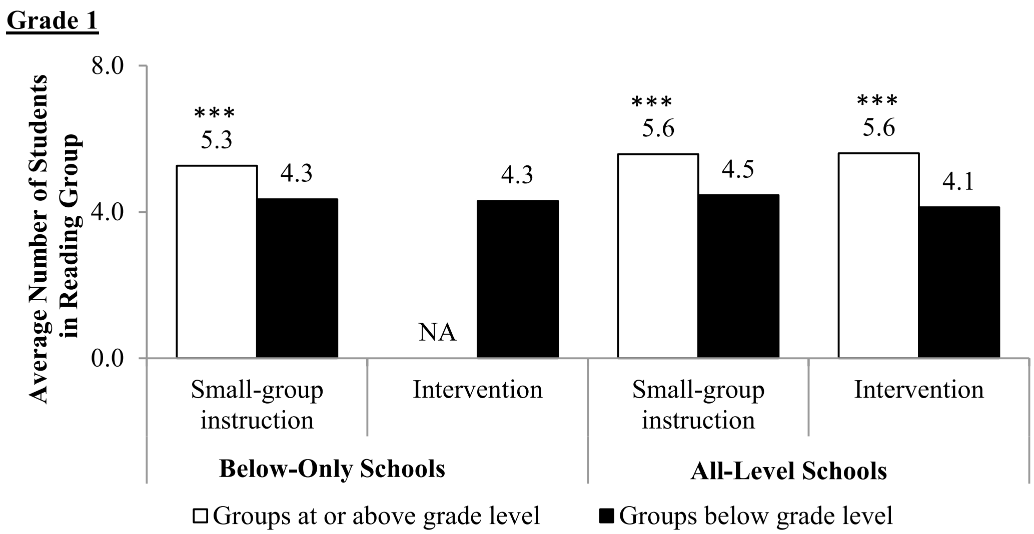
---

[98]Mixed-level groups were excluded from the analysis. Groups of any size from 2 to 10 as well as occasional cases of one-on-one interventions were included in the group-level analysis and description of small-group interventions.

[99]It is not possible to make a corresponding comparison for Below-Only schools, which have no AA intervention groups.

Figure 4.4

Service Contrast for Group Size: Difference Between Groups At or Above Grade Level and Below Grade Level in Below-Only and All-Level Schools, for Grade 1



SOURCES: Teacher survey and interventionist survey.

NOTES: "Small-group instruction" refers to services provided by teachers during the core reading block to all students. Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. The Below-Only school sample represents schools that had at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services. No tests were performed between intervention groups in Below-Only schools, which do not provide intervention to At or Above grade-level groups.

Statistical significance is indicated as follows: *** at the $p \leq 0.001$ level, ** at the $p \leq 0.01$ level, and * at the $p \leq 0.05$ level.

The school sample sizes are, for small-group instruction, 67 Below-Only schools and 51 All-Level schools; for intervention, 72 Below-Only schools and 59 All-Level schools.

Table 4.5

Service Contrast for Group Size: Difference Between Groups At or Above Grade Level and Below Grade Level in Below-Only and All-Level Schools, by Grade

Average Number of Students	Groups At or Above Grade Level	Groups Below Grade Level	Mean Differences Between Groups	P-Value
Grade 1				
Below-Only schools				
Small-group instruction	5.3	4.3	-0.9	0.000
Intervention	NA	4.3	NA	NA
All-Level schools				
Small-group instruction	5.6	4.5	-1.1	0.000
Intervention	5.6	4.1	-1.5	0.000
Grade 2				
Below-Only schools				
Small-group instruction	5.9	4.7	-1.2	0.000
Intervention	NA	4.2	NA	NA
All-Level schools				
Small-group instruction	5.8	4.3	-1.5	0.000
Intervention	6.4	4.5	-1.9	0.000
Grade 3				
Below-Only schools				
Small-group instruction	6.4	5.3	-1.1	0.000
Intervention	NA	4.5	NA	NA
All-Level schools				
Small-group instruction	6.3	4.8	-1.5	0.000
Intervention	6.2	4.4	-1.8	0.000

SOURCES: Teacher survey and interventionist survey.

NOTES: "Small-group instruction" refers to services provided by teachers to all students during the core reading block. Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. The Below-Only school sample represents schools that had at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services. No tests were performed between intervention groups in Below-Only schools, which do not provide intervention to At or Above grade-level groups.

Group Session Time

Time is calculated as minutes of small-group instruction or intervention per week. It involves combining variables about session length and frequency (minutes per session multiplied by the number of meetings per week)[100] to obtain a "dosage" measure. Because some reading levels met multiple times per week while others met fewer times, a per week measure allowed for fair comparisons between reading levels. Figure 4.5 is organized like Figure 4.4, displaying this intensity measure for Below-Only schools on the left-hand side and for All-Level schools on the right-hand side.

- **Small-group instruction time during the core reading block was longer for readers below grade level than for readers at or above grade level in impact sample schools. In All-Level schools that provide intervention services to both types of reading groups, the intervention time did not differ significantly in Grade 1.**

For Grade 1, in Below-Only schools, small-group instruction time for AA groups was 62 minutes per week, compared with 89 minutes for groups below grade level, reflecting different instruction time by reading level. (This part of Figure 4.5 does not display intervention minutes for AA groups because such groups do not exist in the impact sample schools.) In All-Level schools (the right-hand side of the figure), instruction time for AA groups was 100 minutes per week, compared with 140 minutes for groups below grade level.[101] The difference in intervention time in All-Level schools for Grade 1 is not statistically different between reading levels. The average intervention time reported in both categories of schools falls within the range suggested by the IES Practice Guide on RtI for Elementary Reading: 60 minutes to 200 minutes per week.[102]

Table 4.6 shows results for Grades 2 and 3 as well. In Below-Only schools, small-group instruction time was offered for about 31 minutes per week more to readers below grade level in

---

[100]Details about the conversion of time-span categories into a continuous measure of minutes are discussed in Appendix C.

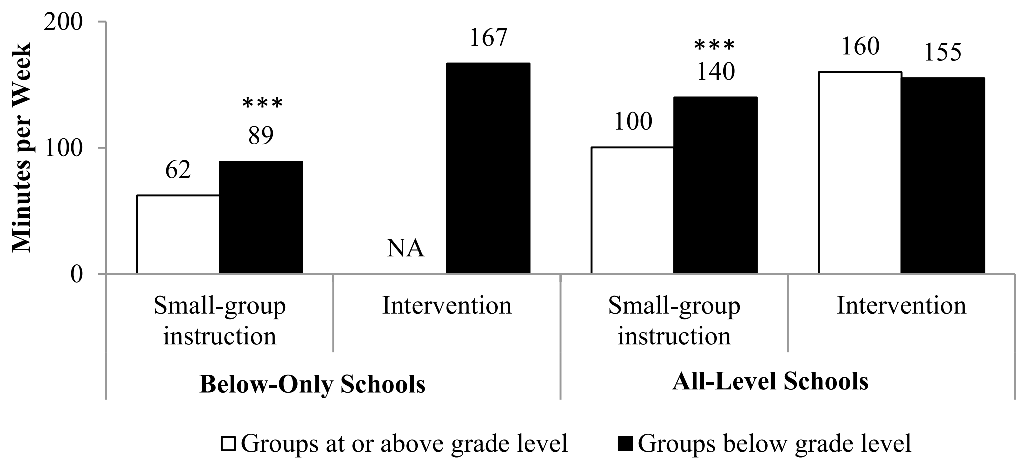
[101]The core reading block is about 120 minutes in All-Level schools — 20 minutes longer than in Below-Only schools, which may explain the difference in the amount of time spent on small-group instruction during the core between the two categories of schools. The proportion of All-Level schools that were eligible for Title I funds is greater than the proportion of Below-Only schools eligible for these funds, which may be related to having available staff to serve students during the core.

[102]The IES Practice Guide on RtI for Elementary Reading suggests that small groups for struggling readers meet three to five times a week for approximately 20 to 40 minutes each. Taking the minimum value of three sessions of 20 minutes each yields 60 minutes per week, while the maximum value is obtained for sessions of 40 minutes each at five times per week.

Figure 4.5

Service Contrast for Minutes per Week: Difference Between Groups At or Above Grade Level and Below Grade Level in Below-Only and All-Level Schools, for Grade 1

<u>**Grade 1**</u>



SOURCES: Teacher survey and interventionist survey.

NOTES: "Small-group instruction" refers to services provided by teachers during the core reading block to all students. Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. The Below-Only school sample represents schools that had at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services. No tests were performed between intervention groups in Below-Only schools, which do not provide intervention to At or Above grade-level groups. Means reflect rounding.

Statistical significance is indicated as follows: *** at the $p \leq 0.001$ level, ** at the $p \leq 0.01$ level, and * at the $p \leq 0.05$ level.

The school sample sizes are, for small-group instruction, 67 Below-Only schools and 51 All-Level schools; for intervention, 72 Below-Only schools and 59 All-Level schools.

Table 4.6

Service Contrast for Minutes per Week: Difference Between Groups At or Above Grade Level and Below Grade Level in Below-Only and All-Level Schools, by Grade

Minutes per Week	Groups At or Above Grade Level	Groups Below Grade Level	Mean Differences Between Groups	P-Value
Grade 1				
Below-Only Schools				
Small-group instruction	62	89	27	0.000
Intervention	NA	167	NA	NA
All-Level Schools				
Small-group instruction	100	140	39	0.000
Intervention	160	155	-5	0.353
Grade 2				
Below-Only Schools				
Small-group instruction	59	90	31	0.000
Intervention	NA	182	NA	NA
All-Level Schools				
Small-group instruction	89	119	30	0.000
Intervention	151	176	25	0.000
Grade 3				
Below-Only Schools				
Small-group instruction	61	93	31	0.000
Intervention	NA	184	NA	NA
All-Level Schools				
Small-group instruction	76	101	26	0.000
Intervention	143	165	23	0.001

SOURCES: Instruction questions from the teacher survey and intervention questions from the interventionist survey.

NOTES: "Small-group instruction" refers to services provided by teachers to all students during the core reading block. Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. The Below-Only school sample represents schools that had at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services. No tests were performed between intervention groups in Below-Only schools, which do not provide intervention to At or Above grade-level groups. Means reflect rounding.

(continued)

Table 4.6 (continued)

The school sample sizes were as follows:

Grade 1: Below-Only schools had 67 schools in small-group instruction and 72 schools in intervention. All-Level schools had 51 schools in small-group instruction and 59 schools in intervention.

Grade 2: Below-Only schools had 76 schools in small-group instruction and 81 schools in intervention. All-Level schools had 41 schools in small-group instruction and 45 schools in intervention.

Grade 3: Below-Only schools had 72 schools in small-group instruction and 85 schools in intervention. All-Level schools had 36 schools in small-group instruction and 39 schools in intervention.

Grades 2 and 3. In All-Level schools, intervention time in Grades 2 and 3 was offered for more than 20 additional minutes per week to groups below grade level than to AA reading groups.

Note that, for several reasons, this analysis does not add small-group instruction time to intervention time to obtain a total "dosage" of reading instruction or intervention time. First, intervention groups cannot be linked to small instruction groups, to determine who received instruction plus intervention and who received instruction only. Second, some intervention groups occurred during the core and may have displaced some instruction time, rather than supplementing it.[103] Third, intervention groups served fewer students than small instruction groups, so the intensity of time was greater for the students in those groups than in small instruction groups. Thus, a simple addition of time would be misleading.

Mechanisms to Vary Support Services

Staff Specialization of Interventionists

Of six staff types listed on the survey that interventionists could have chosen, four of the types are specialized: reading specialist, special educator, teacher of English Language Learners, and speech pathologist. The two remaining staff types are paraprofessionals and classroom teachers. In the latter case, teachers completed both a teacher survey and an interventionist survey if they served in both roles. When teachers and others completed an interventionist survey, they were asked to describe services provided above and beyond core instruction services (thereby distinguishing it from the teacher survey). In prior studies, more specialized staff provided more specific services, and nonteaching staff potentially could have provided more time.

[103]To assess whether reading intervention took time away from other subjects, the study research team examined teacher survey responses about what subjects students with a reading IEP (which include many FB students) miss when they receive additional instruction. No single subject or pattern of subjects was missed more often than others.

In this way, there may have been some interaction between staff type and time that may have changed the service intensity.

- **Specialized staff in impact sample schools served a large proportion of intervention groups below grade level, but classroom teachers and paraprofessionals served the majority of all intervention groups.**

Teachers and paraprofessionals served the majority of reading groups in All-Level schools — up to 77 percent of groups at or above grade level and up to 64 percent of groups below grade level. In Below-Only schools, up to one-third (36 percent) of groups below grade level were served by classroom teachers or paraprofessionals.[104] These results suggest that non-specialized staff may have played more of a role in All-Level schools, while specialized staff may have played more of a role in Below-Only schools, which follow a more traditional system of delivering interventions.

The presence of teachers as interventionists could have been related to such factors as some schools organizing intervention during the core and providing intervention for at least some students at all reading levels; and, as shown in Chapter 3, that teachers were used in a variety of roles.

The contrast between groups at or above and below grade level was restricted to All-Level schools, because only they used interventionists to serve both levels of reading groups. In Grade 1, the differences in the percentage of groups served by particular staff differ for three staff types. Table 4.7 shows that 31 percent of groups at or above grade level were served by paraprofessionals, compared with 37 percent of groups below grade level. The next row shows that classroom teachers served 42 percent of AA groups, compared with 26 percent of groups below grade level. Reading specialists served 6 percent of AA groups, compared with 14 percent of groups below grade level. The remaining specialized staff types served similar percentages of groups at both reading levels; the percentages are not significantly different.

Progress Monitoring

Impact sample schools were selected based on whether they used progress monitoring to assess the effectiveness of interventions for readers below grade level. This process involved a brief[105] but frequent assessment of struggling students' progress in oral reading fluency or word identification fluency in Grade 1 or Grade 2. The IES Practice Guide on RtI for Elementary Reading recommends at least monthly progress monitoring for readers below grade level,

[104]Results are shown in a footnote of Appendix Table C.9.
[105]These assessments took no longer than five minutes.

Table 4.7

Service Contrast for Interventionist Staff Specialization: Difference Between Groups At or Above Grade Level and Below Grade Level in Intervention Groups for All-Level Schools, for Grade 1

	Groups At or Above Grade Level	Groups Below Grade Level	Mean Difference Between Groups	P-Value
Percentage of groups as served by staff type				
Paraprofessional	30.7	37.1	6.4	0.078
Classroom teacher	41.6	25.7	-15.9	0.000
Reading specialist	5.8	14.1	8.3	0.001
Special educator	4.3	7.3	2.9	0.121
English Language Learner teacher	6.5	5.9	-0.6	0.695
Speech pathologist	0.0	0.7	0.7	0.314
Other	11.1	9.3	-1.8	0.490

SOURCE: Interventionist survey.

NOTES: Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services.

Out of 59 All-Level schools, 57 had responses for this item.

with more monitoring for those reading Far Below grade level. (Because the guide does not recommend monitoring AA groups, the survey did not ask about monitoring them.)[106]

- **Progress monitoring in impact sample schools occurred as often as three times per month for students Far Below grade level.**

Table 4.8 shows the percentage of schools using key tests to conduct progress monitoring for Grade 1, and the frequency of monitoring students on those tests, as reported by interventionists. Key monitoring tests include oral reading fluency tests, curriculum-embedded tests (which are integrated into classroom tasks), and the so-called "running records test," in which a student reads out loud a text considered appropriate for that student's age or grade and an

[106]Gersten et al. (2009).

The Response to Intervention (RtI) Evaluation

Table 4.8

Average Frequency of Progress Monitoring for Intervention Groups Below Grade Level, for Grade 1

	Below-Only Schools						All-Level Schools					
		Of Schools Using Each Test, the Frequency of Monitoring per Month						Of Schools Using Each Test, the Frequency of Monitoring per Month				
	Schools Using This Test (%)	Somewhat and Far Below Grade Level	Somewhat Below Grade Level	Far Below Grade Level	Mean Difference Between Groups	P-Value	Schools Using This Test (%)	Somewhat and Far Below Grade Level	Somewhat Below Grade Level	Far Below Grade Level	Mean Difference Between Groups	P-Value
<u>Number of Times per Month for Each Type of Test</u>												
Oral reading fluency	93	3.3	3.0	3.5	0.6	0.002	95	3.3	3.4	3.4	0.0	0.931
Curriculum-embedded tests	63	1.2	1.2	1.3	0.1	0.559	80	1.8	1.9	2.1	0.2	0.161
Running records	78	2.3	2.1	2.8	0.7	0.002	83	2.6	2.5	2.6	0.1	0.506

SOURCE: Interventionist survey.

NOTES: The means presented for Far Below groups are regression-adjusted, and school fixed effects are used to account for clustering of groups within schools. The means presented for Somewhat Below groups are simple means. The percentages of schools using the test and the means reflect rounding.

At least one interventionist had to report using any of the three tests in order for a school to be included in the analysis. There were 72 Below-Only schools serving Grade 1 groups and 59 All-Level schools serving Grade 1 groups.

assessor monitors how many words the student reads correctly and how many specific challenges the student encounters. Oral reading fluency was used by more than 90 percent of schools in both samples. Curriculum-embedded tests and running records tests were used to a lesser extent, but by the majority of schools in both samples.

For below-grade-level readers, monitoring of student reading progress using oral reading fluency tests occurred three or more times per month in both Below-Only and All-Level schools. Monitoring of below-grade-level readers using curriculum-embedded tests occurred less than twice per month, while running records tests were used more than twice per month in both types of schools. In Below-Only schools, students Far Below grade level were monitored slightly more frequently for oral reading fluency (3.5 times per month) and by using running records tests (2.8 times per month) than were students Somewhat Below grade level (3.0 times and 2.1 times per month, respectively). This monitoring occurred more often than is recommended by the IES Practice Guide on Elementary Reading, and it confirms increased monitoring for readers Far Below grade level in schools providing intervention services to only the reading groups that were below grade level.

Reading Skills Covered

The surveys asked about five reading skills, often addressed in Grades 1 to 3: fluency, reading comprehension, vocabulary, phonics, and phonemic awareness. In both the teacher and the interventionist survey, respondents could have selected multiple reading skills for a given session. *Note that these surveys did not collect information on the amount of session time spent on any particular reading skill and, thus, cannot answer whether groups spent more time on one skill than another.*[107] The analysis treats each reading skill separately and compares the percentages of groups between reading levels that addressed that skill. Discussion focuses only on statistically significant differences.

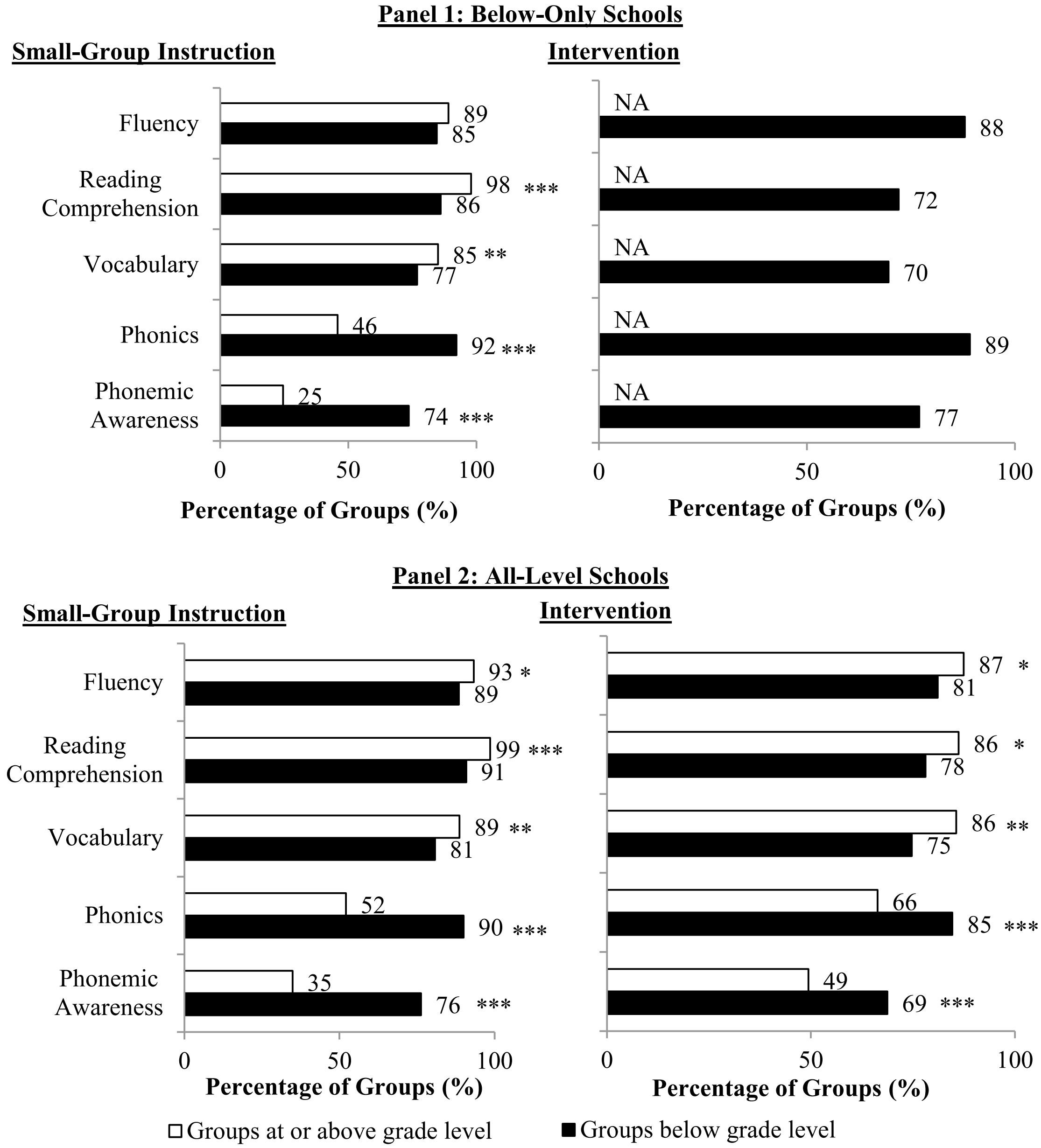
- **In impact sample schools, a larger percentage of intervention groups below grade level addressed phonics and phonemic awareness, while a larger percentage of instruction groups at or above grade level addressed fluency, reading comprehension, and vocabulary. These latter skills, however, were addressed in most reading groups, regardless of level.**

Figure 4.6 displays the percentage of groups that touched on each of the five reading skills during a group session for Grade 1. Panel 1 shows results for Below-Only schools,

[107]Teachers were asked about the "content focus" of the instruction group session, and interventionists were asked which reading components were "emphasized" during the session.

Figure 4.6

Service Contrast for Reading Skills: Difference Between Groups At or Above Grade Level and Below Grade Level in Below-Only and All-Level Schools, for Grade 1, by Reading Skill Targeted



(continued)

Figure 4.6 (continued)

SOURCES: Teacher survey and interventionist survey.

NOTES: "Small-group instruction" refers to services provided by teachers during the core reading block to all students. Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. Respondents to each survey could indicate any one or any combination of content foci. As a result, the percentage of groups indicating one content area could overlap with another content area. Means reflect rounding. The Below-Only school sample represents schools that had at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services. No tests were performed between intervention groups in Below-Only schools, which do not provide intervention to At or Above grade-level groups.

Statistical significance is indicated as follows: *** at the $p \leq 0.001$ level, ** at the $p \leq 0.01$ level, and * at the $p \leq 0.05$ level.

The numbers of schools represented in small-group instruction are 67 Below-Only schools and 51 All-Level schools. The numbers of schools represented in intervention are 72 Below-Only schools and 59 All-Level schools.

including core reading small-group instruction (left-hand side) and intervention (right-hand side). Panel 2 shows a similar breakout for All-Level schools. In each panel, the five reading skills are shown, with results for AA groups (white bars) and for groups below grade level (black bars).

Among small instruction groups in Grade 1 in Below-Only schools (Panel 1), a larger percentage of groups below grade level reported touching on phonemic awareness and phonics than the percentage of AA groups (74 percent and 25 percent, respectively, and 92 percent, compared with 46 percent). While 85 percent of AA groups addressed vocabulary, 77 percent of groups below grade level did. And while nearly all AA groups addressed reading comprehension (98 percent), 86 percent of groups below grade level touched on this skill. Among intervention groups in Below-Only schools, the magnitude of the percentage of groups below grade level that addressed these skills is similar to the mean levels in instruction groups (the black bars on the left-hand and right-hand portions of that panel).

In All-Level schools for Grade 1 (Panel 2), the magnitude of the means and differences between groups at or above and below grade level are similar to the differences in the Below-Only schools for small-group instruction. The mean proportion of intervention groups below grade level that touched on each reading skill is similar in All-Level and Below-Only schools. The percentage of below-grade-level groups that addressed fluency and reading comprehension exceeded 78 percent of instruction and intervention groups in all the impact sample schools. The majority of Grade 1 AA groups in all schools reported that they addressed vocabulary, reading comprehension, and fluency — the last outcome was one of those tested in the impact analysis for that grade. A higher percentage of groups below grade level than at or above grade level addressed phonics and phonemic awareness. These findings suggested that there were differences in emphasis of some skills between reading levels and that groups below grade level were not focusing on just one skill.

Discussion and Conclusion

This chapter shows that, despite variation, impact sample schools followed key practices consistent with the RtI framework and prior literature. Schools followed the RtI framework described in prior literature in several ways: they distributed students across reading levels according to the concept of multiple tiers (which permits intensification of services for struggling readers); they moved students between reading levels after universal screening assessments; and they offered more intense services to readers below grade level. Across a variety of intensity dimensions, the evidence is consistent with the hypothesis of increasing intensity of services for those with reading difficulties.

The variation in organizing services for reading groups reflects three major differences from prior literature, which Figure 1.2 displays. First, prior studies that designed or monitored the delivery of intervention services usually served students placed in Tiers 2 or 3, that is, students reading below grade level. This study found that some schools offered intervention services to at least some reading groups at all reading levels (though not necessarily intervention services for all at-or-above-grade-level students). A greater percentage of schools in the Grade 1 sample did this than in Grades 2 and 3. However, the study was not able to link reading group data to individual-level service data to determine the characteristics of students receiving intervention services (including students reading at or above grade level), exactly what services were provided to these students, or whether services differed for At or Above, Somewhat Below, and Far Below grade-level reading groups.

Second, previous studies of small-group intervention services often designed intervention as supplemental to the core reading block time. This study found that the majority of schools offered at least some intervention services during the core. In such schools, intervention may have displaced instruction time and replaced some small-group or other instruction services with intervention services. As a result, not all students were receiving intervention *in addition to* the full core reading time.

Third, in contrast to more controlled studies of RtI that have relied on non-classroom teaching staff to provide intervention services, the current study included intervention services provided by whoever was designated by schools to provide these services. This study found that up to 47 percent of schools in Grade 1 used classroom teachers to do so. Schools that used this model used teachers in multiple roles. The reason this occurred — whether because it was a strategy to help align instruction with intervention, or for other reasons — could not be answered with the available data. These results suggest that schools adapted time and staff resources to address students' needs within an RtI framework.

These differences may have reflected either limitations of the RtI framework described in prior literature or adaptation of that model to reflect the needs of students and availability of resources. Schools faced constraints during the study year of 2011-12, when many states and school districts experienced budget cuts or reductions in services.[108] At least half the states in the study experienced a decline in per pupil expenditures between Fiscal Years 2010 and 2011.[109] Despite the widespread adoption of RtI documented in Chapter 3, there was not a prescribed RtI curriculum or a single vision of RtI against which to monitor implementation of the framework at the school level.

This chapter has described how schools delivered reading-group services — which depended on staff and time allocations — in real-world settings during 2011-12. Survey questions regarding the nature and purpose of these interventions were focused on a select number of key characteristics (group size, minutes offered, reading skills addressed, staff specialization, and progress monitoring) that prior literature had identified as critical. As a result, the analysis is limited in several ways. It cannot describe other aspects of the interventions, such as intervention quality and alignment with the core curriculum. In addition, data limitations do not permit the study to link reading-group data to individual-level service data to determine the characteristics of students receiving intervention services (including students reading at or above grade level), exactly what services were provided to these students, or whether services differed for At or Above, Somewhat Below, and Far Below grade-level reading groups. Finally, the data collected do not provide school leaders' reasons for providing interventions to At or Above reading groups. Further data collection would be required to address these issues.

The descriptive results showing increased service intensity for readers below grade level provide motivation for the impact analysis in Chapter 5. The impact analysis narrowed its focus to those students who scored near a threshold value (cut point) on a screening test and compared those who received intervention services with those who did not. While the contrast in services described in this chapter is not limited to those students near the threshold value, the averages for groups at or above grade level and below grade level shed light on services offered to students near either side of the cut point.[110]

---

[108]Johnson, Oliff, and Williams (2011).

[109]Corman (2013).

[110]The findings in this chapter are based on reading-group-level data, which is not linked to individual student-level data. Thus, the analysis cannot restrict comparisons to reading groups serving students near the cut point. In a randomized controlled trial, one would not need to restrict the comparisons, because generally all students are included in the estimated average service contrast as well as the average impact.

This page has been left blank for double-sided copying.

Chapter 5

Primary Impact Findings

Previous chapters in this report describe the prevalence of Response to Intervention (RtI) practices in the impact sample schools and how the intensity of reading interventions varied with student reading levels, which correspond to the second and third columns of the logic model for this evaluation (Chapter 1, Figure 1.3). Chapter 5 examines the effect on reading achievement outcomes of actually being assigned to receive such interventions for students on the margin of being identified as at risk of reading failure, and tests whether reading outcomes improved for these students, as hypothesized in the rightmost column of the logic model presented in Figure 1.3 (p. 12). Specifically, this chapter addresses the following question:

> ***Research question:*** *For students who fell just below school-determined standards for each grade on screening tests: What were the effects on reading achievement of actual assignment to receive reading intervention services (in addition to core instruction)?*

This question was addressed by comparing children who were assigned to Tier 2 or Tier 3 as their highest tier placement with similar children just above the cut point who were assigned to receive only Tier 1 reading instruction during school year 2011-12.[111]

The primary findings from the present study show that:

- Actual assignment to receive Tier 2 or Tier 3 intervention services had a negative impact on a comprehensive reading measure for Grade 1 students near the cut point.[112] The magnitude of the negative effect is roughly equivalent to about one month of learning for first-graders. The estimated impact on a decoding fluency measure for Grade 1 students who were close to the cut point is also negative but is not statistically significant.

[111] As discussed in Chapter 1 and depicted in Figure 1.1, RtI is often portrayed as a pyramid: all students receive core instruction (Tier 1) as the foundation, and students who read below grade level receive more support (Tier 2 or Tier 3) in addition to Tier 1, as needed.

[112] Although there is a rough equivalence between Tier 2 and reading Somewhat Below grade level, the cut point that individual schools set does not always correspond to standards of reading at grade level. As a result, the cut point to assign students to Tier 2 may not always correspond to reading just below grade level. Similarly, the cut point used by some schools to assign students to Tier 3 may not always correspond to reading far below grade level.

- The estimated impacts on the decoding fluency measure for Grade 2 students near the cut point and on the comprehensive reading measure for Grade 3 students near the cut point are not statistically significant.

This chapter starts by briefly describing the design, the analytic methods, and the samples used for the impact analysis. It then presents the primary findings for the overall student sample, and it concludes by placing the findings in the context of recent literature on this topic.

Study Design, Analytic Approach, Data, and Sample

Study Design and Interpretation of Results

This study used a Regression Discontinuity (RD) design to assess the effect of assignment to reading intervention services on the performance of early-grade students reading below grade level. The way that RtI schools in the study's impact sample identified students who struggled with reading and assigned them to more intense reading interventions created the opportunity to use such a design. As discussed in Chapter 1, in the impact sample schools, student reading performance was assessed by a screening test at the beginning of a school year. A set of predetermined rules was then used to decide students' tier placement. According to these rules, students in each school whose screening test score (known as the "rating variable" in the RD design literature) fell at or below a certain cutoff score (the "cut point") set forth by each school were considered at risk and, therefore, were assigned in the fall to either Tier 2 or Tier 3 to receive increasingly more intense reading interventions above and beyond the core reading instruction that serves all students. These at-risk students constitute the *treatment group* in the study. On the other hand, students in each school whose screening scores were above the cut point were assigned to Tier 1 to receive only the core reading instruction; these are the *comparison group* students.[113] This RD design allows assessment of the program impact at the cut point for Tier 2 intervention between students placed only in Tier 1 and students placed in Tier 2 or Tier 3. Most but not all of the students with scores just below the cut point were placed in Tier 2, and most of the students with scores just above the cut score were placed in Tier 1 only. The analysis does not distinguish between the impacts of Tier 2 and Tier 3 services because the criteria used for Tier 3 placement (in other words, an additional cut point or other criteria used to determine Tier 3 rather than Tier 2 placement) were not clearly documented by schools and because the number of Tier 3 students is too small to provide adequate statistical power for meaningful estimation.

[113]The study research team screened schools to make sure that all schools in the impact analysis sample used such a process to assign students into treatment or comparison conditions. As a result, an RD design that pools students across schools is feasible for this analysis. Appendix D provides details on the screening tests and decision rules used by sample schools to assign students to tiers.

In an ideal situation where a student's assignment to intervention is completely determined by the decision rules, the *intended* assignment to intervention services (as determined by the rules) is equivalent to the *actual* assignment to intervention, and this design could be considered a "sharp" RD. In this case, the impacts of *being assigned to* reading intervention on students at the cut point can be obtained by comparing the outcomes of students just below and just above the cut point, after adjusting for the rating variable values.[114] Under plausible assumptions, this approach can generate unbiased estimates of the impacts of assignment to Tier 2 or Tier 3 intervention services for students near the cut point in this situation.

All impact sample schools were selected largely based on whether they had clearly stated decision rules for a student's tier assignment and how well they followed their stated rules in placing students into different tiers. (See Chapter 2 and Appendix B for the sample selection process.) Nonetheless, not all schools in the analysis sample adhered to their stated decision rules when actually assigning students to tiers. The rules were amended in some schools due to resource constraints, teachers' judgments, or other reasons.[115] As a result, students' actual assignment to Tier 2 or Tier 3 intervention services deviated from their intended assignment. Figure 5.1 presents the relationship between the intended assignment to intervention and the actual assignment of students' tier placement in the fall, as defined in Chapter 4. As mentioned above, the intended assignment of students to treatment or comparison conditions depended entirely on their screening test score (rating) and the school's decision rules. The actual tier placement for each student in the fall semester, on the other hand, could deviate from the intended assignment. The students whose actual assignments were consistent with their intended assignments are referred to as the "compliers." Some students had screening test scores that fell below the cut point, but they did not end up in Tiers 2 or 3. These students are referred to as the "no-shows." Other students scored at or above the cut point, but they ended up in Tiers 2 and 3. These students are referred to as the "crossovers" with regard to the decision rules. The existence of no-shows and crossovers in this context means that membership in Tier 2 or 3 reading intervention groups was not determined by screening test scores only but also by such additional factors as the professional judgment of school staff.

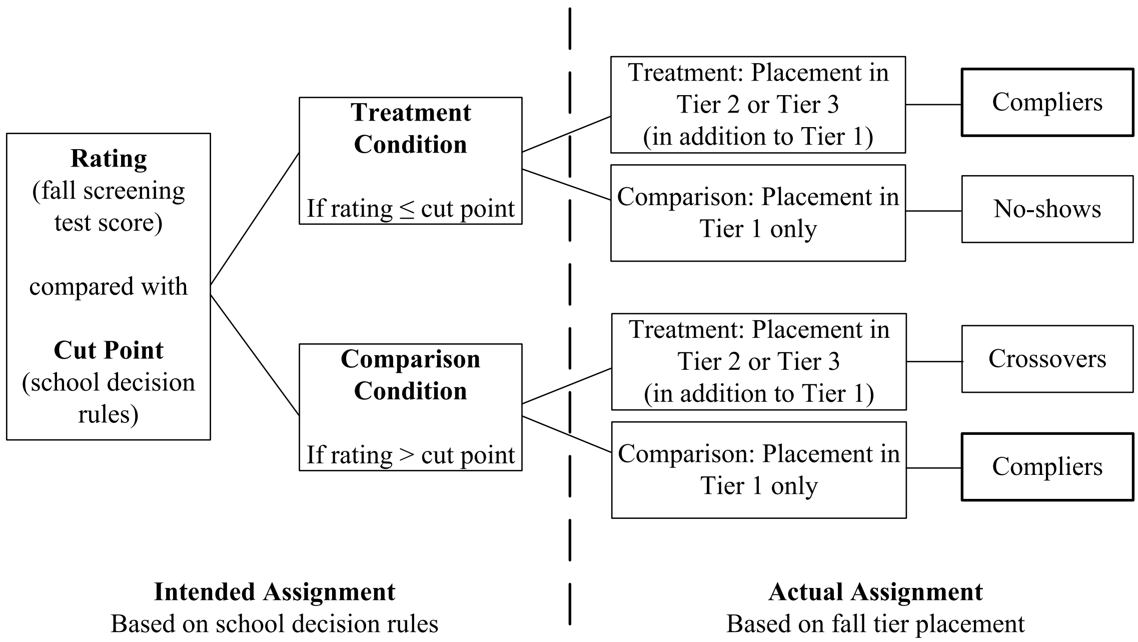
Figure 5.2 shows the percentages of no-shows and crossovers in the full impact sample, by grade and by intended treatment assignment. The two bars in each grade's graph represent the intended comparison group (on the left) and the intended treatment group (on the right). The

[114]Specifically, for schools using multiple screening tests to determine a student's tier placement in the fall, a "binding score" approach (Reardon and Robinson, 2012; Wong, Steiner, and Cook, 2013) is used. This approach allows the collapsing of multiple dimensional ratings into one dimension and, hence, allows the pooling of data across schools using different decision rules. See Appendix D for more detailed discussion of this approach and other related issues.

[115]Appendix D examines the schools' decision rules.

Figure 5.1

Intended Assignment Relative to Actual Assignment to Receive Tier 2 or Tier 3 Intervention Services

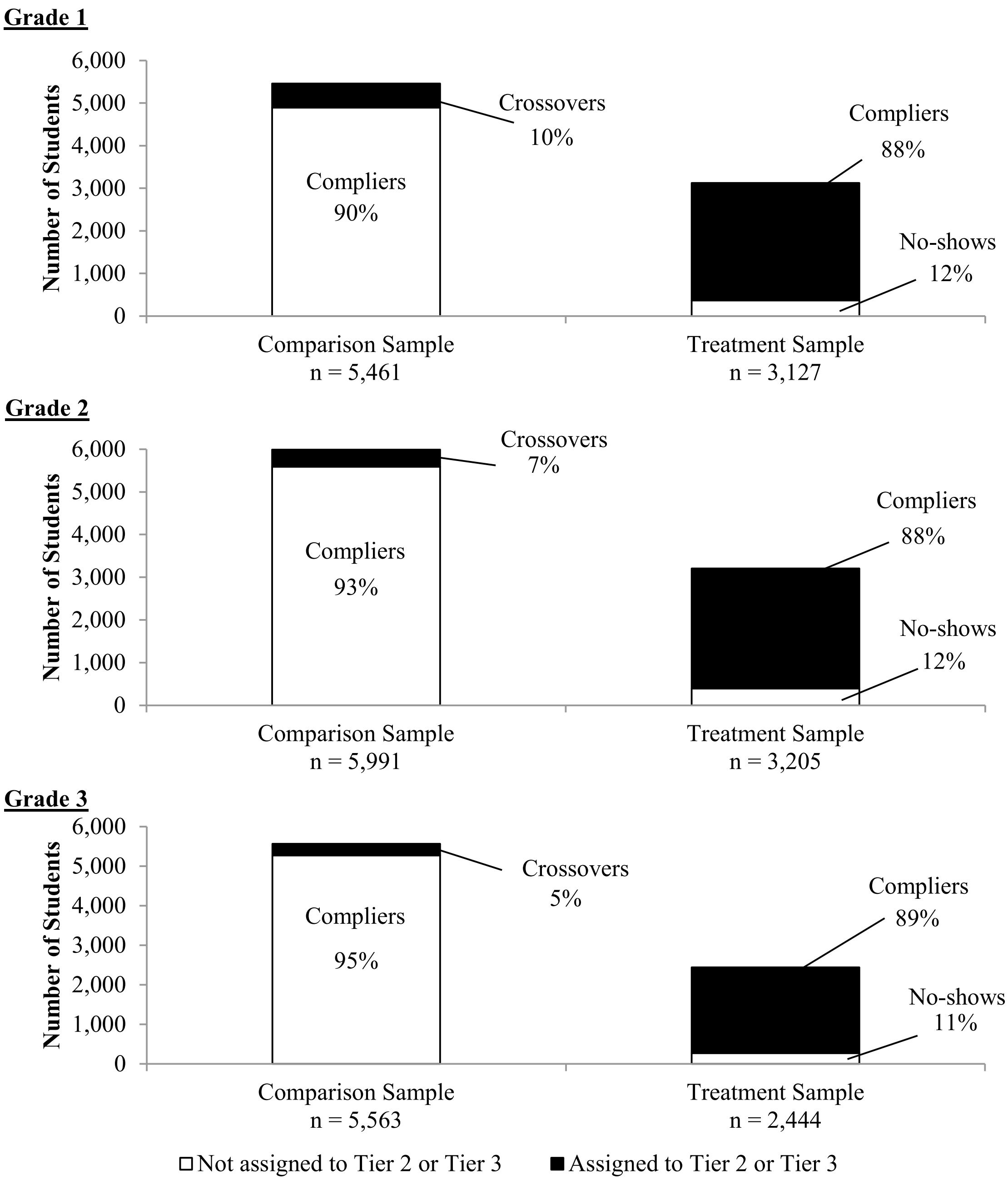


NOTE: The students whose actual assignments were consistent with their intended assignments are referred to as the “compliers.” Some students had screening test scores that fell at or below the cut point, but they did not end up in Tiers 2 or 3. These students are referred to as the “no-shows.” Other students scored above the cut point, but they ended up in Tiers 2 and 3. These students are referred to as the “crossovers” with regard to the decision rules.

height of the bars represents the sample size for each group. The white portion of each bar stands for the percentage of students within that group who were not actually in Tier 2 or Tier 3, and the dark portion stands for the percentage of students within the group who were actually in Tier 2 or Tier 3 in the fall. In other words, for the comparison groups, the dark portion represents the crossovers; for the treatment groups, the white portion represents the no-shows. This figure shows that the average percentage of no-shows and crossovers varies for the full sample across grades, with the percentage of no-shows being stable at around 11 percent to 12 percent and the percentage of crossovers ranging from 5 percent to 10 percent.

Figure 5.2

Percentage of Compliers, Crossovers, and No-Shows, by Intended Treatment Status, by Grade



(continued)

Figure 5.2 (continued)

SOURCES: Fall screening scores and tier placement data from schools in the sample.

NOTES: The data report on students in the analytic sample.

Grade 1 numbers represent students in the sample who completed the ECLS-K Reading Assessment.

The numbers of schools are 119 for Grade 1, 127 for Grade 2, and 112 for Grade 3. The number of students per sample is indicated in the figure as "n."

Percentages reflect rounding.

The existence of no-shows and crossovers makes the design of the present study deviate from the sharp RD design.[116] Fortunately, under certain additional assumptions,[117] one can reliably estimate the effect of actual assignment to Tier 2 or Tier 3 intervention services for students close to the cut point whose assignment complied with the intended assignment to intervention services based on the decision rules.

The nature of this design puts limitations on the generalizability of the impact findings. First, unlike a randomized controlled trial (RCT), in which the estimated average treatment effect applies to all students in the study, the RD impact applies only to students near the cut point of the rating variable. In other words, the RD analysis examines whether there is a discontinuity in the relationship between the rating variable (fall screening test score) and the outcome (follow-up spring reading test score) *at the cut point value.* This estimate does not necessarily represent the impact of intervention on students far away from the cut point of the rating variable.[118] However, this is not to say that the results should be generalized to Tier 2 students only. While, for each grade from 1 to 3, Tier 2 students' rating values were centered closer to the cut point than those of Tier 3 students, there were both Tier 2 students and some Tier 3 students whose rating values were close to the cut point.[119]

Second, the deviation of students' *actual* assignment to Tier 2 or Tier 3 intervention services from their *intended* assignment means that the estimated local average treatment effect

[116]This design is referred to in the RD literature as the "fuzzy" RD design. See Bloom (2012) and Lee and Lemieux (2010) for detailed discussions of this design and the estimation methods suitable for it.

[117]The assumptions focus on the validity of the instrumental variables (in this case, the treatment assignment) and are described in Angrist, Imbens, and Rubin (1996). Briefly, the assumptions are (1) a monotonicity condition: scoring at or below the cut point on the screening test increases the probability of actually being placed into intervention and scoring above it increases the probability of being placed in Tier 1; (2) a strong instrument: the correlation between the instrument and placement into intervention is sufficiently large; and (3) the exclusion restriction: that the effect of treatment can occur only through actually being placed in intervention. Appendix E provides more detailed discussion of this assumption and demonstrates that the data used here satisfy this requirement.

[118]Bloom (2012); Schochet et al. (2010); Cook (2008); Imbens and Lemieux (2008).

[119]The last section of Appendix E provides more detailed discussion on this issue and displays the distribution of rating values by tier and by grade (Appendix Figure E.3).

(LATE) of actual assignment to reading intervention is applicable only to the compliers; in other words, it applies to students who were identified for assignment to Tiers 2 and 3 and who, indeed, ended up in those tiers. As a result, the impact findings reported in this chapter are applicable only to students who were close to the cut point of being identified for intervention and whose assignment complied with the decision rules for tier assignment.

These limitations on the generalizability of impacts also affect the interpretations and policy implications of the findings. Given that the impact findings based on this RD design are generalizable only to students at or close to the cut point whose assignment complies with the tier-assignment decision rules, they provide estimates of the average effect of intervention for students who would be added or dropped by marginally changing the eligibility criterion. In this sense, these results are relevant for decisions about expanding or reducing the scope of intervention (for example, perhaps by shifting the cut point) but are not necessarily relevant for decisions about offering or not offering intervention.

Analytic Approach

Intuitively, under the RD design, the impact of Tier 2 or Tier 3 reading intervention services can be obtained by comparing the outcomes (follow-up test scores) of students just below and just above the cut point (based on school decision rules), after adjusting for the rating variable (fall screening test scores). Analytically, it is challenging to correctly adjust for the rating variable without causing bias in the impact estimates. One often-used and recommended approach to deal with this issue is to choose a small neighborhood (known as "bandwidth") to the left and right of the cut point and use only data within that range to estimate the jump in outcomes at the cut point.[120] Following this approach, for each grade and outcome measure, an *optimal bandwidth* was pre-selected to minimize the potential bias in estimation and to maximize the statistical precision of the estimation.[121] Appendix E provides details about this recommended approach and the modifications that the research team made to adapt the method for the present study.

It is important to note that even though the impact estimations presented here use all students above or below the cut point whose rating values fell within the pre-selected optimal bandwidth, it does not mean that the estimated impacts can be generalized to all students. As stated above, the estimated RD impact represents a *local* effect that applies only to a subset of all students within the optimal bandwidth whose ratings were at or close to the cut point. There-

[120]See, for example, Schochet et al. (2010) and Bloom (2012).

[121]In general, choosing a bandwidth in this context involves finding the best balance between precision and bias. Although using a larger bandwidth yields more precise estimates — given that more data points are used in the estimation — the model specification is less likely to be accurate, which could lead to larger bias; using a smaller bandwidth is less prone to bias but also less precise. The recommended approach provides a mathematical solution for the bandwidth that minimizes a particular function of bias and precision.

fore, it may not necessarily apply to students whose ratings (fall screening test scores) are far away from the cut point. Given the fact that Tier 2 students tend to have higher fall screening test scores — and hence ratings closer to the cut point — than Tier 3 students, whose fall screening test scores are more likely to be farther away from the cut point, the impact findings reported below are least likely to be applicable to any student whose rating is far from the cut point, a group who are mostly Tier 3 students. (See Appendix E for detailed distribution of rating values by actual tier placement.) To provide context to the primary findings, the study research team used this approach to estimate the average effects of the *intended* assignment and the *actual* assignment to Tier 2 or Tier 3 reading intervention services. The latter estimate is the focus of discussion below.[122]

The study research team assessed the validity of the RD design used in this study by examining the assumptions of the RD design. These validity checks addressed (1) the continuity of the rating variable, (2) the integrity of the rating variable and the treatment assignment process, and (3) the possible influence of data "heaping."[123,124] Appendix F describes these verification tests in more detail and demonstrates the validity of the current design.

The study research team also assessed the robustness of the primary impact findings by estimating the impacts with alternative bandwidth selections, with alternative model specifications, and with alternative sample specifications. The results from these tests are presented in Appendix G and generally suggest that they are not sensitive to these alternative specifications.

Student Samples

As described in Chapter 2, there are 146 RtI schools with one or more grade levels among Grades 1 to 3 eligible for the RD design. By grade, the numbers of eligible schools are 119 for Grade 1, 127 for Grade 2 , and 112 for Grade 3. All students in a given grade in the eligible schools with nonmissing values for the grade-specific outcome measures, the rating variable, and the treatment receipt status are included in the analysis sample for that grade.[125]

---

[122]The realized minimum detectable effect sizes (MDES) based on this model and the realized samples are reported in Appendix E.

[123]In this context, "data heaping" refers to the phenomenon that multiple observations share a unique rating value due to data rounding or data discretization.

[124]Barreca, Lindo, and Waddell (2011).

[125]Note that a total of 325 students in self-contained special education classes were also excluded from the analysis sample. These students were removed from analyses because the kind of reading instruction that they experienced was different from instruction for students in regular classrooms. They did not receive core instruction that all other students received and, therefore, could not be used for the comparison of Tier 1 (core instruction only) with Tier 2 and Tier 3 students (core instruction plus intervention services).

Table 5.1 presents the baseline characteristics of the three student samples in each grade. The first pair of columns shows the mean characteristics of the full sample; the middle pair of columns reports information for the subsample of all students within the selected optimal bandwidth. This is the sample that was used for the preferred impact estimation method. The rightmost pair of columns presents information about the subsample of compliers to the decision rules within the selected optimal bandwidth. The subsample of compliers at or around the cut point is the sample for which the primary impact estimates are most relevant.[126]

Table 5.1 shows that student characteristics are fairly consistent across these samples as well as across grades. Overall, more than 60 percent of students in the sample are white, non-Hispanic; about 35 percent to 45 percent of them had low-income status; 6 percent to 13 percent of the students were English Language Learners; around 10 to 12 percent had an Individualized Education Program; and a small portion (5 percent to 7 percent) were overage for grade.[127]

Primary Impact Findings

The focus of the impact analysis in this study is to assess the effects of actual assignment to Tier 2 or Tier 3 reading intervention services on students with difficulty reading in early grades. Two kinds of effects — the effect of intended assignment to intervention (based on the decision rules) and the effect of actual assignment to intervention — were examined for this purpose. Table 5.2 presents the estimation results for both effects, using the analytic approach described above. The findings are fairly consistent across these two sets of estimates, so the discussions below focus on the estimates for the impact of *actually* being placed in intervention services.

- **There is a statistically significant and negative effect of assignment to Tier 2 or Tier 3 intervention services on the comprehensive reading measure for Grade 1 students whose ratings were around the cut point. The estimated effect on the measure of decoding fluency is also negative but not statistically significant.**

The estimate for the effect of assignment to Tier 2 or Tier 3 intervention services on the ECLS-K Reading Assessment is –0.17 standard deviation and is statistically significant (p-value = 0.002). The estimate for the effect of treatment assignment on the TOWRE2 Sight Word

[126]Note, however, that this is *not* the sample to which the findings should be generalized. As discussed above, the impact findings are applicable only to the compliers whose rating values are at or around the cut point.

[127]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

The Response to Intervention (RtI) Evaluation

Table 5.1

Average Background Characteristics of the Full Sample and Two Subsamples Within Optimal Bandwidth, by Grade

	All Students		Students Within Optimal Bandwidth		Compliers[a] Within Optimal Bandwidth	
Grade/Characteristic	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Grade 1						
Age (years)	6.5	0.34	6.5	0.35	6.5	0.34
Low-income students (%)	42.3	49.40	43.8	49.62	42.9	49.49
Race/ethnicity (%)						
White, non-Hispanic	61.9	48.58	62.5	48.42	62.3	48.48
Black, non-Hispanic	6.5	24.62	6.1	23.84	5.9	23.61
Hispanic	22.0	41.43	22.4	41.68	22.7	41.88
Asian/Pacific Islander	5.6	23.01	4.8	21.45	5.1	21.89
Other	3.8	19.18	4.0	19.66	3.9	19.31
Male (%)	50.8	50.00	50.6	50.00	50.6	50.00
English Language Learners (%)	12.8	33.43	12.8	33.36	13.2	33.87
Students with IEPs[b] (%)	9.6	29.40	9.8	29.77	10.0	29.94
Overage for grade[c] (%)	5.4	22.64	5.4	22.63	5.4	22.68
Number of observations	8,342		6,049		5,219	
Grade 2						
Age (years)	7.5	0.35	7.5	0.35	7.5	0.35
Low-income students (%)	38.6	48.68	42.4	49.43	42.3	49.40
Race/ethnicity (%)						
White, non-Hispanic	66.3	47.28	65.9	47.42	65.6	47.52
Black, non-Hispanic	6.2	24.06	7.1	25.63	6.7	25.01
Hispanic	17.8	38.21	18.6	38.90	18.9	39.16
Asian/Pacific Islander	5.4	22.65	3.9	19.28	4.1	19.82
Other	4.0	19.70	4.3	20.19	4.4	20.47
Male (%)	50.9	49.99	51.6	49.98	51.5	49.98

(continued)

Table 5.1 (continued)

Grade/Characteristic	All Students Mean	All Students Std. Dev.	Students Within Optimal Bandwidth Mean	Students Within Optimal Bandwidth Std. Dev.	Compliers[a] Within Optimal Bandwidth Mean	Compliers[a] Within Optimal Bandwidth Std. Dev.
Grade 2						
English Language Learners (%)	9.2	28.84	9.4	29.16	9.8	29.70
Students with IEPs[b] (%)	10.8	31.00	9.8	29.68	9.5	29.38
Overage for grade[c] (%)	5.7	23.12	5.4	22.63	5.6	23.05
Number of observations	8,956		4,195		3,582	
Grade 3						
Age (years)	8.5	0.36	8.5	0.36	8.5	0.36
Low-income students (%)	36.4	48.12	39.5	48.89	39.4	48.88
Race/ethnicity (%)						
White, non-Hispanic	64.2	47.95	63.2	48.22	62.8	48.34
Black, non-Hispanic	6.7	24.97	7.3	25.99	7.2	25.85
Hispanic	18.5	38.83	19.8	39.82	20.0	40.01
Asian/Pacific Islander	6.4	24.44	5.4	22.64	5.6	23.03
Other	3.8	19.22	3.8	19.23	3.9	19.45
Male (%)	50.7	50.00	50.8	50.00	50.5	50.00
English Language Learners (%)	6.2	24.11	6.8	25.15	6.8	25.14
Students with IEPs[b] (%)	11.7	32.11	11.5	31.84	11.0	31.26
Overage for grade[c] (%)	6.6	24.78	6.8	25.22	6.9	25.37
Number of observations	7,868		6,360		5,816	

(continued)

Table 5.1 (continued)

SOURCES: Fall screening test scores from schools in the sample; student demographic data from district records.

NOTES: The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

Grade 1 data are based on the students in the sample who completed the ECLS-K Reading Assessment and are within optimal bandwidth.

The numbers of observations in the table represent the total number of students with data for at least one baseline characteristic. Individual numbers for specific baseline characteristics, in the full sample, vary as follows: 7,108 - 8,277 for Grade 1; 7,746 - 8,884 for Grade 2; 7,150 - 7,809 for Grade 3. Individual numbers for specific baseline characteristics, within the optimal bandwidth, vary as follows: 5,082 - 6,002 for Grade 1; 3,565 - 4,165 for Grade 2; 5,771 - 6,311 for Grade 3. Individual numbers for specific baseline characteristics, for complier students within the optimal bandwidth, vary as follows: 4,411 - 5,183 for Grade 1; 3,056 - 3,555 for Grade 2; 5,251 - 5,768 for Grade 3.

The sample to whom the impact findings apply is a subset of the sample used for impact estimation.

[a]Compliers were students whose actual assignment to intervention was the same as their intended assignment to intervention as determined by the decision rules.

[b]"IEP" represents Individualized Education Plan. This classification does not distinguish between reading IEPs and other IEPs.

[c]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

The Response to Intervention (RtI) Evaluation

Table 5.2

Estimated Impacts of Assignment to Tier 2 or Tier 3 Intervention Services for Students Within Optimal Bandwidth, by Grade

	Impact of Intended Assignment to Tier 2 or Tier 3 Intervention Services		Impact of Actual Assignment to Tier 2 or Tier 3 Intervention Services	
Grade	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value
Grade 1				
ECLS-K Reading Assessment	-0.13	0.000	-0.17	0.002
	(0.036)		(0.054)	
TOWRE2	-0.12	0.002	-0.11	0.057
	(0.039)		(0.058)	
Grade 2				
TOWRE2	0.04	0.298	0.10	0.084
	(0.034)		(0.061)	
Grade 3				
State reading achievement test	-0.01	0.722	-0.01	0.823
	(0.032)		(0.046)	

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement test scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

All outcomes were standardized to have a standard deviation of 1, so impact estimates are reported in effect-size units. The impact of intended assignment to Tier 2 or Tier 3 intervention services was estimated using an OLS regression of the outcome on treatment status as determined by the school decision rule. The impact of actual assignment to Tier 2 or Tier 3 intervention services was estimated using a 2SLS regression of the outcome on the indicator of student receiving intervention at least in the fall semester, using treatment status as determined by the school decision rule interacted with school indicators as the instrument variables. A complete description of the estimation model, including the use of covariates and standard error adjustments, can be found in Appendix E.

A two-tailed t-test was applied to the estimated effect.

First-stage F-statistics are 92.0 for Grade 1 ECLS-K; 79.3 for Grade 1 TOWRE2; 60.4 for Grade 2; 121.2 for Grade 3.

The numbers of students are 6,224 for Grade 1 ECLS-K; 5,448 for Grade 1 TOWRE2; 4,305 for Grade 2; 6,478 for Grade 3.

Efficiency test[128] is –0.11 standard deviation and is close to being statistically significant (p-value = 0.057).[129] These are the estimated effects for students in the vicinity of the cutoff value of the rating variable.

The magnitude of the impact on the ECLS-K Reading Assessment is not trivial. Based on estimates from Hill, Bloom, Black, and Lipsey,[130] the average annual gain in reading for Grade 1 — calculated based on national norming-sample scores from seven major standardized comprehensive reading tests — is 1.52 standard deviations. Therefore, an effect of –0.165 is equal to approximately 11 percent, or one-tenth of a year, less learning for students who were close to the cut point and were actually placed into tiers according to their intended assignment. These estimated impacts are not sensitive to the specific analytic approach used here.[131]

Even though the two outcomes differ in the range of reading skills that they assess — with the ECLS-K Reading Assessment for Grade 1 being a comprehensive assessment of a range of reading skills such as decoding, vocabulary, and passage comprehension while the TOWRE2-Sight Word Efficiency Test is narrowly focused on decoding fluency — the scores from these two tests correlate fairly highly (correlation coefficient = 0.85). As a result, it is not surprising to see similar impact estimates for these two outcomes.

- **For Grade 2, the estimated impact of assignment to Tier 2 or Tier 3 intervention services on student's fluency skill is positive but not statistically significant.**

The estimated effect is +0.10 standard deviation (p-value = 0.084). An effect of this magnitude is equivalent to about 10 percent more learning for students in the vicinity of the cut point who were assigned to intervention as intended, compared with their counterparts in the comparison group.

- **For Grade 3, the estimated impact of assignment to Tier 2 or Tier 3 intervention services on students' general reading skills, as measured by the state reading achievement test scores, is near zero and not statistically significant.**

[128] Torgesen, Wagner, and Rashotte (1999).

[129] The significance level of these findings remained the same with the Benjamini-Hochberg adjustment for multiple hypothesis testing (Benjamini and Hochberg, 1995).

[130] Hill, Bloom, Black, and Lipsey (2008).

[131] Appendix G presents the impact estimates for actual assignment under a variety of alternative model and sample specifications, including alternative bandwidths, alternative model specifications, and exclusion of certain observations.

Grade 3 students' reading performance was measured by their test scores on the high-stakes comprehensive state reading achievement test. The estimate on this outcome is small (effect size = –0.01) and not statistically significant (p-value = 0.823).

To summarize, results presented in Table 5.2 show that early-grade elementary students *at the margin* of being considered at risk by current screening measures failed to benefit from Tier 2 or Tier 3 intervention services provided to them. In first grade, these students actually fell further behind their counterparts who, because they scored just above the cut point on the screening variable for intervention, were placed to receive only Tier 1 services.

There are a few caveats to keep in mind when interpreting these results. First, the treatment condition tested here is "assignment to the reading intervention services provided to Tier 2 and Tier 3 students." This is a key feature of the RtI system but does not represent the entire RtI system in and of itself. Hence, one cannot draw conclusions about the effectiveness of the RtI system based on the findings reported in this study.

Second, the estimate of the effect of actual *assignment* to Tier 2 or 3 intervention is related, but is not equivalent, to the effect of *receiving any* intervention. Chapter 4 reports that some schools in the impact sample provided intervention to at least some students in all reading levels (the All-Level schools). This implies that, at least in some schools, some students in the comparison condition also received reading interventions (that is, treatment).[132] Consequently, the service contrast between treatment and comparison conditions was reduced, and so was the estimated difference in outcome between these two conditions. In other words, the estimated effect of *assignment* to Tier 2 or Tier 3 intervention for the full sample was expected to be smaller in magnitude in the All-Level schools than would be the case if the students who were reading at or above grade level did not receive reading interventions. Discussions in this chapter focus on the estimated effect for the full sample, while Chapter 6 explores the difference in the impact estimates between the Below-Only schools and the All-Level schools.

Third, the findings do not apply to *all* students in the sample who received Tier 2 or 3 services. As discussed in detail above, given the limitations of the RD design, these findings can be generalized only to students whose rating values were close to the cut point and whose assignment complied with their intended tier assignment based on the decision rules.

Fourth, these findings show no consistent pattern in the impacts of assignment to Tier 2 or Tier 3 intervention across grades. This might be related to different levels of prior exposure to RtI practices for students in these grades. Given that all the impact sample schools were required to have at least three years' experience in implementing the RtI framework at the begin-

[132]Note that the intervention services for comparison group students could be different from the kind of intervention services provided to Tier 2 or Tier 3 students. Chapter 4 provides some evidence of that. However, these two kinds of intervention services cannot be definitively distinguished from each other.

ning of the study year, students in Grades 1 to 3 could have been exposed to RtI for varying amounts of time, with third-graders having the most exposure and first-graders the least.[133] Due to lack of data from prior years, however, the analysis cannot disentangle the effect of prior exposure from the grade-specific effect of intervention. Therefore, these results should be considered as the average incremental effect of being assigned to intervention in the study year (2011-12) on students with varying degrees of prior exposure to RtI.

Discussion

To put the primary impact findings in context, this section first describes recent reading intervention studies in terms of the treatment being evaluated, the method, and the sample and then contrasts the current study with this existing research.

What Key Earlier Studies Say About the Effects of Reading Intervention

Over the past two decades, the field has seen an increase in studies addressing the effect of interventions delivered to early readers in need of help within an RtI framework. A targeted survey of the recent literature (since 1999) yields 27 studies that report the impact of providing certain types of interventions to students with reading difficulties on a range of reading skill measures. The remainder of this section and Table 5.3 summarize the features of these studies and their findings.

- **Targeted Population.** Grade 1 is the center of the research literature. Of the 27 studies, interventions in 18 studies focused on Grade 1 students, with another 5 including Grade 1 among the targeted grades; only 4 studies had other grade levels (Grades 2 or 3) as the research target (for a total of 9 mixed-grade studies).
- **Types of Intervention Studied.** Two types of interventions — small-group intervention and one-on-one tutoring to students in Tier 2 — emerged as the target of these studies. Small-group intervention is the focus of 12 out of 18 studies of Grade 1 only and the focus of 6 out of 9 mixed-grade studies. The rest of the studies evaluate one-on-one intervention or tutoring. Both types of interventions tended to have the following features:

[133]This is true on average. Within each grade level, individual students' prior exposure to RtI could also vary with their mobility between schools in past school years. Also note that, in most of the sample schools, students may have been exposed to RtI since Kindergarten.

Table 5.3

Summary of Key Studies Published Since 1999

Type of Intervention and Evaluation Method	Number of Studies	Number of Studies with Sample > 100	Number of Studies (Significant Findings)[a]: Comprehensive Reading Outcomes	Number of Studies (Significant Findings)[a]: Specific Reading Outcomes
Grade 1 Only	**18**			
Small-group randomized controlled trial (RCT)[b]	9	3	+: 1 (1) 0: 1 (1)	+: 3 (11) 0: 7 (28)
Small-group quasi-experimental design (QED)[c]	3	1	+: 1 (1) 0: 0 (0)	+: 2 (9) 0: 3 (6)
One-on-one tutoring RCT[d]	2	0	+: 0 (0) 0: 0 (0)	+: 0 (0) 0: 2 (6)
One-on-one tutoring QED[e]	4	1	+: 0 (0) 0: 0 (0)	+: 4 (20) 0: 3 (11)
Other or Mixed Grades	**9**			
Small-group RCT[f]	2	2	+: 1 (3) 0: 0 (0)	+: 2 (7) 0: 2 (6)
Small-group QED[g]	4	2	+: 1 (2) 0: 0 (0)	+: 2 (6) 0: 1 (2)
One-on-one tutoring RCT[h]	1	0	+: 0 (0) 0: 0 (0)	+: 1 (2) 0: 0 (0)
One-on-one tutoring QED[i]	2	0	+: 0 (0) 0: 0 (0)	+: 2 (5) 0: 2 (3)

SOURCE: Authors' summarization of existing studies.

NOTES: [a]The first number in each row in the findings columns refers to the number of studies, and the number in parentheses refers to the number of significant findings. "+" refers to findings that are positive and statistically significant at the 0.05 level; "0" refers to findings that are not statistically significant at the 0.05 level; no reviewed study has a negative and statistically significant finding at the 0.05 level. Comprehensive reading outcomes evaluate all or most of the important aspects for reading at a given grade level. In early grades, emphasis is on comprehension, vocabulary, and accuracy. Specific reading outcomes measure only one facet of reading proficiency.

[b]Burns (2011); Case et al. (2010); Denton et al. (2010); Gilbert et al (2013); Kerins, Trotter, and Schoenbrodt (2010); Mathes et al. (2005); Vaughn et al. (2006); two studies in Wanzek and Vaughn (2008).

(continued)

Table 5.3 (continued)

[c]Baker et al. (2015); Ebaugh (2000); Harn, Linan-Thompson, and Roberts (2008).
[d]Gibbs (2001); McMaster, Fuchs, Fuchs, and Compton (2005).
[e]Ehri, Dreyer, Flugman, and Gross (2007); Jenkins, Peyton, Sanders, and Vadasy (2004); Vadasy, Jenkins, and Pool (2000); Vadasy, Sanders, and Peyton (2005).
[f]Gunn, Biglan, Smolkowski, and Ary (2000); Ransford-Kaldon et al. (2010).
[g]O'Connor et al. (2014); O'Connor et al. (2013); O'Connor, Fulmer, Harty, and Bell (2005); Vaughn, et al. (2009). O'Connor et al. (2013) and O'Connor, Fulmer, Harty, and Bell (2005) did not report statistics on the findings in their studies.
[h]Vadasy, Sanders, and Tudor (2007).
[i]Two studies in Vadasy, Sanders, and Peyton (2006).

- They were designed by the researchers and were delivered under controlled conditions. None of these 27 studies examine "home-grown" or preexisting interventions.
- They focused on students placed in Tier 2, and the duration of the intervention ranged from 7 to 26 weeks.
- They tended to use specific curricula selected by the researchers and usually had one or more specific reading skills, rather than comprehensive reading skills, as targeted outcomes.
- They usually supplemented, rather than supplanted, the core reading instruction provided to all students.
- Professional development, training, and coaching of interventionists were usually provided by the study research team.

- **Research Designs.** Of these 27 studies, 14 were randomized controlled trials (RCTs). The rest employed different types of quasi-experimental designs (QEDs) that ranged from Regression Discontinuity (RD) design to historical comparison group design. It is worth noting that two studies had RD designs; one focused on Grade 1 small-group intervention,[134] and the other studied small-group intervention for Grade 2 students who did not respond to initial Tier 2 interventions.[135]
- **Sample Sizes.** The sample size for each of the reviewed studies tended to be small. Of the 27 studies, 18 had an overall sample size of 100 students or fewer,

[134]Baker et al. (2015).
[135]Vaughn et al. (2009).

and 9 studies had a sample size larger than 100 students.[136] None of the summarized studies had a student sample size in a given grade approaching this study's sample size, which has more than 8,000 students for each grade.

- **Findings.** The two rightmost columns in Table 5.3 summarize the findings reported by the existing studies. The symbols "+" and "0" represent, respectively, positive and significant findings and nonsignificant findings. The first number in each cell refers to the number of studies, and the number in the parenthesis refers to the total number of statistically significant findings in a given category.[137]
  - Only a handful of studies estimated the impact of intervention on a comprehensive reading test. Among the five studies that did, four showed positive and significant estimates while one yielded nonsignificant results.
  - Most studies tested for the impact of intervention on specific reading skills, such as fluency, decoding, and others. Findings are either positive and significant (16 of 27 studies) or nonsignificant (20 of 27 studies).[138] For the studies of Grade 1 only that focus on small-group intervention — the category with an intervention structure most like the one in the present study — the reviewed literature yields 35 nonsignificant findings and 22 statistically significant ones, considering both comprehensive and specific reading outcomes.

In summary, these recent studies support the conclusion that well-designed and closely monitored supplemental reading interventions provided in a small-group setting (either within small groups or one-on-one) could be beneficial to early-grade readers in terms of improving their specific reading skills. The evidence is stronger for second and third grades than for first grade. The effect of such intervention on students' more comprehensive reading skills is less clear. Also not clear is the impact of such interventions if they were to be implemented at a larger scale.

---

[136]The small sample size of many of these studies determines that they were only able to detect impacts of a fairly large size.

[137]There can be multiple findings from a given study, and results reported here are corrected for multiple hypotheses testing using the Benjamini-Hochberg adjustment, as recommended by the What Works Clearinghouse (Benjamini and Hochberg, 1995).

[138]Some studies have findings on some outcomes that are positive and significant as well as findings on other outcomes that are nonsignificant. These studies are counted in both categories.

How the Current Impact Findings Fit in the Literature

Compared with these recent studies, the current evaluation is unique in several ways. To start, this study assesses the mechanisms that the impact sample of experienced RtI schools chose to use in assigning students to tiers. As discussed more in Chapter 6, the methods that they used to screen and place students in tiers varied across the impact sample schools, in contrast to many of the prior studies. In addition, this study found natural practice variation in the organization and delivery of reading interventions in small groups, which previously had been understudied. Finally, this study uses an RD design that estimates the impact of assignment to intervention by comparing outcomes of students just above and just below the cut point. Findings based on this design, therefore, cannot be generalized to all students served by interventions. This is different from the RCT studies — which account for the majority of the prior evaluations — whereby similar eligible students were randomly assigned to receive interventions or not. As a result, and as discussed in detail above, the interpretation of findings from this study is quite different from the interpretation of findings from earlier RCT evaluations.

A limitation of this study is that it does not measure the impact of assignment to intervention services using a consistently defined population of children at risk of reading difficulties. Each school in the impact sample determined its own cut point on its preferred screening test. This heterogeneity in cut points and tests implies heterogeneity in the reading level of students placed in Tier 2 across schools in the impact sample.

In sum, this study differs from what has been done before in terms of how students were assigned to tiers, how the interventions were provided, and how the impact findings should be interpreted, based on the research design — all of which could have contributed to the differences in the findings between this study and others.

Chapter 6 explores these differences in search of mechanisms that may explain the current findings or shed light on future research. Specifically, it examines the relationships between the impact findings and factors related to tier assignment, the organization of intervention delivery, and the school context and student body composition within which the intervention took place.

Chapter 6

Exploring the Relationships Between Estimated Impacts and School and Student Characteristics

The primary impact findings from the Response to Intervention (RtI) evaluation that are reported in Chapter 5 suggest that actual assignment to intense reading intervention services did not improve the reading skills of early readers who were just below the cut point over the course of one school year. In fact, for first-graders who were on the margin of being identified for intervention, the estimated impact is negative and significant for one outcome measuring broad reading skills. There are negative but not statistically significant impacts for first-graders on a separate outcome measuring decoding fluency. The impact finding is positive but not statistically significant for second-graders and is essentially zero for third-graders. To the extent that there is variation in estimated impacts across schools within a grade, these average impact estimates may mask important differences in the effectiveness (or lack of effectiveness) of the intervention under different conditions.

To this end, Chapter 6 explores hypotheses for potential explanations of the primary findings. By exploiting variation in the estimated impact of assignment to Tier 2 or Tier 3 intervention services across the study schools, this chapter examines associations between the magnitude of the impact estimates on student reading achievement and certain school and student characteristics. Note, however, that this study was not designed to detect subgroup or differential subgroup effects; therefore all analyses reported in this chapter are exploratory and for hypothesis-generating purposes only. Specifically, the chapter addresses the following research questions:

1. What is the extent of variation in estimated impacts across RtI schools?

2. How is the estimated impact associated with certain school features or student characteristics? (This question is answered in separate sections of the chapter because the estimation methods used to explore school features differ from methods used for student characteristics.)

The school-level factors explored below capture RtI reading practices in the schools, including how students were assigned to intervention services and how interventions were organized and delivered. They also include the school context and student body characteristics that may capture average student needs in the school. The student-level characteristics could have served as proxies for individual students' reading needs. The findings show that:

- Across the impact sample schools, there is significant variation in the estimated impacts of reading interventions on reading outcomes. This is true for the estimated impacts on all four outcomes across Grades 1 to 3.
- At the school level, certain RtI practices and student body characteristics are associated with the varying magnitude of the estimated impacts across schools. However, these findings are not consistent across all grades.
- At the student level, students in specific learning circumstances (with an Individualized Learning Program [IEP] or who are overage for grade) near the cut point in some grades appear to have been affected by the treatment more negatively, but these findings are not consistent across grades or reading outcomes.

This chapter presents the impact variation across sample schools first. It then assesses the association between school-level factors and the estimated impact of assignment to intervention, as well as the relationship between student characteristics and the estimated impact. While the former explores whether any factors predict *school-level* estimated impacts, the latter explores the association between certain types of individual *students* and the impact experienced by these students. The chapter concludes with a discussion of the findings and their limitations. For ease of presentation and relevance to practitioners, this chapter focuses the analyses on the effect of *actual assignment* to Tier 2 or Tier 3 intervention services.

What Is the Extent of Variation in Estimated Impacts Across RtI Schools?

This section examines whether the impacts differ across schools in a grade. Of particular interest is (1) whether assignment to reading intervention had a *larger* estimated impact on reading outcomes for students in some schools than in others and (2) whether the programs had a *positive or negative* impact estimate for students in some schools, even though, on average, the estimated impact is negative or not statistically significant.

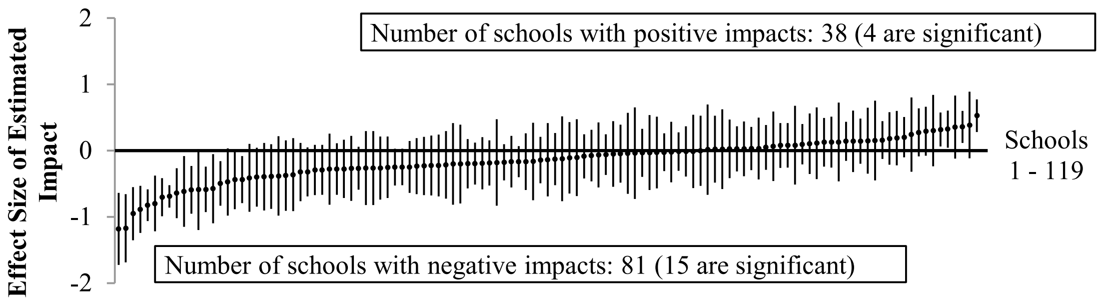
The four panels of Figure 6.1 present the estimated impact of assignment to intervention on students' reading performances, by school, for each of the four outcomes, respectively.[139] The estimates are ordered by their magnitude. The figure plots the impact estimate for each school with a solid dot and represents the respective 95 percent confidence interval with a

[139]These are the Empirical Bayesian estimates for school-level impacts. Appendix H explains why this is the preferred estimate of the program effect for a given site (Raudenbush and Bryk, 2002; Bloom, Raudenbush, Weiss, and Porter, under review).

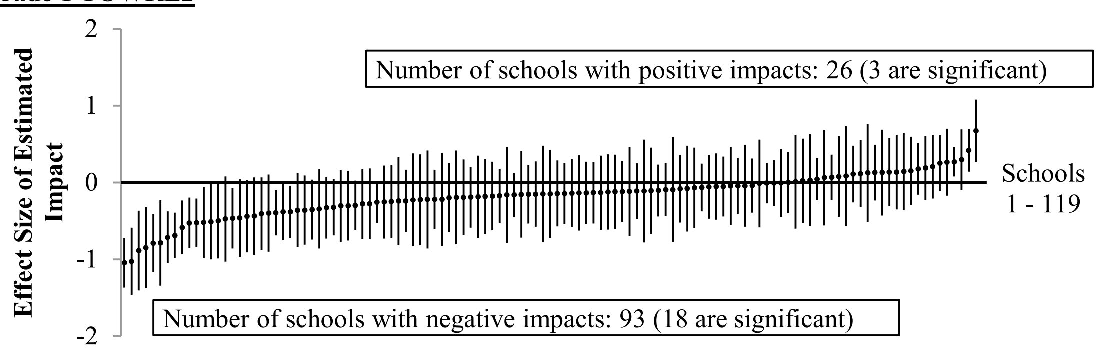
Figure 6.1

Distribution of School-Level Impact Estimates of Actual Assignment to Tier 2 or Tier 3 Intervention Services, by Outcome

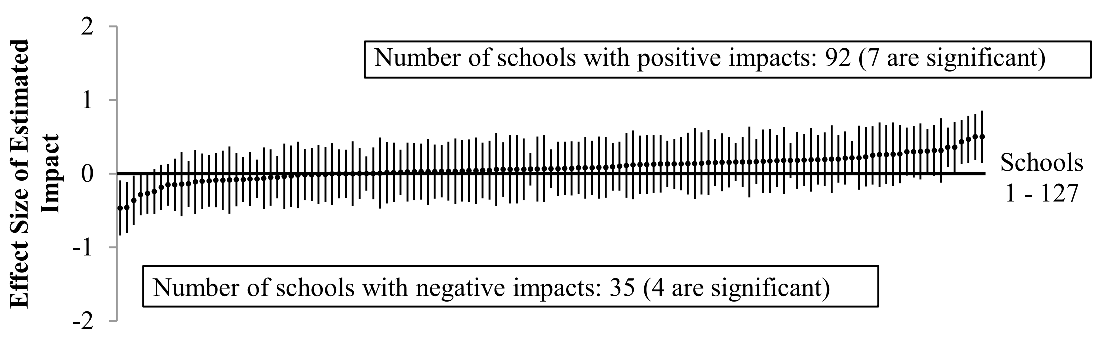
Grade 1 ECLS-K Reading Assessment



Grade 1 TOWRE2



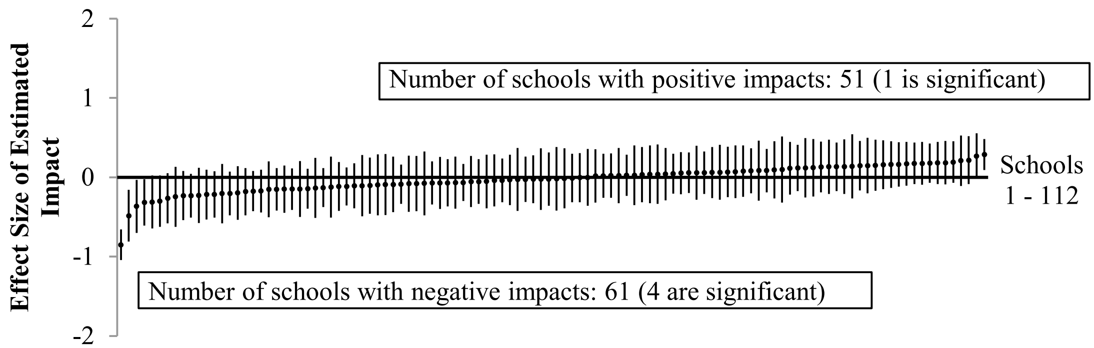
Grade 2 TOWRE2



(continued)

Figure 6.1 (continued)

<u>Grade 3 State Reading Achievement Test</u>



SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: All outcomes were standardized to have a standard deviation of 1, so impact estimates are reported in effect-size units. The fixed-effect impact for each school was estimated using a 2SLS regression of the outcome on the indicator of actual assignment to intervention interacted with school indicators, using the intended treatment status interacted with school indicators as the instrument variables. A complete description of the estimation model can be found in Appendix H.

A chi-squared test was used to test the statistical significance of the variation in the empirical Bayes impact estimates for each outcome. The Q-statistics for this test are 498 for Grade 1 ECLS-K Reading Assessment, 373 for Grade 1 TOWRE2, 239 for Grade 2, and 285 for Grade 3. The corresponding p-values are below 0.001 for all four outcomes.

vertical line running through each impact estimate. The wider the confidence interval, the broader the margin of error, and the greater the uncertainty about the impact estimate. The impact estimates with confidence intervals that do not include zero are statistically significant (p-values are less than or equal to 0.05). Several patterns that emerge from this figure are discussed below.

- **There is observable variation in the estimated impacts across schools, and this is true regardless of whether the average impact estimate is statistically significant or not.**

For example, the estimated school-level impacts on the ECLS-K Reading Assessment score for Grade 1 (first panel) range from –1.18 to +0.53 standard deviations in effect size, and the school-level impact estimates on the third-grade achievement test score (last panel) range from –0.82 to +0.29. Out of a total of 119 schools included in the impact analysis for Grade 1, there are 15 schools with significant negative findings and 4 schools with positive and significant findings. Similar patterns of variation were found for the estimated impacts on the other three reading outcomes. Statistical tests show significant variation in impact estimates across schools — for all four outcomes across three grade levels. This finding indicates that the estimated impact could be more negative or more positive in some schools than in others, regardless of the overall average impact estimate.

- **The distributions of school-level impact estimates are consistent with the primary impact findings for these outcomes.**

For the two Grade 1 outcomes with negative overall average estimated impacts of actual assignment to intervention, a majority of the sample schools — 81 and 93 out of a total of 119 schools, for the ECLS-K Reading Assessment and TOWRE, respectively — registered negative impact estimates. For second grade, where the overall impact is positive but not statistically significant, 92 out of 127 schools in the sample had positive impact estimates. Lastly, the schools split roughly equally between positive and negative estimates for third grade (51 positive and 61 negative), for which the overall impact estimate is closer to zero.

Finally, to assess the variability in impact estimates across schools more systematically, a chi-squared test was used to assess whether the variation in school-level impact estimates is larger than would be expected due to chance.[140] Results from this test show that the school-to-school impact variation is statistically significant within each grade; this is true for all four outcomes (p-values are less than 0.001). The next part of this chapter exploits this variation to see whether and how certain school or student characteristics are associated with estimated school-level impacts.

The correlations among school-level impact estimates across grades are, with one exception, low. Table 6.1 presents the correlation coefficient across the four outcomes in three grades. Across grades, correlations between the impact estimates on different outcomes are 0.15 and below, with a zero (–0.00) correlation between first grade ECLS-K Reading Assessment scores and second-grade TOWRE scores. Within Grade 1 (the only grade for which there were two separate reading outcome measures), there is a stronger correlation (correlation coefficient = 0.80) between the two outcomes. This is not surprising, because the student sample underlying these estimates is essentially the same, and the two outcome measures are highly correlated with each other.

This lack of correlation across grades in the estimated school-level impacts could be related to the fact that the impact estimates are based on different cohorts of students. First-

[140]Appendix H describes how this test is carried out.

The Response to Intervention (RtI) Evaluation

Table 6.1

Correlation Between Estimated School-Level Impacts Across Outcomes

	Grade 1		Grade 2	Grade 3
	ECLS-K Reading Assessment	TOWRE2	TOWRE2	State Reading Achievement Test
Grade 1				
ECLS-K Reading Assessment	1.00			
TOWRE2	0.80	1.00		
Grade 2				
TOWRE2	-0.00	0.06	1.00	
Grade 3				
State reading achievement test	0.14	0.14	0.15	1.00

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: All outcomes were standardized to have a standard deviation of 1, so impact estimates are reported in effect-size units. The fixed-effect impact for each school was estimated using a 2SLS regression of the outcome on the indicator of actual assignment to intervention interacted with school indicators, using the intended treatment status interacted with school indicators as the instrument variables. A complete description of the estimation method can be found in Appendix H.

graders could have been different from third-graders in terms of demographic composition, academic development, prior exposure to RtI (as discussed earlier), and other factors, even within the same school. These differences may be associated with the extent to which students were affected by the reading intervention services. It also could be that the intervention services differed by grade and, as such, may have differed in effectiveness. The next section in this chapter discusses these issues further and tries to identify factors that might be related to the impact estimates.

How Is the Estimated School-Level Impact Associated with Certain School Features?

This section examines the association between the estimated school-level impacts of actual assignment to Tier 2 or Tier 3 intervention (as shown in Figure 6.1) and the characteristics of the RtI schools in the study. The purpose of this analysis is to better understand the circumstances under which reading intervention provided by these schools was more successful or less successful in helping students who were reading below grade level, which might inform decision-making regarding RtI practices.

How to Understand the Reported Results

This analysis explores the relationship between certain school features and the estimated school-level impacts on assignment to intervention services with a multivariate regression approach. By including multiple features in the model simultaneously, it allows for an assessment of the relationships between each factor and the estimated impacts while controlling for the values of other factors. Joint statistical significance of these variables was also tested to see whether these features are associated with the impact estimates as a group.[141]

In general, a positive and statistically significant estimate for a given feature from the model implies that an RtI school with that feature (for dichotomous measures), or with a higher value of that feature (for continuous measures), tended to have less negative or more positive impacts than schools without, or with a lower value of, that factor. In contrast, a negative and statistically significant estimate indicates that an RtI school with (or with a higher value of) that feature tended to have more negative or less positive impacts than schools without (or with a lower value of) it. A nonsignificant estimate indicates that the given feature is not likely to be associated with the impact estimates. The discussion below focuses on findings that are statistically significant at the 5 percent (0.05) level but also mentions findings that are significant at the 10 percent (0.10) level if they are consistent across different models (reported here or in Appendix H) — for they might also provide useful information for the interpretation of the primary impact findings.

Results from the analyses of associations between school-level characteristics and the effects of actual assignment to interventions should be considered exploratory and should be interpreted with caution, for several reasons. First, these analyses are nonexperimental because school factors were not randomly assigned to schools and are likely to be correlated with factors that are not captured in this study. Therefore, these results cannot be interpreted as causal. Second, some statistically significant findings may have occurred by chance, given the number of hypotheses tested here. Hence, the results from these analyses may be suggestive of factors that could contribute to the success or failure of the treatment on the margin and are worthy of further investigation, but they do not conclusively answer questions about the kinds of schools for which the intervention can produce more positive or less negative effects. Third, as with the

[141]The relationship between individual features and the school-level impact estimates is also assessed through a bivariate regression model. This allows for an assessment of this relationship for one feature at a time, independent of other factors. For dichotomous factors, this approach is equivalent to a subgroup analysis wherein subgroups are defined by this factor. Results from that analysis are consistent with those from the multivariate regressions and are reported in Appendix H.

confirmatory analyses presented in Chapter 5, the estimated effects apply only to students who were close to the cut point of being identified for intervention and whose assignment complies with the school's decision rules for tier assignment. And lastly, note that a statistically significant association between school characteristics and impacts does not imply a statistically significant overall impact.

School Features Examined in the Exploratory Analysis

Substantively, a wide range of school-level factors could have been associated with the RtI program's effectiveness. This analysis focuses on the school-level features that are related to the intervention delivery and the school context within which the interventions were implemented. The characteristics of the student population composition are also included in the analysis, as control variables to gauge the level and type of students' overall reading needs. This section describes these features, each of which is measured at the school-by-grade level to allow separate exploratory analysis for each grade. Table 6.2 provides descriptive statistics for these features, by grade. Note that the following discussion about these school features presents potential hypotheses on how a given feature might be related to the estimated school-level impact. However, this discussion is by no means exhaustive — there could be any number of additional hypotheses relating the school features and the impact findings. The categories of the school features are summarized below.

1. **RtI Reading Practices:** These factors capture how the reading intervention services were organized in the RtI schools. They were chosen either because they represent the differences between the current study and previous literature, as discussed at the end of Chapter 5 (factors related to tier assignment in the impact school sample), or because they could have led to different levels of service contrast across schools, as discussed in Chapter 4 (factors related to the organization of intervention services).[142] It is hypothesized that the impacts of assignment to receive Tier 2 or Tier 3 reading intervention that were found in the current study could have been related to the way that the services were delivered. The following specific features were measured and are included in this category.

   - ***Whether a school used a single or multiple screening test scores to assign students to tiers.*** This factor is related to how students were identified for intervention. Across grades, about 75 percent to 79 percent of the sample schools primarily used one screening test score to determine a

[142]Potentially, many other implementation factors could be associated with school-level impact estimates, but they are not included in this analysis due to lack of available data.

The Response to Intervention (RtI) Evaluation

Table 6.2

Descriptive Statistics of School Features Examined in the Exploratory Analysis, by Grade

Grade/Characteristic	Mean (%)	Standard Deviation
Grade 1		
RtI Reading Practices		
Grade uses a single screening test to assign students to intervention	79.0	40.91
Grade provides intervention to some students at all reading levels	47.7	50.18
Percentage of intervention groups meeting outside the core	59.0	35.38
Percentage of students identified for intervention	37.6	17.32
School Context		
School's prior Grade 3 reading performance relative to state mean[a]	3.5	14.35
Title I eligible school	69.7	46.13
School uses behavioral RtI in Grade 1	30.7	46.33
Student Body Composition		
English Language Learners	12.5	16.15
Students with an Individualized Education Program[b]	9.4	10.13
Male students	50.7	6.84
Low-income students	42.7	27.14
Students overage for grade[c]	5.4	5.49
Grade 2		
RtI Reading Practices		
Grade uses a single screening test to assign students to intervention	74.8	43.59
Grade provides intervention to some students at all reading levels	33.6	47.45
Percentage of intervention groups meeting outside the core	61.0	35.52
Percentage of students identified for intervention	35.6	15.66
School Context		
School's prior Grade 3 reading performance relative to state mean[a]	3.5	13.40
Title I eligible school	67.7	46.94
School uses behavioral RtI in Grade 1	27.5	44.84
Student Body Composition		
English Language Learners	9.8	13.56
Students with an Individualized Education Program[b]	10.3	11.03
Male students	50.9	5.86
Low-income students	39.4	24.26
Students overage for grade[c]	5.6	4.50

(continued)

Table 6.2 (continued)

Grade/Characteristic	Mean (%)	Standard Deviation
Grade 3		
RtI Reading Practices		
Grade uses a single screening test to assign students to intervention	79.5	40.58
Grade provides intervention to some students at all reading levels	30.2	46.16
Percentage of intervention groups meeting outside the core	60.1	36.27
Percentage of students identified for intervention	31.4	13.86
School Context		
School's prior Grade 3 reading performance relative to state mean[a]	3.8	12.98
Title I eligible school	65.2	47.85
School uses behavioral RtI in Grade 1	28.8	45.52
Student Body Composition		
English Language Learners	7.0	10.14
Students with an Individualized Education Program[b]	11.2	10.60
Male students	50.9	6.44
Low-income students	37.8	23.89
Students overage for grade[c]	6.5	5.75

SOURCES: Fall screening test information from schools in the sample; interventionist and teacher survey responses about reading groups; state achievement data downloaded from 13 state websites for which links are provided in Appendix D.

NOTES: The number of schools with these characteristics can be found in Appendix Table H.1.
[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and is measured as the deviation from the state mean.
[b]This classification does not distinguish between reading Individualized Education Programs (IEPs) and other IEPs.
[c]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

student's tier placement at the beginning of the fall semester.[143] If having used multiple screening tests improved the reliability of student identification and led to less false identification of students and a better match between a student's reading needs and intervention, then schools having used multiple tests could have been associated with a less negative or more positive impact.

[143]Composite screening measures are considered a single assessment in this analysis, if only the score from the composite is used to place students into a tier.

- ***Proportion of students assigned to Tier 2 and Tier 3 based on their screening test scores and the decision rules.*** This percentage decreased by grade level, with the highest percentage (38 percent) in Grade 1 and the lowest percentage (31 percent) in Grade 3. On the one hand, this measure could have been an indicator of the amount of resources that were available to help struggling readers; if more students were identified, there would have been fewer resources per student in need. Therefore, a higher proportion of identified Tier 2 or Tier 3 students could have led to more negative or less positive impacts (the "constrained resource" hypothesis). On the other hand, the measure could have indicated the kinds of students that a school was targeting for service.[144] A higher proportion likely included students scoring higher on screening tests. This could have meant that the performance level of the intervention group was better than if fewer students — only those with serious reading problems — were identified for intervention. In other words, the peer group for students in intervention had a higher performance level. To the extent that higher-performing peers led to better performance for all students in Tier 2 or 3 intervention, one may expect to see positive links between this proportion and the impacts (the "peer effect" hypothesis).[145]

- ***Whether a school provided intervention to at least some students in all reading levels in the corresponding grade.*** This feature determined whether a school was a Below-Only school or an All-Level school, as discussed in Chapter 4. In Grade 1, about 48 percent of the schools did provide intervention to some students in Tier 1 only (students in the comparison group). This percentage dropped to 34 percent and to 30 percent for second grade and third grade, respectively. Conceptually, providing intervention services to some Tier 1-only students, as in the All-Level schools, may have reduced the service contrast between the treatment and comparison groups. Findings in Chapter 4 show that the service contrast between intervention groups serving students with different reading levels was, indeed, smaller for the All-Level schools than for the Below-Only schools, along several dimensions of the intensity

[144]For example, 30 schools used DIBELS Nonsense Word Fluency – Correct Letter Sounds (NWF-CLS) as the screening test in first grade, and the cut point for each school ranges from 17 to 40 in raw scores, which corresponds to 39 percent to 63 percent in percentile rank, based on a nationally normed sample.

[145]The direction of this hypothesized "peer effect" is not entirely clear, as other things could be at play. For example, while higher proportions of Tier 2 or 3 students could mean higher-performing peers for the treatment students in intervention, it could also mean higher-performing peers for the comparison students who are in Tier 1. As a result, the positive peer effects for these two groups of students might cancel out each other.

measure. These differences in service contrast may have led to differences in the estimated impacts for these two groups of schools.[146]

- ***Proportion of intervention groups that were served outside the core reading block.*** In the impact sample schools, this percentage stayed at around 60 percent for all three grades. As discussed in Chapter 4, intervention services could have occurred during or outside the core reading block. When intervention services occurred during the core, there was an increasing likelihood that they replaced rather than supplemented instructional services, given the fact that most schools designate a fixed amount of time to the core reading block. Therefore, schools with higher proportions of intervention groups that were served outside the core were likely to have a larger service contrast than schools with a lower proportion, which could have led to less negative or more positive impacts.

2. **School Context:** These factors reflect the school context within which the reading interventions were delivered. They were selected because they might represent factors that could affect the implementation of RtI, such as baseline school reading performance, school resources, and other factors.

    - ***Overall school performance in reading at baseline.*** This was measured by the proportion of students reading at or above proficiency level, based on third-grade state reading achievement test scores from school year 2010-11 (the year immediately before the study year). To compare schools across states where different tests and standards were used, the variable was expressed as the difference between the school's proportion proficient and the average proportion proficient in the respective state. The impact sample schools, on average, outperformed the average schools in their respective states by about 3.5 percentage points in terms of the proportion of students at or above proficient reading level in the year before the study year. While this study did not include direct measures of the quality of reading instruction or intervention, this baseline measure of reading performance could have reflected the level of core reading instruction, as well as the reading backgrounds of students and the quality of reading interventions in the school, all of which could

[146]There is no obvious difference in the proportion of Tier 2 and Tier 3 students between All-Level schools and Below-Only schools. The percentage of Tier 2 and Tier 3 students ranges from 33 percent to 41 percent in All-Level schools and from 31 percent to 37 percent in Below-Only schools. In addition, as reported in Appendix Tables H.2-H.4, the correlation between a school's status as All-Level or Below-Only and the proportion of students in the school placed in Tier 2 or Tier 3 is generally low, ranging from 0.05 in Grade 3 to 0.14 in Grade 1.

be positively or negatively associated with school-level impacts for a variety of reasons. The current study therefore examines this relationship.

- ***Title I eligibility status.*** This is an indicator of a school's eligibility for Title I funding, as reported in the Common Core of Data (CCD). Around 65 percent to 70 percent of the schools in each grade's sample were schools eligible for Title I status. This status could have indicated the amount of school resources available to serve students in need, with more resources available for Title I schools than for others, allowing them to better address students' needs. Hence, this factor could lead to less negative or more positive impacts. It could have also reflected the proportion of students who came from economically disadvantaged backgrounds and who may have needed reading support at school. In this sense, a school's Title I status may have indicated higher demand for services at these schools than at comparison schools and could have led to more negative or less positive impacts.

- ***Whether a school implemented RtI practices for behavior-related interventions in Grade 1.*** About 30 percent of the impact sample schools reported using such practices to influence student behavior in first grade. Behavioral disruptions could have limited the benefits of reading intervention. This was especially true for early-grade students whose learning difficulties may have been associated with behavioral problems.[147] By addressing behavior through a separate intervention channel, schools could have alleviated the difficulty of delivering the reading intervention. Though data were collected only for Grade 1, this measure could have served as a proxy for a school's focus or philosophy about dealing with behavior in general; therefore, it is included in the analysis for Grades 2 and 3 as well.[148]

3. **Student Body Composition:** These measures of average demographic characteristics of the grade-specific student population within a school are based on the impact analysis sample and include the following factors. They reflect the peer environment within which students experienced the interventions.

   - ***Proportion of students who were English Language Learners (ELL).*** On average, the impact sample schools had about 13 percent ELL stu-

[147]Wehby et al. (2003).

[148]Many RtI programs for behavior have schoolwide components that could affect all grades. See material online at www.pbis.org.

dents in Grade 1, and this percentage dropped to about 10 percent and 7 percent for Grades 2 and 3, respectively.

- ***Proportion of students with an Individualized Education Program (IEP).*** This characteristic measures the percentage of students who were identified as needing individualized support because of a Specific Learning Disability (SLD). Note that this measure does not distinguish between an SLD and other learning, behavioral, or physical disabilities. For all three grades, about 9 percent to 11 percent of students in the sample schools had an IEP.

- ***Proportion of male students.*** The average of this characteristic is around 51 percent for all three grades.

- ***Proportion of students with low-income status.***[149] On average, this ratio ranges from 38 percent to 43 percent for Grades 1 to 3.

- ***Proportion of students who were overage for grade.***[150] In the sample schools, the percentage of overage students is between 5 percent and 7 percent in all three grades.

School-Level Findings

Table 6.3 summarizes findings on associations between school characteristics and estimated effects of reading interventions on each of the four outcomes.[151] The results labeled "Model 1" include all seven school features that were related to RtI practice and school context, while results labeled "Model 2" add the composition of the student body to see how findings vary with these additional control variables. In these tables, positive and negative estimates are represented by "+" and "–," respectively, and asterisks indicate the statistical significance level of the estimate. Note that discussion of the findings focuses on those that are statistically significant and that results discussed in the text may round up from the tables in some cases. Table 6.3 shows the findings discussed in the remainder of this section.

---

[149]This is defined based on how districts report students' low-income status. Some school districts identified students as eligible for free and reduced-price lunch; others provided an indicator of income or socioeconomic status.

[150]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

[151]These summary tables only present the signs and significance levels of the estimated regression coefficients for the school-level variables. The full set of results from these regressions — including the estimates and corresponding standard errors as well as results from bivariate regression models — can be found in Appendix H.

The Response to Intervention (RtI) Evaluation

Table 6.3

Signs and Significance Levels of Regression Coefficients Associating School Features with School-Level Impact Estimates, by Outcome

	Grade 1				Grade 2		Grade 3	
	ECLS-K		TOWRE2		TOWRE2		State Reading Achievement Test	
School Feature	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Reading RtI Practices (%)								
Grade used a single benchmark assessment to assign students to intervention	+	+	+	+	+	+	+	-
Grade provided intervention to some students at all reading levels	+	+	+	+	+ *	+	+	+
Percentage of intervention groups meeting outside the core	+	+	+	+	+	+	-	-
Percentage of students identified for intervention	+ **	+	+	-	+ **	+ ***	+	+
School Characteristics (%)								
School's prior Grade 3 reading performance relative to state mean[a]	+	-	+	+	+	+	+ *	+ **
Title I eligible schools	- **	-	-	-	-	-	+	+
School used Behavioral RtI in Grade 1	+	+	+	+	+	+	-	-
Student Body Composition (%)								
Percentage of English Language Learner (ELL) students		+		+ **		-		+ **
Percentage of Indivualized Education Program (IEP) students[b]		-		-		-		-
Percentage of male students		-		-		- **		-

(continued)

Table 6.3 (continued)

	Grade 1				Grade 2		Grade 3	
	ECLS-K		TOWRE2		TOWRE2		State Reading Achievement Test	
School Feature	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Percentage of students that had low-income status		-		-		-		-
Percentage of students overage for grade[c]		-		-		-		-
P-Value for Joint Significance Test	0.086	0.701	0.420	0.240	0.097	0.064	0.199	0.075

SOURCES: Study-administered ECLS-K reading assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records; school characteristics information from the 2010-11 Common Core of Data (CCD); interventionist and teacher survey responses; state achievement data downloaded from state websites.

NOTES: The "+" and "-" in the table represent the positive and negative sign of regression coefficient estimating conditional differential treatment effects for a given school feature. A two-step procedure was used for the estimation. First, the fixed-effect impact for each school was estimated using a 2SLS regression of the outcome on the indicator of actual assignment to intervention interacted with school indicators, using the intended treatment status interacted with school indicators as the instrument variables. The estimated impact for each school was then used in a V-known random-effects meta-analysis model whereby school characteristics, as well as their corresponding missing indicators, were used as explanatory variables in the model. Model 1 includes school's reading RtI practices and certain school characteristics as covariates, and Model 2 adds student body compositions as covariates. A complete description of the estimation model can be found in Appendix H.

A two-tailed t-test was applied to the estimated effect. The statistical significance is indicated as follows: *** indicates p-value ≤ 0.01, ** indicates 0.01 < p-value ≤ 0.05, and * indicates 0.05 < p-value ≤ 0.10.

The numbers of schools are 119 for Grade 1; 127 for Grade 2; 112 for Grade 3.

[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and is measured as the deviation from the state mean.

[b]This classification does not distinguish between reading IEPs and other IEPs.

[c]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

- **School-level features explored in these analyses are not associated in a statistically significant way with the impact of assignment to Tier 2 or Tier 3 intervention services on first-graders' comprehensive reading measure.**

None of the examined features show consistent, statistically significant results across these two models for the estimated impacts on the ECLS-K Reading Assessment scores. Even though the estimates for percentage of students identified for Tier 2 and Tier 3 and the school's Title I eligibility status are significant in the first model, their estimated association dissipates once the student body composition is controlled for in the model, indicating that their association with the estimated impacts could have been channeling other factors that were captured in the second model. In addition, these variables (with or without student composition) do not jointly explain the variation in the estimated school-level impacts, as suggested by the joint significance tests reported at the bottom of Table 6.3.

- **The proportion of ELL students in schools may be associated with the estimated impact of assignment to intervention on students' decoding-fluency measure in Grade 1. None of the other school-level features is associated with the impact for this outcome.**

The proportion of English Language Learners (ELL) among Grade 1 students in the school is positively associated with the impact on the TOWRE2 Sight Word Efficiency scores, and this result seems to be robust.[152] This indicates that schools with higher proportions of ELL students had a less negative impact of assignment to intervention services than schools with fewer ELL students. The fact that this association is significant for impacts on the TOWRE2 scores but not on the ECLS-K Reading Assessment scores for essentially the same group of students could suggest that intervention in schools with a higher proportion of ELL students might focus more on decoding fluency.

- **There is evidence that the proportion of students who were identified for intervention in a school is *positively* associated with the estimated impacts on second-graders' decoding-fluency measure, while the percentage of male students in Grade 2 is *negatively* associated with the estimated impacts in this grade.**

The positive and significant estimate for proportion of identified Tier 2 and Tier 3 students is consistent for Grade 2 across all relevant models.[153] These positive findings seem to support the "peer effect" hypothesis rather than the "constrained resource" one (discussed

[152]As shown in Appendix H, this finding is always significant at the 0.05 level, whether it is estimated by itself, together with other measures of student body composition, or with all other school-level features.

[153]This is true for models reported in Appendix H as well.

above). In addition, there is indication of a negative link between the estimated impact and the percentage of male students.

- **A school's baseline reading performance and the percentage of ELL students in Grade 3 may be positively associated with the impact estimates on the comprehensive reading measure for third-graders. None of the other features is associated with the impact for this outcome.**

The results suggest that there is a positive link between the impacts on third-grade state achievement test scores and the school's baseline third-grade reading performance measure, indicating that schools with better reading performance are the ones with higher impacts. Second, there is an indication that the proportion of ELL students in the school is positively correlated with the impact estimates.

In summary, the correlational analysis linking school-level features and the impacts of receiving intervention yields only a few sporadic findings, suggesting a few possible explanations of variation in the impact estimates across schools. The findings include a positive association between the proportion of students identified for intervention and the estimated impact in Grade 2, a positive association between the school's baseline reading performance and the estimated impact in Grade 3, and a positive association between the proportion of ELL students and the estimated impacts on decoding fluency in Grade 1 and on comprehensive reading in Grade 3. Altogether, however, they do not provide a consistent explanation for the impact findings across grade levels.

How Is the Estimated Impact of Assignment to Tier 2 or Tier 3 Intervention Services Associated with Certain Individual Student Characteristics?

This section examines the association between the estimated impacts of placement in Tier 2 or Tier 3 intervention services and certain student characteristics — this time, at the student level. The purpose of this analysis is to assess whether being assigned to receive intervention affected different types of students differently, which could provide useful information about how practitioners might target the intervention services.

How to Understand the Reported Results

Conceptually, to assess the differential impact of actual assignment to Tier 2 or Tier 3 intervention that was experienced by students who had different characteristics, separate impacts were first estimated for subsets of students with or without a certain characteristic. As in the main impact analysis, these impact estimates apply only to students whose ratings were just above or just below the cut point. These impact estimates were then compared with each other to see whether students with the given characteristic benefited more or less from the intervention than students without that characteristic. A regression model similar to the one used for the primary impact analysis is adapted here to serve this goal. Specifically, the reported estimate indicates the relationship between a given individual student characteristic and the impact of actual assignment to intervention. A positive estimate suggests that students with a given characteristic experience a more positive or a less negative impact than average students without this characteristic, and vice versa. Note, however, that a statistically significant estimate of this association does not imply a significant overall impact.

For simplicity, the discussion below focuses on the patterns of findings from the joint regression approach, whereby all five student characteristics (discussed below) were included in the model simultaneously. Results from the bivariate approach are largely consistent with the ones presented below and, therefore, are not presented here. They are available in Appendix H. As discussed above, these results need to be interpreted with caution because this analysis is exploratory and cannot be used for causal inference.

Individual Student Characteristics Examined in the Exploratory Analysis

A wide range of individual student characteristics could have affected students' potential of benefiting from the reading intervention. This analysis focuses on the ones that can be considered as indicators of student needs and examines whether the students with these characteristics stood a better chance or a worse chance of benefiting from the treatment. Specifically, students' status in the following five dimensions are considered:[154]

- **Sex.** Given the different developmental trajectories for boys and girls, especially in early grades, it is of interest to see whether the intervention affected boys and girls in the sample differentially.

- **Low-Income Status.** A student's socioeconomic status is highly correlated with his or her academic performance and could have served as a proxy for the student's reading needs. Between 36 percent and 42 percent of students in the sample had low-income status (were eligible for free or reduced-price lunch).

- **English Language Learner (ELL).** ELL students may exhibit a different kind of need for reading intervention than other students and, therefore, may not have benefited as much from the intervention if the intended target of the intervention was native English speakers. About 13 percent of first-graders in

[154]Descriptive statistics of these student characteristics appear in Chapter 5, Table 5.1.

the sample were ELL students. Among second- and third-graders, this percentage dropped to 9 percent and 6 percent, respectively.

- **Individualized Education Program (IEP).** Compared with other students, IEP students may be in greater need of help, but their needs may differ from the needs of other students. As mentioned above, students with a reading IEP likely already received intervention services and did not respond, or they had severe enough needs that they began school with an IEP. If the reading intervention did not adjust to these students' needs and provide differential treatment, they may not have benefited from the intervention. The percentage of IEP students stayed at around 10 percent to 12 percent for all three grades (though not all of these students had a reading-related IEP).

- **Overage for Grade.** A student could have been held back because of inadequate progress (in reading or other subjects) or because of a parental decision to postpone the child's initial enrollment. If the former, then this status is also a proxy for student need or prior achievement level. The overall percentage of students overage for grade was 5 percent to 7 percent, depending on grade level.

Student-Level Findings

Table 6.4 summarizes the findings for the relationship between student characteristics and the impact of assignment to Tier 2 and Tier 3 interventions, by outcome. Like the previous table, this one uses "+," "–," and asterisks to represent the direction and statistical significance levels of the findings.[155]

The discussion below focuses on findings that are statistically significant at the 5 percent level, but it also mentions results that have a slightly weaker significance level but might be of interest otherwise.

- **Grade 1 students with an IEP and students overage for grade are associated with a more negative effect on their comprehensive reading measure. The joint association between all five student characteristics and the estimated impact is significant.**

In particular, compared with a similar student with no IEP, an IEP student whose screening score was just below the cut point and who participated in the intervention would fall further behind a counterpart who was just above the cut point and had no exposure to the

[155]The detailed estimates for each outcome are presented in Appendix H.

Table 6.4

Signs and Significance Levels of Regression Coefficients Associating Student Characteristics with Impact Estimates, by Outcome

	Grade 1		Grade 2	Grade 3
Student Characteristic	ECLS-K	TOWRE2	TOWRE2	State Reading Achievement Test
Student is male	-	-	-	-
Student had low-income status	-	-	-	-
Student had English Language Learner (ELL) status	+	+	+	+ **
Student had an Individualized Education Program (IEP)[a]	- ***	-	-	- **
Student was overage for grade[b]	- **	-	-	-
P-Value for Joint Significance Test	0.006	0.434	0.301	0.018

SOURCES: Study-administered ECLS-K reading assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: The "+" and "-" in the table represent the positive and negative sign of regression coefficient estimating conditional differential treatment effects for a given student characteristic. The model used a 2SLS regression of outcome on the interactions between student's actual assignment to Tier 2 or Tier 3 intervention and school indicators as well as the four listed student characteristics, using as instrumental variables the interactions between student's intended assignment to Tier 2 or Tier 3 intervention and school indicators as well as the four listed student characteristics. A full description of the model can be found in Appendix H.

A two-tailed t-test was applied to the estimated effect. The statistical significance is indicated as follows: *** indicates p-value ≤ 0.01, ** indicates 0.01 < p-value ≤ 0.05, and * indicates 0.05 < p-value ≤ 0.10.

Student sample sizes are as follows: 6,236 for Grade 1 ECLS-K reading assessment; 5,398 for Grade 1 TOWRE2; 4,301 for Grade 2; and 6,549 for Grade 3.

[a]This classification does not distinguish between reading IEPs and other IEPs.

[b]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 are classified as overage.

intervention in the fall. Similarly, an overage student who was actually placed in intervention would also fall further behind counterparts in the comparison group, relative to a student with similar treatment who was not overage for grade.

These results suggest that the Tier 2 or Tier 3 intervention services as provided in the sample schools may not have been adequate or appropriate for students with specific learning needs. As discussed above, IEP status is an indication for students needing individualized instruction to deal with their learning challenges; and the overage for grade status could mean that a student was held back in school because of inadequate progress. The negative and significant estimates suggest that students' needs may not have been met through the reading intervention services that they received.

- **In both Grade 1 and Grade 2, none of the student characteristics is significantly associated with the impact on students' decoding-fluency measure, and the joint association between the impact estimates and these characteristics is not statistically significant for either outcome.**

- **Jointly, the five student characteristics are significantly associated with the impact estimates for the comprehensive reading measure for Grade 3 students. Specifically, ELL students are associated with more positive impact estimates, while IEP students are associated with more negative impact estimates.**

In Grade 3, the joint significance of these characteristics in explaining the impact of assignment to intervention services is below 0.05 (p-value = 0.018), and two out of the five characteristics examined are significant at the 0.05 level. Specifically, the association between ELL status and the impact estimate is positive, and the correlation between IEP status and the impact estimates is negative. This latter finding is similar to the Grade 1 finding and could imply that the needs of IEP students were not met in third grade either.

Overall, though the findings are not consistent across grade levels, they suggest that the reading interventions, as delivered in the impact sample schools, may not have been appropriate for students in specific circumstances who were near the cut point. This is supported by the significantly negative findings for IEP and overage students on the comprehensive reading measure in Grade 1 and by the negative and close-to-significant finding for overage students and IEP students in Grade 1 (for the decoding-fluency measure) and Grade 3 (for the comprehensive reading measure), respectively.

Implications of the Exploratory Analysis

The exploratory analyses presented in this chapter establish that the estimated impact of actual assignment to Tier 2 or Tier 3 intervention services for students on the margin of being identified varies significantly across schools, and this is true for all four outcomes across three grade levels. The school- and student-level correlational analyses yield suggestive findings for potential mechanisms through which the magnitude of impact was influenced; some of these mechanisms were more pronounced or more consistent than others. However, the overall pattern of the primary findings — especially the negative and significant or close-to-significant impact estimates on both outcomes for first-graders — remains puzzling. There could be unexplored but plausible factors and hypotheses, from initial identification of students to the school-level curriculum structure, that could have led to negative impacts of assignment to Tier 2 or Tier 3 intervention services in Grade 1. The study research team was not able to explore these hypotheses within the scope of this study due to limited availability of data, and therefore does not draw any conclusions. These factors need to be explored in future research.

This page has been left blank for double-sided copying.

References

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91, 434: 444-455.

Baker, Scott K., Keith Smolkowski, Erin A. Chaparro, Jean Louise M. Smith, and Hank Fien. 2015. "Using Regression Discontinuity to Test the Impact of a Tier 2 Reading Intervention in First Grade." *Journal of Research on Educational Effectiveness* 8, 2: 218-244.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. 2011. "Heaping-Induced Bias in Regression-Discontinuity Designs." NBER Working Paper No. 17408. Cambridge, MA: National Bureau of Economic Research.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of Royal Statistical Society.* Series B (Methodological) 57, 1: 289-300.

Blachman, Benita A., Eileen W. Ball, Rochella S. Black, and Darlene M. Tangel. 1994. "Kindergarten Teachers Develop Phoneme Awareness in Low-Income, Inner-City Classrooms." *Reading and Writing: An Interdisciplinary Journal* 6, 1: 1-18.

Black, Paul, and Dylan Wiliam. 2009. "Developing the Theory of Formative Assessment." *Educational Assessment, Evaluation and Accountability* 21, 1: 5-31.

Blackorby, Jose, Ellen Schiller, Sangeeta Mallik, Kathleen Hebbeler, Tracy Huang, Harold Javitz, Camille Marder, Katherine Nagle, Debra Shaver, Mary Wagner, and Cyndi Williamson. 2010. *Patterns in the Identification of and Outcomes for Children and Youth with Disabilities: Final Report.* NCEE 2010-4005. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Bloom, Howard S. 2012. "Modern Regression Discontinuity Analysis." *Journal of Research on Educational Effectiveness* 4, 1: 43-82.

Bloom, Howard S., Stephen W. Raudenbush, Michael Weiss, and Kristin Porter. Under review. "Using Multi-Site Evaluations to Study Variation in Program Effects." Working paper.

Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. "Comprehensive School Reform and Achievement: A Meta-Analysis." *Review of Educational Research* 73, 2: 125-230.

Bradley, M. C., Tamara Daley, Marjorie Levin, Fran O'Reilly, Amanda Parsad, Anne Robertson, and Alan Werner. 2011. *IDEA National Assessment Implementation Study.* NCEE 2011-4027. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Bradley, Renee, Louis Danielson, and Daniel P. Hallahan. 2002. *Identification of Learning Disabilities: Research to Practice.* Mahwah, NJ: Lawrence Erlbaum Associates.

Burns, Darci A. 2011. "Examining the Effect of an Overt Transition Intervention on the Reading Development of At-Risk English-Language Learners in First Grade." Unpublished paper. Eugene: University of Oregon.

Burns, Matthew K., and James E. Ysseldyke. 2005. "Comparison of Existing Response-to-Intervention Models to Identify and Answer Implementation Questions." *California School Psychologist* 10: 9-20.

Callender, Wayne A. 2007. "The Idaho Results-Based Model: Implementing Response to Intervention Statewide." Pages 331-342 in Shane R. Jimerson, Matthew K. Burns, and Amanda M. VanDerHeyden (eds.), *Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention.* New York: Springer.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik. 2014. "Robust Data-Driven Inference in the Regression-Discontinuity Design." *Stata Journal* 14, 4: 909-946.

Carnine, Douglas W. 1997. "Bridging the Research-to-Practice Gap." Pages 363-373 in John Lloyd, Edward J. Kameenui, and David Chard (eds.), *Issues in Educating Students with Disabilities.* Mahwah, NJ: Lawrence Erlbaum Associates.

Case, Lisa Pericola, Deborah L. Speece, Rebecca Silverman, Kristen D. Ritchey, Christopher Schatschneider, David H. Cooper, Elizabeth Montanaro, and Dawn Jacobs. 2010. "Validation of a Supplemental Reading Intervention for First-Grade Children." *Journal of Learning Disabilities* 43, 5: 402-417.

Compton, Donald L., Douglas Fuchs, Lynn S. Fuchs, Bobette Bouton, Jennifer K. Gilbert, Laura A. Barquero, Eunsoo Cho, and Robert C. Crouch. 2010. "Selecting At-Risk First-Grade Readers for Early Intervention: Eliminating False Positives and Exploring the Promise of a Two-Stage Gated Screening Process." *Journal of Educational Psychology* 102, 2: 327.

Compton, Donald L., Douglas Fuchs, Lynn S. Fuchs, and Joan D. Bryant. 2006. "Selecting At-Risk Readers in First Grade for Early Intervention: A Two-Year Longitudinal Study of Decision Rules and Procedures." *Journal of Educational Psychology* 98, 2: 394-409.

Connor, Carol M., Paul A. Alberto, Donald L. Compton, and Rollanda E. O'Connor. 2014. *Improving Reading Outcomes for Students with or at Risk for Reading Disabilities.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Special Education Research.

Cook, Thomas D. 2008. "'Waiting for Life to Arrive': A History of the Regression-Discontinuity Design in Psychology, Statistics and Economics." *Journal of Econometrics* 142, 2: 636-654.

Corman, Stephen Q. 2013. *Revenues and Expenditures for Public Elementary and Secondary Education: School Year 2010-11 (Fiscal Year 2011).* NCES 2013-342. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Denton, Carolyn A., Kim Nimon, Patricia G. Mathes, Elizabeth A. Swanson, Caroline Kethley, Terri B. Kurz, and Minyi Shih. 2010. "Effectiveness of a Supplemental Early Reading Intervention Scaled Up in Multiple Schools." *Exceptional Children* 76, 4: 394-416.

DesJardins, Stephen, and Brian McCall. 2008. "The Impact of the Gates Millennium Scholars Program on the Retention, College Finance- and Work-Related Choices, and Future Educational Aspirations of Low-Income Minority Students." Unpublished paper.

Ebaugh, Jane Cover. 2000. "The Effects of Fluency Instruction on the Literacy Development of At-Risk First Graders." Unpublished paper. New York: Fordham University.

Ehri, Linnea C., Lois G. Dreyer, Bert Flugman, and Alan Gross. 2007. "Reading Rescue: An Effective Tutoring Intervention Model for Language-Minority Students Who Are Struggling Readers in First Grade." *American Educational Research Journal* 44, 2: 414-448.

Fan, Jianqing, and Irene Gijbels. 1996. *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability* 66. Boca Raton, FL: CRC Press.

Fletcher, Jack M., W. Alan Coulter, Daniel J. Reschly, and Sharon Vaughn. 2004. "Alternative Approaches to the Definition and Identification of Learning Disabilities: Some Questions and Answers." *Annals of Dyslexia* 54, 2: 304-331.

Fletcher, Jack M., and Sharon Vaughn. 2009. "Response to Intervention Models as Alternatives to Traditional Views of Learning Disabilities: Response to the Commentaries." *Child Development Perspectives* 3, 1: 48-50.

Fuchs, Douglas, Donald L. Compton, Lynn S. Fuchs, Joan Bryant, and G. Nicole Davis. 2008. "Making 'Secondary Intervention' Work in a Three-Tier Responsiveness-to-Intervention Model: Findings from the First-Grade Longitudinal Reading Study of the National Research Center on Learning Disabilities." *Reading and Writing* 21, 4: 413-436.

Fuchs, Douglas, Lynn S. Fuchs, and Sharon Vaughn. 2008. *Response to Intervention: A Framework for Reading Educators.* Newark, DE: International Reading Association.

Fuchs, Lynn S., Stanley L. Deno, and Phyllis K. Mirkin. 1984. "The Effects of Frequent Curriculum-Based Measurement and Evaluation on Pedagogy, Student Achievement, and Student Awareness of Learning." *American Educational Research Journal* 21, 2: 449-460.

Fuchs, Lynn S., and Douglas Fuchs. 2002. "What Is Scientifically-Based Research on Progress Monitoring?" Website: www.studentprogress.org.

Fuchs, Lynn S., Douglas Fuchs, and Donald L. Compton. 2004. "Monitoring Early Reading Development in First Grade: Word Identification Fluency Versus Nonsense Word Fluency." *Exceptional Children* 71, 1: 7-21.

Fuchs, Lynn S., Douglas Fuchs, Carol L. Hamlett, and Pamela M. Stecker. 1991. "Effects of Curriculum-Based Measurement and Consultation on Teacher Planning and Student Achievement in Mathematics Operations." *American Educational Research Journal* 28, 3: 617-641.

Fuhrman, Susan H., and Richard F. Elmore. 2004. *Redesigning Accountability Systems for Education.* New York: Teachers College Press.

Gallagher, Lawrence, Barbara Means, and Christine Padilla. 2008. *Teachers' Use of Student Data Systems to Improve Instruction: 2005 to 2007.* Washington, DC: U.S. Department of Education.

Gersten, Russell M., Donald L. Compton, Carol M. Connor, Joseph Dimino, Lana Santoro, Sylvia Linan-Thompson, and W. David Tilly. 2009. "Assisting Students Struggling with Reading: Response to Intervention and Multi-Tier Intervention in the Primary Grades." Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
Website: http://ies.ed.gov/ncee/wwc/PracticeGuide.aspx?sid=3.

Gibbs, Susan Elizabeth Lynn. 2001. "Effects of a One-to-One Phonological Awareness Intervention on First Grade Students Identified as At Risk for the Acquisition of Beginning Reading." Unpublished paper. Columbia: University of South Carolina, College of Education.

Gilbert, Jennifer K., Donald L. Compton, Douglas Fuchs, Lynn S. Fuchs, Bobette Bouton, Laura A. Barquero, and Eunsoo Cho. 2013. "Efficacy of a First-Grade Responsiveness-to-Intervention Prevention Model for Struggling Readers." *Reading Research Quarterly* 48, 2: 135-154.

Glennan, Thomas K., Jr., Susan J. Bodilly, Jolene R. Galegher, and Kerri A. Kerr. 2004. *Expanding the Reach of Educational Reform: What Have We Learned About Scaling Up Educational Intervention?* Santa Monica, CA: RAND Corporation.

Gunn, Barbara, Anthony Biglan, Keith Smolkowski, and Dennis Ary. 2000. "The Efficacy of Supplemental Instruction in Decoding Skills for Hispanic and Non-Hispanic Students in Early Elementary School." *Journal of Special Education* 34, 2: 90-103.

Haager, Diane, Janette Klingner, and Sharon Vaughn. 2007. *Evidence-Based Reading Practices for Response to Intervention.* Baltimore: Paul H. Brookes Publishing.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69, 1: 201-209.

Harn, Beth A., Sylvia Linan-Thompson, and Gregory Roberts. 2008. "Intensifying Instruction: Does Additional Instructional Time Make a Difference for the Most At-Risk First Graders?" *Journal of Learning Disabilities* 41, 2: 115-125.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2, 3: 172-177.

IDEA (Public Law 108-446). 2004. Individuals with Disabilities Education Act (IDEA) of 2004 (Public Law 108-446).

Imbens, Guido, and Karthik Kalyanaraman. 2010. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." Unpublished paper. London: Institute for Fiscal Studies.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79, 3: 933-959.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142, 2: 615-635.

International Reading Association. 2009. *Response to Intervention: Guiding Principles for Educators from the International Reading Association.* Newark, DE: International Reading Association.

Jacob, Robin, Pei Zhu, Marie-Andrée Somers, and Howard Bloom. 2012. *A Practical Guide to Regression Discontinuity.* New York: MDRC.

Jenkins, Joseph R., Roxanne F. Hudson, and Evelyn S. Johnson. 2007. "Screening for At-Risk Readers in a Response to Intervention Framework." *School Psychology Review* 36, 4: 582-600.

Jenkins, Joseph R., and Rollanda E. O'Connor. 2002. "Early Identification and Intervention for Young Children with Reading/Learning Disabilities." In Renee Bradley, Louis Danielson, and Daniel P. Hallahan (eds.), *Identification of Learning Disabilities: Research to Practice.* Mahwah, NJ: Lawrence Erlbaum Associates.

Jenkins, Joseph R., Julia A. Peyton, Elizabeth A. Sanders, and Patricia F. Vadasy. 2004. "Effects of Reading Decodable Texts in Supplemental First-Grade Tutoring." *Scientific Studies of Reading* 8, 1: 53-85.

Jimerson, Shane R., Matthew K. Burns, and Amanda M. VanDerHeyden. 2007. "Response to Intervention at School: The Science and Practice of Assessment and Intervention." Pages 3-9 in Shane R. Jimerson, Matthew K. Burns, and Amanda M. VanDerHeyden (eds.), *Handbook of Response to Intervention.* New York: Springer.

Johnson, Evelyn, Daryl Mellard, Douglas Fuchs, and Melinda McKnight. 2006. *Responsiveness to Intervention (RtI): How to Do It.* Lawrence, KY: National Research Center on Learning Disabilities.

Johnson, Nicholas, Phil Oliff, and Erica Williams. 2011. *An Update on State Budget Cuts: At Least 46 States Have Imposed Cuts That Hurt Vulnerable Residents and Cause Job Loss.* Washington, DC: Center on Budget and Policy Priorities.

Kerins, Marie R., Dawn Trotter, and Lisa Schoenbrodt. 2010. "Effects of a Tier 2 Intervention on Literacy Measures: Lessons Learned." *Child Language Teaching and Therapy* 26, 3: 287-302.

Konstantopoulos, Spyros, and Larry V. Hedges. 2004. "Meta-analysis." Pages 281-297 in David Kaplan (ed.), *Handbook of Quantitiative Methodology for the Social Sciences*. New York: Sage.

Lee, David S., and David Card. 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics* 142, 2: 655-674.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48, 2: 281-355.

Mathes, Patricia G., Carolyn A. Denton, Jack M. Fletcher, Jason L. Anthony, David J. Francis, and Christopher Schatschneider. 2005. "The Effects of Theoretically Different Instruction and Student Characteristics on the Skills of Struggling Readers." *Reading Research Quarterly* 40, 2: 148-182.

McCardle, Peggy, Hollis S. Scarborough, and Hugh W. Catts. 2001. "Predicting, Explaining, and Preventing Children's Reading Difficulties." *Learning Disabilities Research & Practice* 16, 4: 230-239.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142, 2: 698-714.

McMaster, Kristen L., Douglas Fuchs, Lynn S. Fuchs, and Donald L. Compton. 2005. "Responding to Nonresponders: An Experimental Field Trial of Identification and Intervention Methods." *Exceptional Children* 71, 4: 445-463.

Mellard, Daryl, Melinda McKnight, and Joanna Jordan. 2010. "RTI Tier Structures and Instructional Intensity." *Learning Disabilities Research & Practice* 25, 4: 217-225.

National Association of State Directors of Special Education. 2007. *Response to Intervention Blueprints: State Level Edition.* Alexandria, VA: National Association of State Directors of Special Education, Inc., and Council of Administrators of Special Education.

National Association of State Directors of Special Education. 2008a. *Response to Intervention Blueprints: District Level Edition.* Alexandria, VA: National Association of State Directors of Special Education, Inc., and Council of Administrators of Special Education.

National Association of State Directors of Special Education. 2008b. *Response to Intervention Blueprints: School Building Level Edition.* Alexandria, VA: National Association of State Directors of Special Education, Inc., and Council of Administrators of Special Education.

National Center on Response to Intervention. 2010. *Essential Components of RtI: A Closer Look at Response to Intervention.* Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention.

National Reading Panel. 2000. "Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction." Washington, DC: U.S. Department of Health and Human Services, National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development. Website: http://www.nichd.nih.gov/publications/pubs/nrp/pages/smallbook.aspx?renderforprint=1.

National Research Council. 2002. *Minority Students in Special and Gifted Education. Committee on Minority Representation in Special Education.* M. Suzanne Donovan and Christopher T. Cross (eds.). Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Nichols, Austin. 2007. "Causal Inference with Observational Data." *State Journal* 7, 4: 507-541.

Nomi, Takako, and Stephen W. Raudenbush. 2012. "Understanding Treatment Effects Heterogeneities Using Multi-Site Regression Discontinuity Designs: Example from a 'Double-Dose' Algebra Study in Chicago." Paper presented at the Society for Research on Educational Effectiveness conference, Washington, DC.

O'Connor, Rollanda E., Kathleen M. Bocian, Kristen D. Beach, Victoria Sanchez, and Lindsay J. Flynn. 2013. "Special Education in a 4-Year Response to Intervention (RtI) Environment: Characteristics of Students with Learning Disability and Grade of Identification." *Learning Disabilities Research & Practice* 28, 3: 98-112.

O'Connor, Rollanda E., Kathleen M. Bocian, Kristen D. Beach, Victoria Sanchez, and Lindsay J. Flynn. 2014. "Access to a Responsiveness to Intervention Model: Does Beginning Intervention in Kindergarten Matter?" *Journal of Learning Disabilities* 47, 4: 281-297.

O'Connor, Rollanda E., Deborah Fulmer, Kristin R. Harty, and Kathryn M. Bell. 2005. "Layers of Reading Intervention in Kindergarten Through Third Grade: Changes in Teaching and Student Outcomes." *Journal of Learning Disabilities* 38, 5: 440-455.

O'Connor, Rollanda E., and Joseph R. Jenkins. 1999. "Prediction of Reading Disabilities in Kindergarten and First Grade." *Scientific Studies of Reading* 3, 2: 159-197.

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. 2009. *What to Do When Data Are Missing in Group Randomized Controlled Trials.* NCEE 2009-0049. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Ransford-Kaldon, Carolyn R., E. Sutton Flynt, Cristin L. Ross, Louis Franceschini, Todd Zoblotsky, Ying Huang, and Brenda Gallagher. 2010. *Implementation of Effective Intervention: An Empirical Study to Evaluate the Efficacy of Foutas & Pinnell's Leveled Literacy Intervention System (LLI).* Memphis, TN: University of Memphis, College of Education, Health and Human Sciences, Center for Research in Educational Policy.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thousand Oaks, CA: SAGE Publications, Inc.

Reardon, Sean F., and Stephen W. Raudenbush. 2013. "Under What Assumptions Do Site-by-Treatment Instruments Identify Average Causal Effects?" *Sociological Methods and Research* 42, 2: 143-163.

Reardon, Sean F., and Joseph P. Robinson. 2012. "Regression Discontinuity Designs with Multiple Rating-Score Variables." *Journal of Research on Educational Effectiveness* 5, 1: 83-104.

RTI Action Network. 2015. "RTI Action Network: A Program of the National Center for Learning Disabilities." Website: http://www.rtinetwork.org/.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* Hoboken, NJ: John Wiley & Sons.

Rubin, Donald B. 1996. "Multiple Imputation After 18+ Years." *Journal of the American Statistical Association* 91, 434: 473-489.

Scarborough, Hollis S. 1998. *Early Identification of Children at Risk for Reading Disabilities: Phonological Awareness and Some Other Promising Predictors.* Timonium, MD: York Press.

Schafer, Joseph L. 1999. "Multiple Imputation: A Primer." *Statistical Methods in Medical Research* 8, 1: 3-15.

Schochet, Peter Z., Thomas D. Cook, John Deke, Guido Imbens, J. R. Lockwood, Jack Porter, and J. Smith. 2010. "Standards for Regression Discontinuity Designs." Website: http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

Shapiro, Edward S., Naomi Zigmond, Teri Wallace, and Doug Marston. 2011. *Models for Implementing Response to Intervention: Tools, Outcomes, and Implications.* New York: Guilford Press.

Snow, Catherine E., M. Susan Burns, and Peg Griffin. 1998. *Preventing Reading Difficulties in Young Children.* Washington, DC: National Academy of Sciences, National Research Council.

Speece, Deborah L., Christina Mills, Kristen D. Ritchey, and Elgen Hillman. 2003. "Initial Evidence That Letter Fluency Tasks Are Valid Indicators of Early Reading Skill." *Journal of Special Education* 36, 4: 223-233.

Speece, Deborah L., Kimberly Palombo, and Jamey Burho. 2013. "From FAPE to FEPE: Toward an Excellent Public Education for Children and Youth with Learning Disabilities." Pages 104-118 in H. Lee Swanson, Karen R. Harris, and Steve Graham (eds.), *Handbook of Learning Disabilities,* 2nd ed. New York: Guilford Press.

Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." Pages 80-108 in Donald W. K. Andrews and James H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. New York: Cambridge University Press.

Torgesen, Joseph K. 2009. "The Response to Intervention Instructional Model: Some Outcomes from a Large-Scale Implementation in Reading First Schools." *Child Development Perspectives* 3, 1: 38-40.

Torgesen, Joseph K., Ann W. Alexander, Richard K. Wagner, Carol A. Rashotte, Kytja K. S. Voeller, and Tim Conway. 2001. "Intensive Remedial Instruction for Children with Severe Reading Disabilities: Immediate and Long-Term Outcomes from Two Instructional Approaches." *Journal of Learning Disabilities* 34, 1: 33-58.

Torgesen, Joseph K., Richard K. Wagner, and Carol A. Rashotte. 1999. "TOWRE: Test of Word Reading Efficiency." Austin, TX: Pro-Ed.
Website: http://www.mindresources.com/pdfs/SR2011/pages/PAGE_202.pdf.

U.S. Department of Education. 2015. "Office of Special Education Programs (OSEP)." Website: http://www2.ed.gov/about/offices/list/osers/osep/index.html.

U.S. Department of Education, Office of Special Education and Rehabilitative Services. 2002. *A New Era: Revitalizing Special Education for Children and Their Families.* Washington, DC: U.S. Department of Education.

Vadasy, Patricia F., Joseph R. Jenkins, Lawrence R. Antil, Susan K. Wayne, and Rollanda E. O'Connor. 1997. "The Effectiveness of One-to-One Tutoring by Community Tutors for At-Risk Beginning Readers." *Learning Disability Quarterly* 20, 2: 126-139.

Vadasy, Patricia F., Joseph R. Jenkins, and Kathleen Pool. 2000. "Effects of Tutoring in Phonological and Early Reading Skills on Students at Risk for Reading Disabilities." *Journal of Learning Disabilities* 33, 6: 579-590.

Vadasy, Patricia F., Elizabeth A. Sanders, and Julia A. Peyton. 2005. "Relative Effectiveness of Reading Practice or Word-Level Instruction in Supplemental Tutoring: How Text Matters." *Journal of Learning Disabilities* 38, 4: 364-380.

Vadasy, Patricia F., Elizabeth A. Sanders, and Julia A. Peyton. 2006. "Paraeducator-Supplemented Instruction in Structural Analysis with Text Reading Practice for Second and Third Graders at Risk for Reading Problems." *Remedial and Special Education* 27, 6: 365-378.

Vadasy, Patricia F., Elizabeth A. Sanders, and Sarah Tudor. 2007. "Effectiveness of Paraeducator-Supplemented Individual Instruction Beyond Basic Decoding Skills." *Journal of Learning Disabilities* 40, 6: 508-525.

VanDerHeyden, Amanda M., Joseph C. Witt, and David W. Barnett. 2005. "The Emergence and Possible Futures of Response to Intervention." *Journal of Psychoeducational Assessment* 23, 4: 339-361.

Vaughn, Sharon. 2008. “Tier 2: The Role of Intervention.” In Douglas Fuchs, Lynn S. Fuchs, and Sharon Vaughn (eds.), *Response to Intervention: A Framework for Reading Educators.* Newark, DE: International Reading Association.

Vaughn, Sharon, Paul T. Cirino, Sylvia Linan-Thompson, Patricia G. Mathes, Coleen D. Carlson, Elsa Cardenas Hagan, Sharolyn D. Pollard-Durodola, Jack M. Fletcher, and David J. Francis. 2006. “Effectiveness of a Spanish Intervention and an English Intervention for English-Language Learners at Risk for Reading Problems.” *American Educational Research Journal* 43, 3: 449-487.

Vaughn, Sharon, Jeanne Wanzek, Christy S. Murray, Nancy Scammacca, Sylvia Linan-Thompson, and Althea L. Woodruff. 2009. “Response to Early Reading Intervention Examining Higher and Lower Responders.” *Exceptional Children* 75, 2: 165-183.

Vellutino, Frank R., Donna M. Scanlon, Edward R. Sipay, Sheila G. Small, Alice Pratt, RuSan Chen, and Martha B. Denckla. 1996. “Cognitive Profiles of Difficult-to-Remediate and Readily Remediated Poor Readers: Early Intervention as a Vehicle for Distinguishing Between Cognitive and Experiential Deficits as Basic Causes of Specific Reading Disability.” *Journal of Educational Psychology* 88, 4: 601-638.

Wanzek, Jeanne, and Sharon Vaughn. 2008. “Response to Varying Amounts of Time in Reading Intervention for Students with Low Response to Intervention.” *Journal of Learning Disabilities* 41, 2: 126-142.

Wanzek, Jeanne, and Sharon Vaughn. 2010. “Tier 3 Interventions for Students with Significant Reading Problems.” *Theory Into Practice* 49, 4: 305-314.

Wehby, Joseph H., Katherine B. Falk, Sally Barton-Arwood, Kathleen L. Lane, and Cristy Cooley. 2003. “The Impact of Comprehensive Reading Instruction on the Academic and Social Behavior of Students with Emotional and Behavioral Disorders.” *Journal of Emotional and Behavioral Disorders* 11, 4: 225-238.

Wing, Coady, and Thomas Cook. 2013. “Strengthening the Regression Discontinuity Design Using Additional Design Elements: A Within-Study Comparison.” *Journal of Policy Analysis and Management* 32, 4: 853-877.

Wong, Vivian C., Peter M. Steiner, and Thomas D. Cook. 2013. “Analyzing Regression Discontinuity Designs with Multiple Assignment Variables: A Comparative Study for Four Estimation Methods.” *Journal of Educational and Behavioral Statistics* 38, 2: 107-141.

Appendix A

Members of the Technical Working Group

This page has been left blank for double-sided copying.

The following experts attended or participated in one or more meetings of the Technical Working Group (TWG) advising the study research team.

Donald Compton, Vanderbilt University

Carol Connor, Arizona State University and Florida Center for Reading Research

Judy Elliott, Los Angeles Unified School District

David Francis, University of Houston

Paul McDermott, University of Pennsylvania

Rollanda O'Connor, University of California Riverside

Sean F. Reardon, Stanford University

Amy Sichel, Harvard Family Research Project and Abington School District

Jeffrey Smith, University of Michigan

Deborah Speece, University of Maryland (prior to her appointment as IES Commissioner of the National Center for Special Education Research)

Sharon Vaughn, The Meadows Center for Preventing Educational Risk and University of Texas at Austin

This page has been left blank for double-sided copying.

Appendix B

Data Collection

This page has been left blank for double-sided copying.

Appendix B of this report on the Response to Intervention (RtI) evaluation discusses the selection of research samples and the collection of data used in the impact analyses.

Sample Selection

This section provides additional detail on the selection of the impact sample and the reference sample described in Chapter 2.

Impact Sample Selection

Based on nominations received from experts, the study research team sent an introductory letter or email about the RtI study to the nominated school districts or to districts with at least one nominated school in the 2010-11 school year. During a follow-up phone call, the study research team confirmed that the nominated schools in that district met the study's criteria. The team also asked whether other schools in the district might meet the study criteria. After obtaining district permission to contact the school leadership in potential sites, the study research team collected details about the schools' RtI implementation in Grades 1 to 3 by email or phone.

Schools that said that they use universal screening tests and had prespecified cut scores on those tests were asked to complete a brief questionnaire before the site visit. The form sought more details about the reading curricula used in various tiers, the measures used for screening and progress monitoring and their frequency, and the year that each of the four key RtI practices listed in Chapter 2 was first implemented in Grades 1 to 3. This process verified that schools met the eligibility criteria for experience with implementation of key aspects of the RtI system.

During site visits in 2011, the study research team discussed a memorandum of understanding regarding obligations involved in participation in the evaluation study — including providing data on fall benchmark-test scores and tier placements so that the study research team could verify whether the school uses a quantifiable rule to assign students to tiers.

The analysis to verify a school's use of a decision rule — described in Chapter 2 as Stage 4 of the selection process — determines the degree to which a school "complies" with its decision rule. For example, if a school says that students scoring below 24 (the cut point) on a screening test are assigned to receive intervention services, and all students scoring below 24 are, indeed, placed in Tiers 2 or 3 and no students scoring at or above 24 are placed in those tiers, then that school "complies" 100 percent with its decision rule. This means that the intended assignment to tiers as determined by the decision rules completely predicts actually being assigned to receive intervention services. This relationship between the intended assignment and the actual assignment should ensure that students who score below the cut point should have a 100 percent probability of being in the treatment group, while students who score at or above the cut point should have a 100 percent probability of being in the comparison group.

However, the study research team recognized the possibility that most schools would likely not demonstrate 100 percent compliance. To be able to use an analytic approach called "Two-Stage Least Squares (2SLS) regression" to estimate the impact of being assigned to receive intervention,[1] there needs to be an instrumental variable that can predict the compliance well.[2]

In the context of this study, the *instrumental variable* is the intended treatment assignment; the compliance is based on the actual treatment assignment; and the strength of the instrumental variable is assessed by (1) the magnitude of the estimated compliance rate and (2) the significance level of the instrumental variable in predicting the actual assignment. The study research team estimated these two factors for each school-by-grade combination for 15 percent of students right above the cut point and for 15 percent of students right below the cut point and then calculated the compliance rate and the strength of the instrument for this subsample.[3]

Specifically, the study research team set up three categories to classify school-by-grade units based on their F-statistics and their compliance rates at the cut point.[4] The table at the top of the next page lists these categories. Note that even though Category 1 allowed any compliance rate, when the F-statistic was greater than 10, the compliance rate at the cut point was usually fairly sizable.

As a result of this screening process, some schools dropped out of the sample because of low compliance or weak predicted power of the intended assignment. For schools that remained in the sample after this process, the average compliance rate at the cut point was around 63 percent for Grade 1, 64 percent for Grade 2, and 69 percent for Grade 3. The rates vary by school as well, ranging from around 10 percent to 100 percent in each grade.[5]

---

[1]For more details about this method, see Appendix E.

[2]This is usually referred to as a "strong instrument" in the literature. Stock and Yogo (2005) discuss the potential bias in the estimates as a result of weak instrument in the 2SLS approach.

[3]Since this analysis was carried out on a rolling basis as each school submitted fall screening data, and since the study research team did not have the outcome data at that point, it was not possible to do optimal bandwidth selection at that point. To mimic the impact analysis, the study research team decided to use a subsample that consists of 15 percent of students who are closest on either side of the cut point.

[4]The compliance rate was calculated as the mean difference in compliance between observations on either side of the cut point within the subsample close to the cut point. The F-statistics (or their corresponding p-values) came from a set of regressions that control for the screening test scores using different functional forms. Note that this analysis was preliminary and was used only for site recruitment purposes.

[5]Note that these numbers are based only on the fall screening data submitted by schools by January of the 2011-12 school year. They do not necessarily reflect the realized compliance rates for the final analysis sample, which are defined based on the availability of outcome data as well.

Category (by Priority, High to Low)	Compliance Rate at Cut Point	Strength as Instrument
1	any	F ≥ 10 (p-value of around 0.002, depending on degrees of freedom)
2	≥15%	p-value ≤ 0.05 (F of 3.9 to 4.0, depending on degrees of freedom)
3	≥20%	p-value ≤ 0.10 (F of 2.7 to 2.8, depending on degrees of freedom)

Even though the overall compliance rate is fairly high, the observed variation in compliance across schools based on preliminary data could be an indication that other factors, such as resource constraints, teachers' judgments, or other issues, might have been at play when students' tier placements were determined. Unfortunately information on these factors is not available to the study. Also note that the use of teacher discretion in actual assignment of students to intervention would not necessarily bias the estimated impacts of intervention on those receiving it, as long as the rating variable was not manipulated. Appendix F examines that possibility and finds no conclusive evidence of rating manipulation.

In sum, a school was included in the study if:

- It had been implementing all four key RtI practices since 2009-10.
- It had a quantifiable practice of assigning students to tiers based on a decision rule and complied with the decision rule to a certain degree.
- It had a minimum of 30 students in each grade and/or at least 8 students in Tier 1 or Tiers 2 and 3 combined, to ensure a sufficient number of students to conduct the impact analysis.
- It was located in a state or region of the country required for geographic diversity.
- It was in a district with more than one school. (This allowed the study research team to conserve travel resources and concentrate recruitment and data collection in districts that brought multiple schools to the study.)

As a result of this purposive selection, the impact sample should not be considered a random sample of all nominated and/or screened schools.

Reference Sample Selection

The study's sampling frame included all public elementary schools, charter schools, and magnet schools serving students in Grades 1 to 3 (other than the schools in the impact evaluation sample) that were present in the 2010-11 Common Core of Data Public School Universe in the 13 states from which the impact sample schools had been recruited.[6] A random sample of 100 schools was drawn from each state, with 40 additional schools available as replacements for schools that refused to participate or in case a school within the initial 100 was closed or no longer serving Grades 1 to 3 in the 2011-12 school year. In each state, at least 100 eligible schools were sent surveys in fall 2012, retrospectively asking about services provided in 2011-12. Overall, 1,105 (85 percent) of the 1,300 surveyed schools responded. This means that the reference sample schools represent 85 percent of the universe of public elementary schools serving Grades 1 to 3, wherein the universe excludes the impact sample schools. Appendix Table B.1 shows that the response rates in all states but one exceed 80 percent.

Data Collection and School Samples Used in the Analysis

The study research team began surveying school staff and testing students in April 2012. Impact sample schools agreed to have data collectors field school, teacher, and interventionist surveys. They also agreed to have the study research team test Grade 1 and 2 students on an end-of-year test. In addition, they agreed to provide winter screening test scores for students in Grades 1 to 3, demographic records for those students, and state achievement test scores for Grade 3 students by the end of summer 2012. They also agreed to provide counts of students who were identified for special education and were ages 6 to 10 (or Grades 1 to 5).

Survey Data

The study research team fielded three surveys. School-level response rates are shown in Appendix Table B.1.

- **School Survey.** All but one of the 146 impact sample schools completed the school survey. Of 1,300 randomly sampled schools, 1,105 completed it.

- **Interventionist Survey.** Of the 146 schools in the study, 5 declined to field the survey, citing concerns about staff time. Of the remaining 141 schools, 4 did not serve any reading groups that served exclusively one reading level. This left 137 unique schools used in analysis.

- **Teacher Survey.** All 146 of the impact sample schools completed the teacher survey.

[6]Website: http://nces.ed.gov/ccd/pubschuniv.asp.

The Response to Intervention (RtI) Evaluation

Appendix Table B.1

Data Collection and Response Rates from Schools for 2011-12

	Respondent/Level of Data	Number of Sites Possible	Number of Sites That Submitted Data	Response Rate (%)
Data Item				
Fall screening tests				
Grade 1	Student	142	141	99.3
Grade 2	Student	143	142	99.3
Grade 3	Student	142	142	100
Winter screening tests[a]				
Grade 1	Student	119	113	95.0
Grade 2	Student	127	123	96.9
Grade 3	Student	116	110	94.8
ECLS-K - Grade 1	Student	119	119	100
TOWRE - Grade 1	Student	119	119	100
TOWRE - Grade 2	Student	127	127	100
Student-level state achievement test - Grade 3	Student	116	112	96.6
Student-level demographics				
Grade 1	Student	119	118	99.2
Grade 2	Student	127	126	99.2
Grade 3	Student	116	113	97.4
Teacher logs				
Grade 1	Teacher/student	142	137	96.5
Grade 2	Teacher/student	143	138	96.5
Grade 3	Teacher/student	142	137	96.5
Interventionist logs				
Grade 1	Interventionist/student	142	133	93.7
Grade 2	Interventionist/student	143	133	93.0
Grade 3	Interventionist/student	142	132	93.0
Special education identifications by age				
Total counts	School	146	128	87.7
New identifications	School	146	116	79.5

(continued)

Appendix Table B.1 (continued)

	Respondent/Level of Data	Number of Sites Possible	Number of Sites That Submitted Data	Response Rate (%)
Teacher survey	Teacher	146	146	100
Interventionist survey	Interventionist	146	141	96.6
School survey				
Impact schools	School	146	145	99.3
Reference schools, by state	School	1,300	1,105	85.0
Arizona	School	100	90	90.0
California	School	100	81	81.0
Florida	School	100	84	84.0
Illinois	School	100	90	90.0
Massachusetts	School	100	84	84.0
Minnesota	School	100	81	81.0
Missouri	School	100	85	85.0
Montana	School	100	87	87.0
Pennsylvania	School	100	87	87.0
Texas	School	100	88	88.0
Utah	School	100	75	75.0
Washington	School	100	85	85.0
Wyoming	School	100	88	88.0

SOURCE: Response rates tabulated by study research team.

NOTE: [a]Sites possible for winter screening tests are based on schools that qualified for the study based on meeting fall data requirements for the Regression Discontinuity Design impact analysis. The "number of sites possible" reflects the number of schools admitted to the study based on fall data; the "number of sites that submitted data" reflects the number that submitted winter screening-test results.

Details about the construction of variables, coding, and the final analysis sample are described in Appendix C.

Screening Test Score and Tier Placement Data

Schools provided fall and winter screening test scores and tier placement information. Schools within the sample used a variety of terminology and systems to classify student tier placements. Schools submitting fall data used such classifications as "Benchmark," "Green," "Basic," or "Core" to indicate a Tier 1 placement. "Strategic," "Yellow," and "Emerging" were associated with Tier 2, while "Intensive," "Red," and "Deficient" indicated a Tier 3 placement. Across all three grades, only 20 students in the fall data lacked any tier placement information whatsoever. These students were dropped from the analysis sample.

The impact analysis sample for Grade 1 includes 119 schools with a minimum number of students in Tier 2 or 3. However, for the tier movement analysis in Chapter 4, a significant number of schools needed to be dropped. Appendix Table B.2 describes the schools that are either used in or excluded from the tier movement analysis. That analysis is restricted to schools that showed at least one student in each of the three tiers in the fall and winter data. This restriction allowed the study research team to describe movement or stability among schools with a full range of tiers; inclusion of schools without Tier 3 in either fall or winter, for example, may have misrepresented the proportion of students moving into or out of Tier 3. For Grade 1, 12 schools did not submit winter tier placement data. Of the remaining schools that did, 14 schools assigned students to only two different tier levels in the winter. Another 4 assigned students to only two different tier levels in the fall.

After removing schools for the reasons mentioned above, the tier movement sample for Grade 1 drops to 89 schools. Similar attrition occurs in the other grades. The Grade 2 sample drops from 127 to 102. The Grade 3 sample drops from 116 to 93. In Appendix Table B.2, the top row for each grade shows the number of schools that met the required conditions; the other rows show the number of schools that were dropped under each condition.

National Center for Education Statistics' Common Core of Data (CCD)

These data are used for two purposes in the report.

1. The first purpose in using the CCD is to describe the characteristics of the schools in the impact sample and reference sample and to present characteristics of all schools in the 13 states in the study. The Public School Universe data set provides information on school-level characteristics.

Appendix Table B.3 describes the variables used in the analysis. Characteristics are presented for 1,105 reference sample schools and for 144 impact sample schools. (This is not the full impact sample of 146 schools because one school did not complete the school survey and one was a split-grade school whose characteristics are represented by its lower-grade counterpart.) Appendix Table B.4 presents results for the census of schools in the 13 study states as a companion to Table 3.1 (Chapter 3). For most variables, the rates of missing data on variables from the CCD are less than 3 percent, and the difference in rates between samples is not statistically significant.

Some information, such as the proportion of students who are English Language Learners (ELL students) or who are identified with an Individualized Education Program (IEP), is reported at the district level and appears in a separate data set — the Local Education Agency (School District) Universe Survey — in order to preserve anonymity of students. These data

The Response to Intervention (RtI) Evaluation

Appendix Table B.2

Sample Definition for the Tier Movement Analysis in Chapter 4

	Winter Data Submitted	Full Range of Tiers in Fall Data	Full Range of Tiers in Winter Data	Number of Schools
Grade 1				
Schools used in analysis	Yes	Yes	Yes	89
Schools excluded from analysis	Yes	Yes	No	5
	Yes	No	Yes	4
	Yes	No	No	9
	No	Yes	No	11
	No	No	No	--
Grade 2				
Schools used in analysis	Yes	Yes	Yes	102
Schools excluded from analysis	Yes	Yes	No	7
	Yes	No	No	4
	No	Yes	No	14
Grade 3				
Schools used in analysis	Yes	Yes	Yes	93
Schools excluded from analysis	Yes	Yes	No	6
	Yes	No	Yes	--
	Yes	No	No	--
	No	Yes	No	13

SOURCES: Fall 2011 and winter 2012 tier placement data.

NOTE: Schools with only two tiers were classified as missing a full range of tiers. "--" indicates that a value has been suppressed due to small cell size.

The Response to Intervention (RtI) Evaluation

Appendix Table B.3

Variables Used for Sample Characteristics in Chapter 3

Variable Presented in Table 3.1	Data Source	Short Description of Variable Creation
School size (Grades 1-3)	CCD[a]	Number of students in Grades 1-3, reflecting the sum of all the race/ethnic category counts for each grade. (This is also the denominator in the race/ethnic and male percentage items: Stu_GRADES1TO3)
Race/ethnicity: Asian (%)	CCD[a]	Numerator: sum of Asian and Pacific Islander students in Grades 1-3. Denominator: Stu_GRADES1TO3
Race/ethnicity: Black (%)	CCD[a]	Numerator: sum of black students in Grades 1-3. Denominator: Stu_GRADES1TO3
Race/ethnicity: White (%)	CCD[a]	Numerator: sum of white students in Grades 1-3. Denominator: Stu_GRADES1TO3
Race/ethnicity: Hispanic (%)	CCD[a]	Numerator: sum of Hispanic students in Grades 1-3. Denominator: Stu_GRADES1TO3
Race/ethnicity: Other (%)	CCD[a]	Numerator: sum of "Other" students (including Native American and Two or More Races) in Grades 1-3. Denominator: Stu_GRADES1TO3
Male (%)	CCD[a]	Numerator: sum of male students in Grades 1-3. Denominator: Stu_GRADES1TO3
Locale (urban, suburban, town, rural)	CCD[a]	0/100 dummy based on LOCALE variable
Enrollment	CCD[a]	On original data set
Title I eligible school	CCD[a]	On original data set

(continued)

Appendix Table B.3 (continued)

Variable Presented in Table 3.1	Data Source	Short Description of Variable Creation
Charter or magnet school	CCD[a]	0/100 dummy based on Charter and Magnet variables from CCD data set
Low-income students (% of students qualifying for free/reduced-price lunch)	CCD[a]	Numerator: number of students receiving free or reduced-price lunch. Denominator: number of total students in the school (all grades). Schools that are missing number of free/reduced-price lunch were missing for this variable; no values were imputed.
Number of full-time-equivalent staff	CCD[a]	As in the original data set
English Language Learners (% of students)	LEA[b]	Created by dividing the number of ELL students in the district by the total number of students in the district
Individualized Education Programs (% of students)	LEA[b]	Created by dividing the number of special education students in the district by the total number of students in the district

SOURCES: Study research team analyses of responses to the school survey.

NOTES: [a]U.S. Department of Education, National Center for Education Statistics, Public Elementary/Secondary School Universe Survey School Years 2010-11 and 2009-10 (Common Core of Data).
[b]Local Education Agency (School District) Universe Survey 2010-11.

were not available by grade. Missing data are somewhat of a problem in this data set. The percentage of missing data for ELL students is 6.3 percent of impact sample schools and 8.0 percent of reference sample schools (p-value = 0.429). Of the 20,360 schools in the 13-state sample, ELL data were missing for 5,683 (28 percent), and IEP data were missing for 268 (1 percent). Some schools did not have 2010 data available. In order to preserve sample size, 2009 CCD data were used for variables missing in the 2010 data. For variables used in race/ethnicity and gender calculations, this recoding occurred in 8 percent of schools. For calculations of the full-time-equivalent staff, this recoding occurred in about 15 percent of schools. For all other variables, this recoding occurred in less than 3 percent of schools. As

The Response to Intervention (RtI) Evaluation

Appendix Table B.4

Characteristics of the 13-State School Universe Serving Grades 1-3

Characteristic	13-State Mean
Race/ethnicity[a] (%)	
Asian	5.5
Black	12.3
White	44.7
Hispanic	33.3
Other	4.2
Sex[a] (%)	
Male	51.4
Locale (%)	
Urban	34.8
Suburban	33.6
Town	7.9
Rural	23.7
Poverty[b] (%)	53.1
Title I eligible schools[c] (%)	77.6
Charter and magnet schools[c] (%)	9.8
Average school size (number of students in all grades)	506.5
Average school size (number of students in Grades 1-3)	230.5
Number of full-time-equivalent staff	30.0
English Language Learners[d] (%)	8.9
Individualized Education Program[d] (%)	12.3
Deviation from state mean of percentage proficient on Grade 3 standardized state reading test	NA
Number of Schools	20,360

SOURCES: Common Core of Data, including The U.S. Department of Education, National Center for Education Statistics, Public Elementary/Secondary School Universe Survey School Years 2010-11 and 2009-10, and the Local Education Agency (School District) Universe Survey 2010-11. State achievement data downloaded from 13 state websites.

(continued)

Appendix Table B.4 (continued)

NOTES: Omnibus tests were conducted comparing both reference and impact samples to the 13-state sample on all presented measures. For both omnibus tests, the differences were significant (p-value < 0.001).

Some schools did not have 2010 data available. In these cases, 2009 data were used, where available, for variables missing in the 2010 data.

[a]Race/ethnicity and sex calculations are based on school-level student populations in Grades 1 through 3.

[b]"Poverty" indicates school data on proportion of students receiving free or reduced-price lunch.

[c]Title I and charter and magnet status variables took on values of zero or 100. The means represent the percentage of schools of each type in each sample.

[d]English Language Learners (ELL) and Individualized Education Program (IEP) data come from district-level data and thus are based on district-level student populations.

shown in Appendix Table B.5, weighted rates of recoding from 2009-10 data are described in increasing order of recoding rates.

School-level characteristics, such as whether the school was eligible for Title I funds, are also used in the exploratory impact analysis.

2. The second purpose of using the CCD in the study is to provide information on total student enrollment in order to calculate special education identification rates as a proportion of total enrollment, by age group. This is discussed further below.

Special Education Identification Data

To obtain accurate counts of students identified with an Individualized Education Program (IEP) at the school level, and recognizing that counts vary as students get older, the study research team sought the total counts of students identified under each of the federal disability categories from all schools in the impact sample as of fall 2011. These data were collected from individual schools during summer 2012. Statewide data for children with disabilities who were served under the Individuals with Disabilities Education Act (IDEA)-Part B were downloaded from the Office of Special Education Programs (OSEP), Data Accountability Center, on April 30, 2013.[7]

States reported 13 disability categories, as well as an "All disabilities" category, which is the sum of 13 individual disability categories (hearing impairments, speech or language

[7]Website: http://tadnet.public.tadnet.org/pages/712.

The Response to Intervention (RtI) Evaluation

Appendix Table B.5

Weighted Rates of Recoding Data for Sample Characteristics

	Number of Schools Recoded		Percentage of Schools Recoded (%)		P-Value of Difference in Recoding Rates
	Impact	Reference	Impact	Reference	Between Samples
Indicators					
Charter schools	0	0	0.0	0.0	NA
Magnet schools	0	0	0.0	0.0	NA
Locale (urban, rural, etc.)	0	0	0.0	0.0	NA
Number of students in all grades	0	3	0.0	0.2	0.098
Title I eligible schools	0	3	0.0	0.2	0.098
Number of students receiving free or reduced-price lunch	0	4	0.0	0.3	0.066
Total number of students in Grades 1-3	6	95	4.2	7.9	0.054
Number of students in a given racial/ethnic category across Grades 1-3	6	95	4.2	7.9	0.054
Number of male students in Grades 1-3	6	95	4.2	7.9	0.054
Number of full-time-equivalent staff in entire schools	13	171	9.0	25.6	0.000

SOURCE: School survey.

NOTES: As in Table 3.1, the numbers of schools presented here include only schools that responded to the survey and served Grades 1-3. Reference sample percentages and p-values presented are calculated using a sampling weight.

Some schools did not have 2010 data available. In these cases, 2009 data were used, where available, for variables missing in the 2010 data.

Sampling weights are described in Appendix C.

impairments, visual impairments, emotional disturbance, orthopedic impairments, other health impairments, Specific Learning Disabilities [SLDs], deaf-blindness, multiple disabilities, autism, traumatic brain injury, and developmental delay).

Schools were provided a template that corresponded to these categories and were asked to enter the number of students identified within each category as of fall 2011 and the total number of students with an IEP. Up to 131 schools provided this information. They were also asked to provide the total number of students newly identified in the spring. Fewer than 75 schools provided this information, so new identifications are not included in the analysis.

Due to variations in state reporting, total counts can be compared across states, but some categories cannot be compared. For example, reporting in the "developmental delay" category varies across states. California, Florida, and Texas do not report counts for this category. Arizona, Illinois, and Massachusetts report developmental delay for ages 6 to 9. Minnesota, Missouri, Pennsylvania, and Montana report developmental delay for ages 3 to 6. Utah and Wyoming report developmental delay for ages 3 to 9, and Washington reports it for ages 3 to 8. Furthermore, Florida also does not report in the "multiple disabilities" category; students and children with multiple disabilities are reported according to their primary disability. As a result, the study research team decided to focus on total counts as well as counts of students identified with a Specific Learning Disability (SLD). In the state data, if a category had 1 to 5 students, the data were suppressed, to preserve anonymity of the student or school. However, total counts reflect all students identified, even if the count for a specific category is suppressed.

In order to calculate rates of IEP identification, the study research team used the number of students identified with an SLD in the numerator and the total enrollment in the denominator, for each respective age group. All but 16 schools provided total enrollment information. For the state, since 2011 data were not yet released at the time of the analysis, the study research team used the most recent enrollment data (fall 2010) from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data, Public School Universe Data. The calculated rates were then averaged across all study states, by age. Based on availability of enrollment data and IEP counts, up to 128 impact sample schools were included in the analysis.

State Achievement Data

From each state, the study research team obtained data regarding reference and impact sample schools' performance on the respective third-grade state achievement test in 2010-11. These data came either from public files posted on a state website or by request from a state data repository. These data provided an indicator of school-level performance *before the study year,* in terms of the percentage of students reading at or above proficient. To make schools comparable across states, the study research team calculated the difference between a school's percentage of students “at or above proficient” and that state’s mean across all schools of the percentage of students “at or above proficient” on the respective state test.

Of the 1,249 schools from both samples that responded to the survey, state reading test data for spring 2011 were available for 1,209 schools (97 percent). Deviations from the state mean were missing for 1.4 percent of impact sample schools and for 3.4 percent of reference sample schools (p-value of difference in missing rates = 0.078).

This page has been left blank for double-sided copying.

Appendix C

Survey Sample Definition, Coding, and Analysis

This page has been left blank for double-sided copying.

Appendix C of this report on the Response to Intervention (RtI) evaluation supplements Chapters 3, 4, and 5. Appendix Figure C.1 describes the analysis sample used in each chapter to either study RtI practices across schools (Chapter 3), describe services for reading groups at different skill levels (Chapter 4), or analyze the impacts of assignment to intervention on students' reading achievement (Chapter 5). The next sections describe how samples were defined, the creation and coding of variables, and the analytic models used in Chapters 3 and 4.

The Response to Intervention (RtI) Evaluation

Appendix Figure C.1

Size of Impact Study Sample Throughout the RtI Report

Analysis Component	Number of Schools	Student Sample
Admitted to Study	146 Schools	Students receiving services in Tiers 1, 2, or 3
School-Level Analysis *(Chapter 3)*	145 Schools	Students receiving services in Tiers 1, 2, or 3
Group-Level Analysis *(Chapter 4)*	131 Grade 1 respondent schools 126 Grade 2 respondent schools 124 Grade 3 respondent schools	Students receiving services in Tiers 1, 2, or 3, with Tiers 2 and 3 combined
Student-Level Analysis *(Chapter 5)*	119 Grade 1 respondent schools 127 Grade 2 respondent schools 112 Grade 3 respondent schools	Students assigned to services in Tiers 1, 2, or 3 who fall above (Tier 1) or below (Tier 2 or 3) the cut point for assignment

NOTE: Tiers 2 and 3 are combined in the group-level and student-level analyses to align with the research questions and describe the maximum contrast in services. All Tier 3 and all Tier 2 groups are used in the group-level analysis. In the student-level analysis, most students who fell just below the cut point were placed in Tier 2, but some Tier 3 students were included in the analysis because they fell close to the cut point.

Survey Sample Definition and Attrition

The study research team distributed surveys to all 146 impact sample schools that qualified for the impact analysis and to the 1,300 schools that were selected for the reference sample.

The *school survey* sample for Chapter 3 was defined by how many schools responded to the survey. Of the 146 schools in the impact sample, 145 (99 percent) responded. Of the 1,300 randomly sampled schools, 1,105 (85 percent) responded.

Attrition in the number of schools is an issue primarily for the teacher and interventionist surveys, analyzed in Chapter 4. The *analysis sample* was determined not only by the number of schools that fielded these surveys but also by whether teachers or interventionists responded to specific items. Because the analysis in Chapter 4 focuses on reading groups by grade and reading level for each intensity factor, if a reading group was missing data for any one of these three variables, it was excluded from the analysis. Groups were also excluded from analysis if they did not serve just *one* reading level — only At or Above, only Somewhat Below, or only Far Below grade-level students. If all the reading groups in a school were missing any of these data elements, the school was effectively missing in the analysis.

The *interventionist survey* was fielded in only 141 of the 146 schools, based on agreements with the sample schools to minimize the burden on staff. Of the 141 schools, 4 served students only in groups of mixed reading levels and, consequently, were excluded from analysis. Of the 137 unique schools remaining, the school samples to describe intervention groups consist of 131 schools for Grade 1, 126 schools for Grade 2, and 124 schools for Grade 3.

Because the interventionist survey was used to categorize whether a school served all reading levels in intervention or below-only reading levels, the interventionist survey responses determined the school sample size for the teacher survey as well. Therefore, although all 146 impact sample schools completed the *teacher survey,* the school sample to describe small-group instruction during the core reading block is based on whether the school appears in the interventionist survey sample (in order to be categorized in one of the two types of schools) *and* on whether relevant items on the teacher survey were completed. The sample for small-group instruction is consequently smaller than the interventionist survey sample by grade: 118 schools for Grade 1, 117 schools for Grade 2, and 109 schools for Grade 3.

For the *exploratory analysis* in Chapter 6, the sample declines even further. The number of sites used in the impact analysis is based on availability of tier assignment data and meeting eligible requirements discussed in Appendix B. Appendix Table C.1 describes the overlap between those sites and the number of sites that completed relevant survey items.

The Response to Intervention (RtI) Evaluation

Appendix Table C.1

Overlap Between Sites in Impact Analysis and the Number of Sites That Completed Survey Items

				Of the Impact Analysis Sites	
	Number of Sites Eligible for the Impact Analysis	Number of Schools in the Small-Group Instruction Services Sample	Number of Schools in the Intervention Services Sample	Number of Schools Used to Describe Small-Group Instruction Services	Number of Schools Used to Describe Intervention Services
Grade 1	119	118	131	98	109
Grade 2	127	117	126	105	113
Grade 3	112	109	124	87	96

SOURCES: Sample-size definitions and calculations determined by study research team.

NOTES: The breakdown by grade represents the 146 unique schools in the impact sample, as described in Chapters 2 and 5. A school with at least one respondent to the teacher survey and/or the interventionist survey who completed questions about homogeneous reading groups and grade was eligible for the analysis sample in Chapter 4.

Sample Defined by Reading-Group Level for Chapter 4 Analysis

Survey respondents described services provided to each reading group that they served. Thus, the *respondent sample* determines the final reading-group sample. Teachers were asked to describe up to six groups that they served during the core reading block. Interventionists were asked to describe the services that they provided during the most recent full week of school for up to 10 groups that they met with regularly at any time of day. According to survey instructions, intervention groups should exclude students reading At or Above grade level who receive enrichment services. Based on survey coding, all reading groups should also exclude students served in self-contained special education classes. Some teachers did not complete the small-group instruction items or did not indicate what grade they taught, which excluded them from the grade-specific analysis. As a result, the school sample for small-group instruction varies by grade.

The study's data collection team could not identify a central roster or administrative record at each school that listed the number and type of reading groups served by each teacher or interventionist. As a result, it is possible that the self-reported survey descriptions of groups do not include all groups served in a school. It is also possible that two respondents may have served the same group, resulting in some double-counting of groups across respondents. This means that the number of total groups may underestimate or overestimate the number of unique combinations of students.

Missing Data and the Imputation Process

Because it is unknown whether some groups existed in schools but were not described — and, therefore, were omitted from the survey responses — missing rates at the reading-group level should be interpreted with caution. Appendix Table C.2 presents the number and percentage of missing values pooled across school types, grades, and reading levels for small-group instruction and for intervention services analyzed in Chapter 4. Note that "Staff specialization" and "Progress monitoring" are respondent-level items in the interventionist survey that were used to describe intervention groups. The rate of missing data for these domains may appear higher because the missing rate was applied to all groups associated with that respondent. The respondent-level missing rate is less than 10 percent.

When the intervention-related variables are used as predictors in the exploratory analysis, values can be imputed. The imputation process is described in Appendixes G and H.

Appendix Table C.3 presents the number of schools that responded to each exhibit presented in Chapter 3. Appendix Table C.4 presents the number of schools and reading groups used in the analysis for each contrast exhibit presented in Chapter 4 (Tables 4.5, 4.6, 4.7, and 4.8 and Figures 4.4, 4.5, and 4.6).

Recoding of Variables

Converting Multiple-Response Items into Indicator Variables

In the three surveys, a number of questions offered multiple-response options and asked respondents to "mark all that apply." These questions were coded as 0 if no responses were chosen and as 1 if a respondent chose at least one option from the list. Then each response option was treated as a separate *binary indicator variable:* responses to one option are not mutually exclusive of responses to another option. For example, an interventionist who indicates the reading skills covered during the intervention-group sessions can select multiple skills. Each skill is treated as a separate item, but the interpretation is that the percentage of groups covering, say, phonics may overlap with the percentage of groups addressing, say, fluency. Appendix Table C.5 lists the "mark all that apply" survey questions used in the analysis.

The Response to Intervention (RtI) Evaluation

Appendix Table C.2

Missing Values of Mechanisms Analyzed in Chapter 4

	Total Groups	Number of Groups Missing This Variable	Percentage of Groups Missing (%)
Instruction	4,232		
Time (minutes per week)		71	1.7
Group size		5	0.1
Content focus		29	0.7
Intervention	3,017		
Time (minutes per week)		16	0.5
Group size		17	0.6
Content focus		0	0.0
Staff specialization		346	11.5
Progress monitoring	2,530		
Oral reading fluency		292	11.5
Curriculum embedded tests		545	21.5
Running records		507	20.0

SOURCES: Interventionist and teacher survey responses about reading groups in Grades 1, 2, and 3 serving either At or Above, Somewhat Below, or Far Below grade-level students.

NOTE: The question about reading groups that describes a respondent's staff specialization was applied to all groups served by that respondent, and then differences were analyzed at the group level.

The Response to Intervention (RtI) Evaluation

Appendix Table C.3

Number of Schools Used in Chapter 3 Analysis, by Exhibit

	Impact School Sample	Reference School Sample
Figure 3.1: Schools That Have Fully Implemented RtI Practices in Grade 1		
Number of schools eligible to respond	144	986
Reading	140	883
Math	142	937
Writing	140	926
Behavior/social skills	141	932
Table 3.3: Multiple Tiers and Average Minutes per Week		
Number of schools eligible to respond	132	770
Average minutes per week		
Instruction time allocated for all students in core	131	768
Intervention time allocated for students in Tier 2	132	770
Intervention time allocated for students in Tier 3	132	770
Figure 3.3: Types of Staff		
Number of schools eligible to respond	145	1,105
Analyzed universal screening data	145	1,092
Analyzed progress monitoring data	145	1,076
Table 3.4: Data-Based Decision-Making		
Number of schools eligible to respond	145	1,105
Data considered "very important" for whether students will reach benchmarks		
Progress monitoring measures	143	1,085
Teacher observation	143	1,094
Reading diagnostic tests	141	1,074
Standardized reading tests	143	1,080
Curriculum embedded tests	142	1,075
Used publisher's recommendations for universal screening or benchmark scores assessments	140	1,065
Data considered "always used" to inform determinations of eligibility for special education		
Universal screening or a benchmark in reading	113	818
Information for systematic monitoring of student progress	116	776
Cognitive and reading assessments	104	820
Standardized reading tests	84	626
Data from other procedures	112	823

(continued)

Appendix Table C.3 (continued)

	Impact School Sample	State School Sample
Figure 3.4: Students Identified with a Specific Learning Disability (SLD) for the State Sample and the Impact Sample		
Number of sites/states eligible to respond	145	13
Student age		
Age 6	132	13
Age 7	132	13
Age 8	131	13
Age 9	132	13
Age 10	122	12

SOURCE: School survey and IDEA Section 618 Part B child count data available at http://www2.ed.gov/programs/osepidea/618-data/state-level-data-files/index.html.

NOTE: Number of eligible sites is based on responses to the question mentioned in the exhibit or a preceding question that prompted a skip pattern.

The Response to Intervention (RtI) Evaluation

Appendix Table C.4

Number of Schools and Reading Groups Used in Chapter 4 Analysis, by Exhibit

	School Sample Size	Group Sample Size: Above Grade Level	Group Sample Size: Below Grade Level
Table 4.7 - Staff Specialization			
Interventionist survey[a]			
All-Level schools			
Grade 1	57	207	530
Grade 2	45	131	359
Grade 3	39	106	329
Below-Only schools			
Grade 1	70	NA	354
Grade 2	78	NA	336
Grade 3	82	NA	309
Figure 4.5 and Table 4.5 - Group Size			
Interventionist survey[b]			
All-Level schools			
Grade 1	59	219	610
Grade 2	45	139	411
Grade 3	39	112	364
Below-Only schools			
Grade 1	72	NA	401
Grade 2	81	NA	382
Grade 3	85	NA	352
Teacher survey[c]			
All-Level schools			
Grade 1	51	379	246
Grade 2	41	266	227
Grade 3	36	205	212
Below-Only schools			
Grade 1	67	520	341
Grade 2	76	471	335
Grade 3	73	438	333
Figure 4.6 and Table 4.6 - Minutes per Week			
Interventionist survey			
All-Level schools			
Grade 1	59	218	606
Grade 2	45	140	413
Grade 3	39	114	365

(continued)

Appendix Table C.4 (continued)

	School Sample Size	Group Sample Size Above Grade Level	Group Sample Size Below Grade Level
Interventionist survey			
Below-Only schools			
Grade 1	72	NA	401
Grade 2	81	NA	381
Grade 3	85	NA	353
Teacher survey			
All-Level schools			
Grade 1	51	375	246
Grade 2	41	258	223
Grade 3	36	205	210
Below-Only schools			
Grade 1	67	513	339
Grade 2	76	467	328
Grade 3	72	429	327
Figure 4.7 - Reading Skills			
Interventionist survey			
All-Level schools			
Grade 1	59	222	613
Grade 2	45	140	413
Grade 3	39	115	367
Below-Only schools			
Grade 1	72	NA	401
Grade 2	81	NA	382
Grade 3	85	NA	354
Teacher survey			
All-Level schools			
Grade 1	51	379	246
Grade 2	41	266	227
Grade 3	36	206	209
Below-Only schools			
Grade 1	67	517	340
Grade 2	76	465	331
Grade 3	73	433	332

(continued)

Appendix Table C.4 (continued)

	School Sample Size	Group Sample Size	
		Somewhat Below Grade Level	Far Below Grade Level
Table 4.8 - Frequency of Progress Monitoring - Interventionist survey			
Oral reading fluency			
All-Level schools			
Grade 1	57	309	219
Grade 2	45	233	141
Grade 3	39	221	114
Oral reading fluency			
Below-Only schools			
Grade 1	68	189	167
Grade 2	78	193	143
Grade 3	84	201	108
Curriculum embedded tests			
All-Level schools			
Grade 1	56	285	183
Grade 2	43	219	127
Grade 3	39	198	107
Below-Only schools			
Grade 1	66	164	150
Grade 2	70	161	124
Grade 3	76	167	100
Running records			
All-Level schools			
Grade 1	56	283	189
Grade 2	44	221	124
Grade 3	39	196	106
Below-Only schools			
Grade 1	67	170	160
Grade 2	72	166	130
Grade 3	81	178	100

SOURCES: School, teacher, and interventionist survey responses for group-level items were used to determine sample sizes.

NOTES: [a]Each answer option was transformed into its own indicator variable (reading intervention/specialist, special educator, classroom teacher, speech language therapist, paraprofessional, ELL teacher, and other). This allowed each staff type to be analyzed as a separate outcome. The distribution of respondents across staff types is mutually exclusive and should add up to 100 percent.

[b]This is the number of students served in the intervention group.

[c]This is the number of students served in small-group instruction.

The Response to Intervention (RtI) Evaluation

Appendix Table C.5

Coding of Survey Questions Used in the Analysis with a "Mark All That Apply" Structure

"Mark All That Apply" Questions	Implication for Analysis
<u>School Survey</u>	
Who has the primary responsibility for administering the following student assessments? Which individuals in your school have the primary responsibility for analyzing data from the following student assessments? • Universal screening or benchmark reading tests • Curriculum embedded reading tests • Student progress monitoring in reading • State accountability tests in reading • Diagnostic tests to pinpoint specific problems	Each staff type who administers or analyzes universal screening or progress monitoring data is treated as a separate dependent variable. As a result, staff types are not mutually exclusive, because multiple staff types could analyze the same type of data.
<u>Teacher Survey</u>	
What is the content focus of this instruction? • Phonics • Fluency • Reading comprehension • Vocabulary	Each reading skill response option is analyzed as its own dependent variable. As a result, reading skills are not mutually exclusive for a given group, because groups could address multiple skills during the most recent week.
<u>Interventionist Survey</u>	
What components of reading were emphasized for this group? • Phonemic awareness • Phonics • Vocabulary • Fluency • Reading comprehension	Each reading skill response option is analyzed as its own dependent variable. As a result, reading skills are not mutually exclusive for a given group, because groups could address multiple skills during the most recent week.

SOURCES: School, teacher, and interventionist surveys.

Skip Patterns

A *skip pattern* is a series of questions that begins with a filter question; a respondent's answer to the filter question determines whether or not subsequent questions in the skip pattern should be answered. Respondents to all three surveys violated the skip patterns in two ways:

1. Respondents answered the filter question in a way that should have led them to skip subsequent questions, but they answered the items anyway. In this case, the study research team recoded the filter question to match the subsequent response pattern. This ensured that item response rates were sensible — that response rates to the filter question were greater than or equal to the response rates for the follow-up questions.

2. Respondents left the filter question blank or missing, but they completed subsequent questions in the skip pattern. In this case, the study research team recoded the filter question to match the subsequent response pattern.

Appendix Figure C.2 illustrates examples of survey skip patterns and recoding. The "Special case?" label refers to a situation in the teacher survey where a skip pattern spanned multiple pages or a filter question had two follow-up questions that also had skip patterns (a skip within a skip). These situations may have made it difficult for a respondent to follow the skip pattern correctly.

For the school survey and the teacher survey, Appendix Table C.6 lists the skip questions included in the analysis and gives the frequency of recoding each question for any of the reasons listed above.

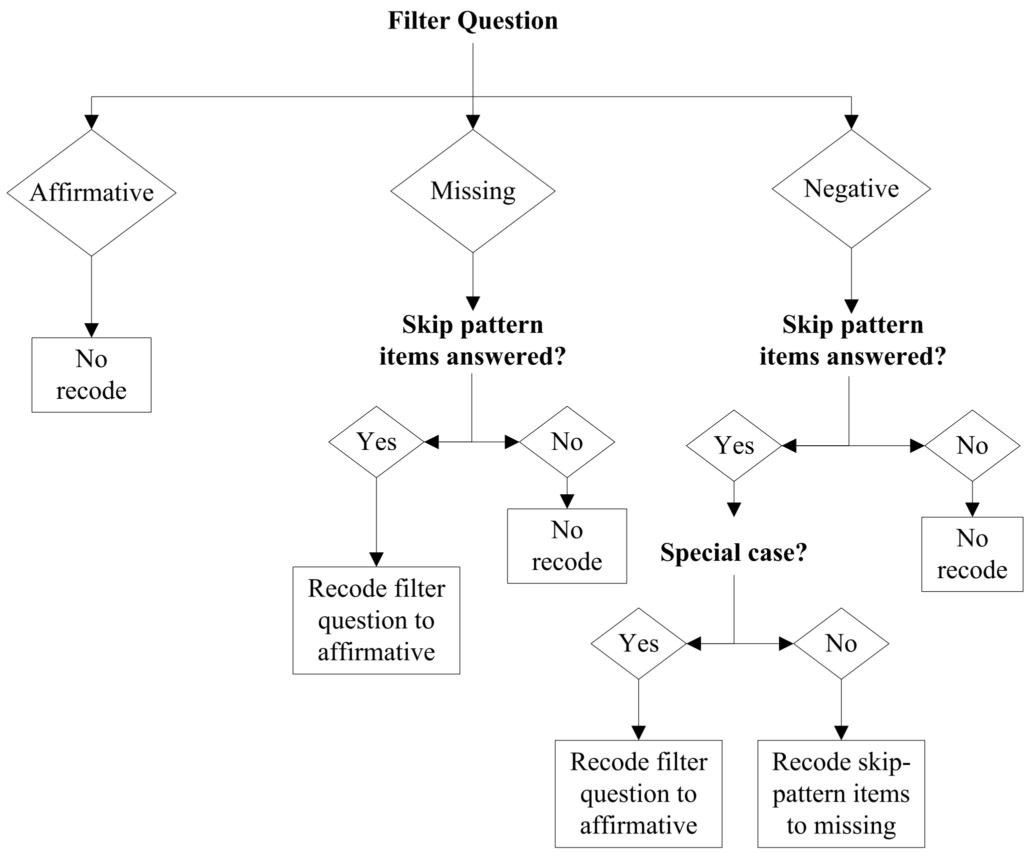
No skip questions on the interventionist survey needed to be recoded.

Missing values are counted only once, and item response rates were not decreased by schools skipping a question that they had been instructed to skip; nor were they additionally penalized for leaving all parts of a skip pattern blank. Item response rates for variables used in the school survey analysis are well above 80 percent, and schools that responded to the survey generally completed key items in the school survey. Item response rates for the teacher and interventionist surveys are reported above.

The Response to Intervention (RtI) Evaluation

Appendix Figure C.2

Recoding of Skip Patterns in Surveys



SOURCES: Teacher and interventionist surveys.

NOTE: The "special case" refers to a situation in the teacher survey where a skip pattern spanned multiple pages of the instrument or a filter question had two follow-up questions that also had skip patterns (a skip within a skip). These situations may have made it difficult for a respondent to follow the skip pattern correctly.

The Response to Intervention (RtI) Evaluation

Appendix Table C.6

Skip Questions Recoded and Used in the Analysis

Data Source and Question	Filter Questions and Condition (“only answered if…”)	Percentage of Filter Questions Recoded
School Survey		
Number of days per week allocated to reading instruction	Respondent answered “no” to the question of whether five days a week are allocated to reading instruction.	0.03
Schools’ first steps when a student scores Somewhat Below grade level in reading; schools’ first steps when student is not progressing in intervention	Respondent answered “yes” to following a prescribed sequence of steps for responding to students below grade level in reading.	2.2
Implementation of RtI in which grades for reading; implementation of RtI in math, writing, and behavior/social skills	Respondent answered “yes” to at least one grade partially or fully implementing RtI.	5.5
Teacher Survey		
Number of students in reading group	Respondent answered “yes” to providing teacher-directed reading instruction to small groups of students.	3.3
Number of minutes of reading instruction from teacher	Respondent answered “yes” to providing teacher-directed reading instruction to small groups of students.	3.3
Number of days per week of reading instruction from teacher	Respondent answered “yes” to providing teacher-directed reading instruction to small groups of students.	3.3
Type of reader in group (At or Above, Somewhat Below, Far Below)	Respondent answered “yes” to providing teacher-directed reading instruction to small groups of students.	3.3
Content focus of instruction	Respondent answered “yes” to providing teacher-directed reading instruction to small groups of students.	3.3
Additional period of time for all students to receive intervention in reading	Respondent answered “yes.”	1.3

SOURCES: School and teacher surveys.

NOTE: No skip questions on the interventionist survey needed to be recoded.

Analysis

School Survey Analysis

For binary outcomes from the school survey, a logistic regression was used to estimate whether the probability of the impact sample reporting a particular characteristic or practice differed significantly from the probability for the reference sample. The regression models do not include covariates or state indicators.

$$\Pr(Y = 1|T) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 T_j \ + \ e_j)}$$

For the reference sample, the study used sampling weights in the analysis to account for the fact that the universe of schools from which reference sample schools were selected was larger in some states than in others. For example, sampled schools in Wyoming received a lower weight than schools in California because California had more eligible elementary schools from which the sample was drawn. The weight, *W*, was constructed for each state, *s*, as follows:

$$W_s = \frac{\#\ of\ magnet, charter, or\ other\ public\ schools\ serving\ grades\ 1-3}{\#\ of\ schools\ sampled\ that\ serve\ grades\ 1-3}$$

The denominator is the same in all cases because 100 schools were sampled from each state. Appendix Table C.7 summarizes the values for each state of the sampling weight applied to reference sample schools. The weights correspond to the size of the state so that schools in the analysis represent the number of eligible schools in the state. This information is found in Appendix Table C.7.

As a check on the creation of sampling weights, the study research team calculated means for the Common Core of Data (CCD) characteristics among all qualifying schools in the 13 states and then calculated means among only the reference sample, with weights and without weights. Weighted versions more closely match the cross-state average, reflecting a correct creation of weights.

No sampling weights were used for the impact sample schools because those schools were not intended to be representative of their states.

School Survey Analysis, by State

The analysis for the impact sample schools pools all impact sample schools within a grade together, because eligible schools all met the same screening criteria. To preserve this

The Response to Intervention (RtI) Evaluation

Appendix Table C.7

Sampling Weights for Reference Sample Schools, by State

State	Number of Public Schools Serving Grades 1-3 (excludes impact schools)	Sampling Weight
Arizona	1,175	11.8
California	5,599	56.0
Florida	2,032	20.3
Illinois	2,121	21.2
Massachusetts	917	9.2
Minnesota	840	8.4
Missouri	1,045	10.5
Montana	332	3.3
Pennsylvania	1,688	16.9
Texas	4,113	41.1
Utah	550	5.5
Washington	1,067	10.7
Wyoming	181	1.8

SOURCE: The sampling weight for each state was determined by the number of schools responding out of the 100 schools approached in each state.

idea throughout the report, the analysis in Chapter 3 that compares the two samples treats the impact sample schools as one group and the reference sample schools as another group. That analysis does not use state indicators (fixed effects) or compare schools only with their counterparts in the same state.

Appendix Table C.8 presents a summary of key RtI practices, by state. It shows that implementation of certain practices varied significantly across states.

The Response to Intervention (RtI) Evaluation

Appendix Table C.8

Summary of Key RtI Practices: Percentage of Reference Schools Reporting Implementation of Each Practice, by State

	State													
School Administrators' Responses	AZ	CA	FL	IL	MA	MN	MO	MT	PA	TX	UT	WA	WY	P-Value
Summary														
Average number of 8 key practices in place (a - h)	5.3	4.6	6.9	5.7	5.1	5.2	5.3	5.4	5.5	5.7	6.0	5.0	6.0	0.000
Multiple Tiers														
a. Provided more than 90 minutes in daily core reading block (Grades 1-3)	62.2	65.4	96.4	66.7	54.8	67.9	63.5	59.8	78.2	67.0	70.7	47.1	59.1	0.000
b. Provided Tier 2 intervention at least 3 times a week	73.3	71.6	96.4	90.0	77.4	72.8	80.0	79.3	70.1	85.2	89.3	76.5	84.1	0.000
c. Provided Tier 3 intervention at least 5 times a week	23.3	28.4	65.5	64.4	47.6	45.7	56.5	69.0	54.0	51.1	46.7	51.8	65.9	0.000
Staffing and Resources														
d. Allocated staff to assist teachers with using data	80.0	61.7	91.7	62.2	63.1	77.8	74.1	67.8	78.2	77.3	81.3	64.7	79.5	0.000
e. Allocated staff to assist teachers with teaching reading	67.8	38.3	85.7	53.3	56.0	45.7	60.0	42.5	49.4	63.6	70.7	45.9	89.8	0.000
Data-Based Decision Making														
f. Conducted universal screening at least 2 times a year in Grades 1 and 3	70.0	53.1	66.7	67.8	56.0	66.7	61.2	70.1	62.1	45.5	74.7	72.9	65.9	0.000
g. Followed a prescribed sequence of steps for responding to students who are below benchmark	82.2	82.7	98.8	87.8	84.5	84.0	78.8	88.5	86.2	97.7	86.7	77.6	90.9	0.000
h. Used data to progress-monitor students following implementation of reading interventions, as part of determining special education eligibility	75.6	61.7	86.9	80.0	66.7	64.2	54.1	59.8	73.6	84.1	80.0	65.9	60.2	0.000

(continued)

Appendix Table C.8 (continued)

SOURCE: School survey; wording of items is listed below.

NOTES: For all items, the number of responding schools is 1,105 for the reference sample. The percentages reported in this table reflect the number of schools out of the total that affirmatively responded that they implemented a practice. For the purposes of summing observations, missing or skipped practices are interpreted as zero values. As a result, the means reported in this table may differ from those reported for individual corresponding items in other tables, because schools that did not answer certain items have a value of zero for each practice, but are excluded from other tables. Sampling weights were applied to each reference school that responded to the survey.

P-values for the average number of practices in place were estimated from a linear regression with number as the outcome.

P-values by feature were based on a chi-squared test over all the states.

The eight practices are defined as follows:

a: "How many total minutes during the school day are allocated to the core reading block (for example, phonemic awareness, phonics, fluency, vocabulary, and reading comprehension, but excluding spelling, grammar, and writing) for students in Grades K-5?" The average of the high end of the time ranges must have been greater than 90 in each of Grades 1, 2, and 3.

b: "How many days per week do most students receive Tier 2 intervention(s)?" Response must have been 3, 4, or 5.

c: "How many days per week do most students receive Tier 3 intervention(s)?" Response must have been 5.

d: "In the 2011-12 school year, is there someone in the building whose role is to assist teachers in using and interpreting assessment data on reading?" Response must have been "yes."

e: "Is there someone in the school who provides coaching to classroom teachers on teaching reading?" Response must have been "yes."

f: "In what months are screening or benchmark measures administered to students in each grade?" Response must have indicated at least two nonconsecutive months of assessment in each of Grades 1 and 3.

g: "Does your school follow a prescribed sequence of steps for responding to students who are below benchmark in reading?" Response must have been "yes."

h: "In your school, which of the following kinds of data are used for informing special education eligibility determinations for students suspected of having a specific learning disability?" For "data and other information from systematic monitoring of student progress following implementation of reading interventions," response must have been "always used."

Analysis of Reading Groups in the Teacher and Interventionist Surveys

For the comparison of reading group services in Chapter 4, the unit of analysis is the adult-led small reading group for intervention and for small-group instruction during the core reading block. Each of five mechanisms is analyzed separately as an outcome of interest. For each one, the analysis estimates the average difference between reading groups serving students reading at or above grade level and those serving students reading below grade level within each school. The study research team estimated this difference separately for each grade. The analysis uses a regression model that accounts for the fact that reading groups are clustered, or nested, within their schools. As shown in the equation below, for reading group i in school j, each of the intensity factors served as a dependent variable. (A similar model was used for the other mechanisms as well.) A binary indicator of reading-group level (at or above or below grade level) is the independent variable *ReadingLevel*. School fixed effects, SCH_j, were included to estimate the within-school average difference between reading groups at or above grade level or below it. The fixed effects also account for the different number of groups in each school that were at each level. Regressions did not adjust for covariates. The error term u_j indicates variation between schools, and the error term e_{ij} reflects the variation within schools between groups.

$$Intensity_{ij} = \beta_1 ReadingLevel_{ij} + SCH_j + u_j + e_{ij}$$

Exhibits in Chapter 4 report averages for the two types of reading groups: the averages for groups serving students at or above grade level are the observed group means, while the averages for groups serving students below grade level reflect the observed group means for above-grade-level groups subtracted from the estimated difference between reading levels (accounting for nested groups within schools).

The study research team explored several approaches to present mean differences between reading levels: simple means across all groups within a reading level, a random-effects model with nesting of groups within schools, and a fixed-effect school intercept model. All produced similar estimated mean differences. The study research team also ran a model that pooled all grades, with grade indicators to account for any between-grade differences. The direction of the estimated differences between reading levels is similar in the all-grade and the grade-specific models.

Appendix Tables C.9, C.10, and C.11 present results for all three grades for the mechanisms to vary support services: specialization of interventionist staff, frequency of progress monitoring, and reading skills addressed in the group session.

The Response to Intervention (RtI) Evaluation

Appendix Table C.9

Service Contrast for Interventionist Staff Specialization: Difference Between Groups At or Above Grade Level and Below Grade Level in Intervention Groups for All-Level Schools, by Grade

	Groups At or Above Grade Level	Groups Below Grade Level	Mean Difference Between Groups	P-Value
Grade 1				
Percentage of groups as served by staff type				
Paraprofessional	30.7	37.1	6.4	0.078
Classroom teacher	41.6	25.7	-15.9	0.000
Reading specialist	5.8	14.1	8.3	0.001
Special educator	4.3	7.3	2.9	0.121
English Language Learner teacher	6.5	5.9	-0.6	0.695
Speech pathologist	0.0	0.7	0.7	0.314
Other	11.1	9.3	-1.8	0.490
Grade 2				
Percentage of groups as served by staff type				
Paraprofessional	20.2	27.5	7.2	0.094
Classroom teacher	53.8	32.5	-21.2	0.000
Reading specialist	6.8	13.3	6.5	0.009
Special educator	6.8	9.8	3.0	0.221
English Language Learner teacher	2.3	4.0	1.7	0.293
Speech pathologist	0.0	0.5	0.5	0.580
Other	10.1	12.5	2.4	0.411
Grade 3				
Percentage of groups as served by staff type				
Paraprofessional	21.7	22.3	0.5	0.908
Classroom teacher	54.9	41.3	-13.6	0.011
Reading specialist	6.9	14.3	7.4	0.007
Special educator	2.0	8.4	6.5	0.020
English Language Learner teacher	1.5	3.5	2.0	0.215
Speech pathologist	0.0	1.0	1.0	0.323
Other	13.1	9.2	-3.8	0.259

SOURCE: Interventionist survey.

NOTES: Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving

(continued)

Appendix Table C.9 (continued)

intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services.

The numbers of All-Level schools are 57 for Grade 1, 45 for Grade 2, and 39 for Grade 3. The numbers of Below-Only schools are 70 for Grade 1, 78 for Grade 2, and 82 for Grade 3. The percentages of below-grade-level groups served by teachers and/or paraprofessionals in the Below-Only schools are 35.4 percent in Grade 1, 37.9 percent in Grade 2, and 32.4 percent in Grade 3.

The findings about service contrast in Chapter 4 are based on data by reading-group level, which are not linked to data by individual-student level. Thus, the analysis could not restrict comparisons to reading groups that served students near the cut score used for the Regression Discontinuity (RD) design; those students formed the basis of the impact analysis sample (described in more detail in subsequent appendixes). In a randomized controlled trial (RCT), one would not need to restrict the comparisons, because generally all students are included in the estimated average service contrast as well as the average impact.

Appendix Table C.10

Average Frequency of Progress Monitoring for Intervention Groups Below Grade Level, by Grade

	Below-Only Schools						All-Level Schools					
		Of Schools Using Each Test, the Frequency of Monitoring per Month						Of Schools Using Each Test, the Frequency of Monitoring per Month				
	Schools Using This Test (%)	Somewhat and Far Below Grade Level	Somewhat Below Grade Level	Far Below Grade Level	Mean Difference Between Groups	P-Value	Schools Using This Test (%)	Somewhat and Far Below Grade Level	Somewhat Below Grade Level	Far Below Grade Level	Mean Difference Between Groups	P-Value
<u>Number of Times per Month for Each Type of Test</u>												
Grade 1												
Oral reading fluency	93	3.3	3.0	3.5	0.6	0.002	95	3.3	3.4	3.4	0.0	0.931
Curriculum embedded tests	63	1.2	1.2	1.3	0.1	0.559	80	1.8	1.9	2.1	0.2	0.161
Running records	78	2.3	2.1	2.8	0.7	0.002	83	2.6	2.5	2.6	0.1	0.506
Grade 2												
Oral reading fluency	95	3.4	3.3	3.5	0.2	0.360	100	3.2	3.2	3.2	0.0	0.929
Curriculum embedded tests	64	1.4	1.3	1.5	0.2	0.245	80	1.4	1.3	1.9	0.6	0.002
Running records	70	2.2	2.2	2.2	0.0	0.891	89	2.7	2.5	3.0	0.5	0.071
Grade 3												
Oral reading fluency	99	3.5	3.3	3.8	0.5	0.030	100	3.4	3.2	3.6	0.5	0.026
Curriculum embedded tests	65	1.7	1.7	2.2	0.4	0.095	90	1.8	1.8	2.0	0.1	0.607
Running records	73	2.3	2.3	2.7	0.3	0.265	87	2.2	2.3	2.2	-0.2	0.546

(continued)

Appendix Table C.10 (continued)

SOURCES: Interventionist survey responses about reading groups in Grades 1-3 concerning Somewhat Below grade-level groups and Far Below grade-level groups.

NOTES: The means presented for Far Below groups were regression-adjusted, and school fixed effects were used to account for clustering of groups within schools. The means presented for Somewhat Below groups are simple means.

The number of schools for Grade 1 Below-Only schools is 72 and for Grade 1 All-Level schools is 59. The number of schools for Grade 2 Below-Only schools is 81 and for Grade 2 All-Level schools is 45. The number of schools for Grade 3 Below-Only schools is 85 and for Grade 3 All-Level schools is 39.

The Response to Intervention (RtI) Evaluation

Appendix Table C.11

Service Contrast for Reading Skills: Difference Between Groups At or Above Grade Level and Below Grade Level in Below-Only and All-Level Schools, by Reading Skill Targeted and Grade

	Below-Only Schools				All-Level Schools			
Reading Skill	Groups At or Above Grade Level (%)	Groups Below Grade Level (%)	Mean Difference Between Groups	P-Value	Groups At or Above Grade Level (%)	Groups Below Grade Level (%)	Mean Difference Between Groups	P-Value
Grade 1								
Small-Group Instruction								
Phonemic awareness	25	74	49	0.000	35	76	41	0.000
Phonics	46	92	46	0.000	52	90	38	0.000
Vocabulary	85	77	-8	0.002	89	81	-8	0.001
Reading comprehension	98	86	-12	0.000	99	91	-8	0.000
Fluency	89	85	-5	0.064	93	89	-5	0.047
Intervention								
Phonemic awareness	NA	77	NA	NA	49	69	19	0.000
Phonics	NA	89	NA	NA	66	85	18	0.000
Vocabulary	NA	70	NA	NA	86	75	-11	0.001
Reading comprehension	NA	72	NA	NA	86	78	-8	0.010
Fluency	NA	88	NA	NA	87	81	-6	0.028
Grade 2								
Small-Group Instruction								
Phonemic awareness	18	60	42	0.000	19	57	38	0.000
Phonics	24	82	58	0.000	30	83	53	0.000
Vocabulary	79	76	-4	0.188	71	69	-2	0.519
Reading comprehension	95	88	-7	0.000	96	86	-10	0.000
Fluency	68	87	20	0.000	78	90	12	0.001
Intervention								
Phonemic awareness	NA	55	NA	NA	19	41	22	0.000
Phonics	NA	79	NA	NA	46	75	30	0.000

(continued)

Appendix Table C.11 (continued)

Reading Skill	Below-Only Schools: Groups At or Above Grade Level (%)	Below-Only Schools: Groups Below Grade Level (%)	Below-Only Schools: Mean Difference Between Groups	Below-Only Schools: P-Value	All-Level Schools: Groups At or Above Grade Level (%)	All-Level Schools: Groups Below Grade Level (%)	All-Level Schools: Mean Difference Between Groups	All-Level Schools: P-Value
Intervention (cont.)								
Vocabulary	NA	72	NA	NA	81	71	-9	0.022
Reading comprehension	NA	81	NA	NA	86	81	-5	0.152
Fluency	NA	89	NA	NA	79	85	6	0.080
Grade 3								
Small-Group Instruction								
Phonemic awareness	9	34	25	0.000	17	30	13	0.000
Phonics	16	62	47	0.000	19	61	42	0.000
Vocabulary	80	84	4	0.141	84	86	2	0.511
Reading comprehension	99	94	-5	0.003	97	97	1	0.724
Fluency	57	88	31	0.000	62	89	27	0.000
Intervention								
Phonemic awareness	NA	35	NA	NA	22	32	10	0.039
Phonics	NA	63	NA	NA	26	53	28	0.000
Vocabulary	NA	86	NA	NA	80	76	-4	0.357
Reading comprehension	NA	89	NA	NA	93	86	-7	0.083
Fluency	NA	75	NA	NA	75	83	8	0.046

SOURCES: Teacher and interventionist surveys.

NOTES: "Small-group instruction" refers to services provided by teachers during the core reading block to all students. Intervention services are provided by either teachers or interventionists to students needing targeted reading support, either during or outside the core reading block. The Below-Only school sample represents schools that had at least one of either a Somewhat Below or a Far Below grade-level group receiving intervention services. The All-Level school sample represents schools that had at least one At or Above grade-level group receiving intervention services and at least one of either a Somewhat Below or a Far Below grade-level group (a below-grade-level group) receiving intervention services. No tests were performed between intervention groups in Below-Only schools, which do not provide intervention to At or Above grade-level groups. The survey item asking about content focus in the teacher survey and in the interventionist survey were both "mark all that apply." As a result, the percentage of groups indicating one content area could have overlapped with another content area.

See sample sizes in Appendix Table C.4.

This page has been left blank for double-sided copying.

Appendix D

Schools' Decision Rules and Data Used for the Impact Analysis

This page has been left blank for double-sided copying.

Appendix D of this report on the Response to Intervention (RtI) evaluation supplements the analyses discussed in Chapter 5 by going into greater detail about (1) the universal screening tests used in the fall to assess student reading level, (2) the decision rules used by schools to assign tier placements, (3) the subsequent construction of the student rating variable, (4) a description of demographic variables used in the analytic sample, and (5) the difference in attrition rates for key variables between treatment and comparison students.

Assessments Used in Fall Screening

Appendix Table D.1 summarizes all the assessments used in the determination of fall tier placements by schools in the impact analysis sample. The table note lists all specific assessments in each category. Among them, the DIBELS Next Nonsense Word Fluency – Correct Letter Sounds test, used in 30 schools, was the most-used screening for the Grade 1 sample.[8] For Grade 2, the AIMSweb Reading Curriculum-Based Measurement was most widely used in the sample.[9] For Grade 3, the DIBELS Next Oral Fluency (Words Correct) was the most popular screening, followed closely by the Measures of Academic Progress.

Decision Rules Used by Schools for Tier Assignment

As described in Appendix B, during the recruitment process, the study research team conducted several rounds of screening (by paper, over the phone, and in person) to ensure that the RtI schools in the study sample had clear and quantifiable decision rules in place to determine students' tier placement. The study research team verified (1) that a school used fall screening test score(s) to determine students' placement; (2) that screening tests were scored systematically (for example, using score systems like DIBELS or AIMSweb); and (3) that cut points were largely determined independently (for example, cutoff scores offered by the testing system, cutoff scores set by a districtwide standard, or cutoff scores determined by available staff resources at a school or district).

For the majority of the schools in the impact sample, the decision rules and the cut points were provided to the study research team during the recruitment process. They were usually found in statements like "If a first-grade student's DIBELS Next NWF-CLS score is at or below 23, then he or she is in Tier 2 at least" or "If a second-grade student scored below the cut point in at least one of the three subtests of DIBELS Next test – LNF, PSF, NWF-CLS, then he or she is in Tier 2 at least." However, there were some schools in the sample where the cut point was not clear, based on screening responses. For those schools, the study research team

[8]Website: https://dibels.uoregon.edu/.
[9]Website: http://www.aimsweb.com/.

Appendix Table D.1

List of Assessments Used in Single- and Multiple-Screening-Test Schools, by Grade

Fall Screening Test	Number of Schools	Single Screening Tests	Multiple Screening Tests
Grade 1			
AIMS system screening tests	29	18	11
DIBELS Next system screening tests	54	43	11
Other screening tests	38	33	5
Total	119	94	25
Grade 2			
AIMS system screening tests	38	24	14
DIBELS Next system screening tests	47	36	11
Other screening tests	52	35	17
Total	127	95	32
Grade 3			
AIMS system screening tests	27	16	11
DIBELS Next system screening tests	42	33	9
Other screening tests	50	40	10
Total	112	89	23

SOURCE: School-reported fall screening data.

NOTES: The data report on schools in the analytic sample. Totals for the columns on multiple screening tests and number of schools count schools only once regardless of the number of screening tests used.

Tests in this category include the following:

AIMSweb (AIMS): Letter Naming Fluency (LNF) - Identification of upper- and lower-case letters; Letter Sound Fluency (LSF) - Verbalization of letter sounds; Maze (MAZE) - Comprehension of passages read silently; Nonsense Word Fluency (NWF) - Identification of nonsense words; Phoneme Segmentation Fluency (PSF) - Identification of spoken phonemes; Reading Curriculum-Based Measurement (R-CBM) - Correctly reading passages aloud; Word Identification Fluency (WIF) - Timed word identification. Nonstandard AIMSweb benchmark. Its use was confined to a single district.

DIBELS Next (DBN): Composite Score (COMP) - A single summary reading ability score constructed from scores on multiple different DIBELS Next subtests. The component subtests vary in type and weighting by grade, school year, and time of year; Daze (DAZE) - Comprehension of silently

(continued)

Appendix Table D.1 (continued)

read passages as measure through the completion of multiple-choice cloze questions; Oral Reading Fluency – Words Correct (DORF) - Speed of reading words aloud; Oral Reading Fluency (DORF Accuracy) - Accuracy of reading words aloud; Oral Reading Fluency (DORF Retell) - Summarization of recently read passages; Letter-Naming Fluency (LNF) - Speed of identification of upper- and lower-case letters identified in one minute; Nonsense Word Fluency – Correct Letter Sounds (NWF CLS) - Speed of reading letter sounds of nonsense-words aloud; Nonsense Word Fluency – Whole Words Read (NWF WWR) - Speed of reading nonsense-words aloud; Phoneme Segmentation Fluency (PSF) - Identification of spoken phonemes.

Other Benchmarks: Basic Phonics Skills Test (BPST) - Sounding out letter sounds and word recognition; Basic Reading Inventory – Oral Reading Fluency (BRI ORF) - Miscues in reading passages aloud; Benchmark Reading Level Assessment (BRLA) - Accuracy, comprehension, and fluency. Assessment system based of Rigby Literacy Levels in which students read from specialized books; Curriculum Based Measurement (CBM) - Assessment used for monitoring progress in reading in accordance with a curriculum. Unspecified assessment system; California Standards Test (CST) - Composite score from tests of word analysis, reading comprehension, literary response and analysis, writing strategies, and written conventions, used by the State of California; DIBELS Oral Reading Fluency – Words Correct (DBL ORF) - Speed of reading words aloud; Letter-Naming Fluency (DBL LNF) - Speed of identification of upper- and lower-case letters identified in one minute; Nonsense Word Fluency – Correct Letter Sounds (DBL NWF CLS) - Speed of reading letter sounds of nonsense-words aloud; Dolch Sight Words (DOLCH) - Identification of common but nonphonetic words; Developmental Reading Assessment, 2nd Edition (DRA2) - A single score composed of measures of oral reading fluency and comprehension; easyCBM Passage Reading Fluency (EASY PRF) - Speed of reading words aloud; easyCBM Reading Comprehension (EASY RCOMP) - Reading comprehension; Fountas & Pinnell Reading Level (F&P) - A reading level selected by an instructor based on student oral reading comprehension and fluency; Fountas & Pinnell Reading Comprehension (F&P COMP) - Demonstrating understanding of a read passage through conversation with an instructor; Fountas & Pinnell Oral Reading Fluency (F&P ORF) - Correctly reading passages aloud; Florida Assessments for Instruction in Reading – Oral Reading Fluency (FAIR ORF) - Speed of reading a passage aloud; Florida Assessments for Instruction in Reading – Reading Comprehension (FAIR RCOMP) - Reading comprehension; Florida Assessments for Instruction in Reading – Vocabulary (FAIR VOCAB) - Identification of pictures using vocabulary words; Florida Assessments for Instruction in Reading – Word Reading (FAIR WR) - Speed of reading words from a list aloud; High Frequency Words (HFW) - Identification of a list of common words. Unspecified assessment system; IDEL Fluidez en Lectura Oral (IDEL SPFLO) - Speed of reading Spanish words aloud; IDEL Fluidez en Nombrar Letras (IDEL SPFNL) - Identification of upper- and lower-case letters in Spanish; Istation Indicators of Progress Early Reading (IS ISIP) - Phonemic Awareness, Alphabetic Knowledge, Vocabulary, Comprehension, and Fluency; iSTEEP – Oral Reading Fluency (ISTEEP ORF) - Speed of reading words aloud; Measures of Academic Progress (MAP) - Print awareness, vocabulary, phonics skills, reading comprehension, and literary concepts as measured by a computerized adaptive assessment; Place-Specific Assessments - Various benchmark assessments unique to specific districts or schools; Rigby Running Records (RIGBY) - Reading accuracy and sophistication of student reading strategies; Reading/Language Arts - Accuracy (RLA ACCURACY) - Accuracy of words read aloud. Unspecified assessment system; Reading/Language Arts - Words per Minute (RLA WPM) - Speed of reading words aloud. Unspecified assessment system; Scholastic Reading Inventory (SRI) - Reading comprehension of short passages; STAR Reading Enterprise Assessment (STAR) - Phonics, fluency, vocabulary measured by computer-adaptive assessment. Not related to California's STAR (Standardized Testing and Reporting) assessment.

conducted analyses to determine a cut point value that maximized the compliance rate both below and above the selected value.[10]

The Rating Variable

Some schools in the impact sample used multiple screening tests to determine students' intended tier placements while other schools used only one test. Appendix Table D.2 shows that the schools that used multiple screening tests to determine students' rating and tier placements account for about a quarter or less of the analytic sample for all three grades.

To be able to pool data across schools to conduct one regression analysis for each outcome, the study research team needed to express all rating variables using the same metric. The approach used depended on whether schools used a single or multiple screening tests to determine a student's tier placements.

For schools that used a single screening test to determine a student's tier placement in the fall, the raw scores from that test were first centered on the cut point for the given school. Next, the study research team calculated the standard deviation in test scores for that certain test (pooling data across all schools that used that test)[11] and scaled the centered score by this within-test, pooled standard deviation. Specifically, the following equation was used for the standardization of the rating variable for a given school *j*:

$$S_{ij} = \frac{(R_{ij} - C_j)}{\sigma_R} \qquad (1)$$

where:

S_{ij} = the standardized rating value for student *i* at school *j*

R_{ij} = the raw score for student *i* at school *j*

C_j = the cut point value at school *j*

σ_R = the standard deviation of raw test scores across all schools using the given screening test.

[10]Some schools in the sample did not provide information to clarify whether they intended to provide Tier 2 or Tier 3 intervention to students scoring exactly at the cut point as well as to students scoring below it. For these schools, if the majority of students with a rating equal to the cut score received Tier 2 or Tier 3 intervention in the fall, then all students with a rating at the cut score were classified as intended to receive Tier 2 or Tier 3 interventions, and vice versa.

[11]This is done to obtain a more reliable estimate of the standard deviation.

The Response to Intervention (RtI) Evaluation

Appendix Table D.2

Counts of Schools with Single or Multiple Screening Tests Used in Construction of Rating Variable, by Grade

Grade	Number of Schools with Single Screening Test	Number of Schools with Multiple Screening Tests	Total Number of Schools
Grade 1	94	25	119
Grade 2	95	32	127
Grade 3	89	23	112

SOURCE: School-reported fall screening data.

NOTE: The data report on schools in the analytic sample.

For most schools using multiple screening tests to determine a student's tier placement in the fall, the decision rules fell into one of the following four categories: a student was placed into at least Tier 2 if (1) the student scored lower than the cut point for at least *one* out of several screening tests; (2) the student scored lower than the cut point for at least *two* out of several screening tests; (3) the student scored lower than the cut point on *all* of several screening tests; or (4) the student's weighted score across several screening tests was below the cut point.

The construction of the rating variable for these cases involved more steps. First, scores from each of the multiple screening tests were standardized using Equation (1). Next, the students at these schools were assigned a single rating based on decision rules provided by the schools. Corresponding to the four categories described above, students were assigned a rating equal to their lowest standardized rating, their second-lowest standardized rating, their highest standardized rating, or the weighted mean of their standardized rating, respectively.[12,13]

[12]This is referred to as the "binding score" approach in the multiple-rating RD design literature. This approach is used because it allows the collapsing of multiple dimensional ratings into one dimension and, hence, allows the pooling of data across schools using different decision rules. For discussion of this method, see, for example, Reardon and Robinson (2012) and Wong, Steiner, and Cook (2013).

[13]One school used a weighted mean for the first-grade decision rule. Two schools had separate decision rules for English speakers and for Spanish speakers.

Indicators of Intended and Actual Treatment Assignment

The intended treatment assignment was determined by a student's rating value relative to the cut point set by the school. Students whose ratings were above the cut point should have been assigned to the comparison condition and, therefore, have a value of 1 for this variable, while those whose ratings were at or below the cut point should have been assigned to the treatment condition and have a value of 0 for this variable.

The actual treatment assignment was determined by a student's actual tier placement in the fall semester. Based on the fall roster data that schools provided to the study research team, students who were actually in Tier 2 or Tier 3 as their highest tier have an actual treatment status of 1, while students who were actually only in Tier 1 have an actual treatment status of 0. Note that this status is completely determined by the school and that the study research team played no role in this decision.

Even though the overall crossover and no-show rates are moderate across grades in the analysis sample, as indicated elsewhere in the report, the rates vary by schools. It is possible that other factors, such as resource constraints, teachers' judgments, or other issues, might have played some role in students' actual tier placement in the fall. Unfortunately, information on these factors is not available to the study.

Outcome Measures

The study uses four outcome measures to capture students' reading performances at different grade levels. A comprehensive reading measure was used for Grades 1 and 3 to capture students' broad set of reading skills. Specifically, the reading assessment test from the Early Childhood Longitudinal Study, Kindergarten Cohort of 2011 (ECLS-K:2011),[14] was used for Grade 1, and state reading achievement tests conducted by each state were used for Grade 3. A decoding-fluency measure — the Sight Word Efficiency subtest from the Test of Word Reading Efficiency-Second Edition (TOWRE2)[15] — was used to assess the specific decoding-fluency skill of students in Grades 1 and 2. These test scores were collected in the spring of the 2012-13 school year.

The ECLS-K assessed skills that are typically taught and developmentally important. The assessment frameworks were derived from national and state standards, including those of the National Assessment of Educational Progress (NAEP), and from the scope and sequence documents from state assessments. The ECLS-K assessments included items that were specifically created for that study, items adapted from commercial assessments with

[14]Website: https://nces.ed.gov/ecls/assessments2011.asp.
[15]Torgesen, Wagner, and Rashotte (1999).

copyright permission, and items from other NCES studies including items from NAEP (disclosed items), NELS:88, and ELS:2002.

The ECLS-K direct cognitive assessments were two-stage adaptive tests; all children began a subject area test with a routing test, which was then followed by a second-stage form. The two-stage, adaptive assessment format helped ensure that children were tested with a set of items most appropriate for their level of achievement and minimized the potential for floor and ceiling effects.

The reading assessment was designed to measure basic skills such as print familiarity, letter recognition, beginning and ending sounds, recognition of common words (sight vocabulary), and decoding multisyllabic words; vocabulary knowledge such as receptive vocabulary and vocabulary in context; and passage comprehension. The kindergarten-first grade assessment began with relatively more emphasis on basic reading skills.[16]

The ECLS-K Reading test was used at the end of the school year as an outcome measure. It was never used as a screening or progress monitoring tool for student tier placement during the school year.

For Grade 1 and Grade 2 outcomes, ECLS-K Reading Assessment and TOWRE2 scores were standardized at the grade level by using the sample mean and the sample standard deviation.[17] Grade 3 scores on state assessments were standardized by using state-level means and standard deviations provided by the states. The following equation was used for the standardization of outcome measures:

$$Y_i = \frac{(X_i - M)}{\sigma} \qquad (2)$$

where:

Y_i = The standardized score for student i

X_i = The raw test score for student i

M = The sample mean for Grade 1 or Grade 2 students or the appropriate state mean for Grade 3 students

σ = The sample standard deviation for Grade 1 or Grade 2 students or the appropriate state standard deviation for Grade 3 students.

[16]Source: http://nces.ed.gov/ecls/kinderassessments.asp.

[17]No national norming sample means and standard deviations are available for these two tests at the time of this report.

Appendix Table D.3 summarizes the outcome measures used in the analysis, by grade. It includes a brief list of key reading skills assessed, the sample means and sample standard deviations of the outcome measures, and, for the state assessments for Grade 3, the state means and standard deviations used for standardization. It also provides abbreviated links to online documentation of these tests.

Demographic Variables

Student-level demographic variables are also used in the impact analysis. MDRC requested several student-level demographic variables from study districts for the 2011-12 school year. Data-keeping practices for such information varied significantly among the states in the study, among districts within the states, and even among schools within the districts. Legends for variables were typically provided by the district, although, for a few cases, legends had to be found on state or district Department of Education websites. The following student-level demographic variables are used in the impact analysis.

- **Race/Ethnicity.** Race and ethnicity were often confounded. A handful of schools reported a separate ethnicity variable for Hispanic or reported multiple races instead of a pan-multiracial category. In the end, all race/ethnicity variables were consolidated into five mutually exclusive categories: white, non-Hispanic; black, non-Hispanic; Hispanic; Asian/Pacific Islander; and other/multiracial. The "other" category includes students with race values of "Unknown." Students with demographic data indicating Hispanic ethnicity were coded as Hispanic regardless of other, racial variables.

- **Age.** Students' ages were calculated from date of birth to August 15, 2011.

- **Sex.** The data for the sex of students were converted to a binary variable for "male."

- **Income Status.** Students were flagged as having low-income status if they received free or reduced-price lunches.

- **English Language Learner (ELL).** Students were flagged as ELL if the data indicated that they were English Language Learners, Limited English Proficient (LEP), or received English as a Second Language (ESL) services.

- **Individualized Education Program (IEP).** The IEP variable was created as a catchall for students in any disability category. No distinction was made

The Response to Intervention (RtI) Evaluation

Appendix Table D.3

Summary of Standardized Reading Outcomes

Outcome	Reading Skills Measured	Sample Mean	Sample Minimum	Sample Maximum	Sample Std. Dev.	State Mean	State Std. Dev.	Online Assessment Documentation
Grade 1								
ECLS-K Reading Assessment	Comprehensive basic reading skills	68.0	15.8	97.6	12.5			http://tinyurl.com/ne9o6ks
TOWRE2	Word reading fluency	43.2	0.00	86.0	16.0			http://tinyurl.com/cccqf59
Grade 2								
TOWRE2	Word reading fluency	57.7	0.00	90.0	12.7			http://tinyurl.com/cccqf59
Grade 3								
Arizona IMS	Comprehension	460	338	640	46.3	463	48.1	http://tinyurl.com/mcj7vt4
California CST	Comprehension, vocabulary, analysis	359	207	559	57.3	347	64.0	http://tinyurl.com/k9k6q29
Florida FCAT 2.0	Comprehension, vocabulary, analysis	193	154	229	18.3	201	22.0	http://tinyurl.com/m7zhq6s
Illinois SAT	Comprehension, vocabulary, analysis	220	126	329	26.7	210	30.3	http://tinyurl.com/o8ro64v
Massachusetts CAS	Comprehension, fluency, composition	39.2	17.5	47.5	5.69	35.9	8.46	http://tinyurl.com/c36d9ya
Minnesota CA	Comprehension, fluency, composition	367	301	399	17.6	365	19.3	http://tinyurl.com/728wp95
Missouri AP	Comprehension, composition	656	455	762	32.9	642	37.7	http://tinyurl.com/pkgzz8j
Montana CAS	Comprehension, fluency, composition	43.9	14.0	54.5	10.8	39.6	10.7	http://tinyurl.com/ptet22n
Pennsylvania SSA	Comprehension, analysis	1,378	1,000	1,929	133	1,333	151	http://tinyurl.com/pxrzpzb
Texas AAR	Comprehension, analysis	27.2	10.0	40.0	7.02	26.0	8.02	http://tinyurl.com/pz95p82
Utah CRT	Comprehension, analysis, composition	168	130	199	10.6	167	10.5	http://tinyurl.com/panrh5h
Washington CAP[a]	Comprehension, analysis	410	303	500	27.2	410	28.1	http://tinyurl.com/nt5rs48
Wyoming PAWS	Comprehension, analysis	601	504	770	47.2	605	47.3	http://tinyurl.com/pafylhh

(continued)

Appendix Table D.3 (continued)

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement test scores from district records for Grade 3.

NOTES: ECLS-K Reading Assessment and TOWRE2 scores were standardized using the sample mean and standard deviation. State achievement scores were standardized using state means and state standard deviations.

All links were retrieved on August 18, 2014.

[a]All Washington districts except for one submitted scaled CAP scores. This district is omitted from the table. Scores from this district were standardized using the mean and standard deviation of state raw scores.

between disability categories when students were flagged, because most districts did not provide detailed IEP categories.

Analysis of Response Rates

The focus of this section is to assess the overall response rates for the data collection activities related to the impact analysis and to compare the proportion of treatment and comparison students who had key data available — namely, an outcome test score, a fall tier assignment, and demographic variables — for the impact analysis.

Differences in the response rates of the treatment and comparison groups were compared in two ways. The first comparison was of differences for the full sample of students in the impact sample schools, to provide an overall picture of the response rate across various data sources. The study research team then checked the difference in response rate at the cut point, using a linear functional form that is similar to the impact estimation model (described in Appendix E) using only observations within the selected optimal bandwidth for the Regression Discontinuity (RD) analysis. This restriction was used because the study uses an RD design wherein the difference at the cut point provides the right check on the equivalence between treatment and comparison groups.

For the full sample of students with at least one fall 2011 universal screening test score, the differences in attrition rates for treatment and comparison groups were estimated, by grade, using the following model:

$$Y_{ij} = \sum_J \alpha_j + \beta T_{intended,ij} + \varepsilon_{ij} \quad (3)$$

where:

Y_i = 1 if student i at school j had a value for the given variable and 0 otherwise

S_{ij} = 1 if student i was in school j and 0 otherwise

$T_{intended,ij}$ = 1 if student i in school j had a rating at or below the cut point and 0 otherwise

ε_{ij} = random error, assumed to be identically and independently distributed

For the subsample of students within the optimal bandwidth, the differences in attrition rates were estimated, by grade and outcome, using the model above but with an additional regressor — the interaction between intended treatment status and a dummy variable that specifies the subsample of students within the selected optimal bandwidth. The model then becomes:

$$Y_{ij} = \sum_j \alpha_{1j} S_{ij} + T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij} T_{intended,ij} + \epsilon_{ij} \qquad (4)$$

where:

R_{ij} = rating for individual i in school j

ϵ_{ij} = random error, assumed to be identically and independently distributed

All other variables are defined as above.

Appendix Table D.4 shows the estimated differences for the full sample and at the cut point, as well as the proportion of students with available data in the full sample. The overall response rate, for both the treatment and the comparison group, is higher than 85 percent in almost all instances, with the only exception being student age. The estimated differences — though showing statistical significance in some instances — are generally less than 4 percent (for the full sample) and 2 percent (at the cut point) in magnitude. These results suggest that the RD analysis is in the "green" zone designated by the What Works Clearinghouse standard and is at low risk in terms of potential bias caused by differential responses between the treatment and comparison groups.[18]

[18]Figure III.1 and Table III.1 in the WWC "Procedures and Standards Handbook," Version 3.0 (2013); website: http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf.

The Response to Intervention (RtI) Evaluation

Appendix Table D.4

Response Rate Analysis for Full Sample and Subsample Within Optimal Bandwidth, by Grade and Key Variables

Key Variable	Treatment Mean (%)	Comparison Mean (%)	Estimated Difference	P-Value	Estimated Difference at Cut Point	P-Value
Grade 1 ECLS-K Reading Assessment						
ECLS-K score	91.8	95.6	-3.8	0.000	-0.5	0.521
Tier placement	99.9	99.9	-0.0	0.452	0.1	0.298
In ECLS-K analysis sample	91.8	95.6	-3.8	0.000	-0.6	0.502
Age	72.6	73.4	-0.8	0.043	-0.3	0.619
English Language Learner (ELL) student	86.6	88.1	-1.5	0.001	-1.3	0.047
Individualized Education Program (IEP) student[a]	92.9	94.6	-1.7	0.000	-1.1	0.135
Sex	92.3	93.3	-1.0	0.018	-0.3	0.601
Low-income status	89.8	91.1	-1.3	0.002	-1.2	0.074
Race/ethnicity	92.4	93.5	-1.1	0.023	-0.9	0.222
Grade 1 TOWRE2						
TOWRE2 score	91.4	95.4	-4.0	0.000	0.2	0.867
Tier placement	99.9	99.9	-0.0	0.452	0.0	0.495
In TOWRE2 analysis sample	91.4	95.4	-4.0	0.000	0.0	0.979
Age	72.6	73.4	-0.8	0.043	-0.4	0.545
ELL student	86.6	88.1	-1.5	0.001	-1.6	0.030
IEP student[a]	92.9	94.6	-1.7	0.000	-1.3	0.101
Sex	92.3	93.3	-1.0	0.018	-0.6	0.420
Low-income status	89.8	91.1	-1.3	0.002	-1.7	0.020
Race/ethnicity	92.4	93.5	-1.1	0.023	-1.5	0.073
Grade 2						
TOWRE2 score	93.8	95.2	-1.4	0.003	-0.1	0.916
Tier placement	99.9	99.9	-0.0	0.813	-0.1	0.470
In analysis sample	93.7	95.1	-1.4	0.003	-0.2	0.813
Age	74.0	74.7	-0.7	0.085	-0.8	0.331
ELL student	86.5	87.8	-1.3	0.002	-1.0	0.256
IEP student[a]	93.9	95.0	-1.1	0.014	-0.6	0.493
Sex	90.9	91.7	-0.8	0.067	-0.9	0.297
Low-income status	89.2	90.2	-1.0	0.016	-1.5	0.090
Race/ethnicity	93.0	93.7	-0.7	0.137	-0.5	0.575

(continued)

Appendix Table D.4 (continued)

Key Variable	Treatment Mean (%)	Comparison Mean (%)	Estimated Difference	P-Value	Estimated Difference at Cut Point	P-Value
Grade 3						
State achievement score	89.1	92.0	-2.9	0.000	-1.6	0.031
Tier placement	99.9	99.9	-0.0	0.403	-0.1	0.346
In analysis sample	89.0	91.9	-2.9	0.000	-1.7	0.024
Age	75.3	76.9	-1.6	0.000	-1.5	0.026
ELL student	85.3	87.3	-2.0	0.000	-1.2	0.082
IEP student[a]	92.5	94.7	-2.2	0.000	-1.8	0.012
Sex	92.2	94.0	-1.8	0.000	-1.5	0.033
Low-income status	87.0	88.3	-1.3	0.003	-0.7	0.334
Race/ethnicity	90.6	91.9	-1.3	0.006	-1.3	0.076

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement test scores from district records for Grade 3. Fall 2011 and winter 2012 tier placement data; student demographic data from district records.

NOTES: The differences in response rates for the full sample were estimated using Equation 3 described in Appendix D. The differences in response rates at the cut point were estimated using Equation 4 described in Appendix D. Grade 1 demographic data were based on the students with ratings within the optimal bandwidth as selected from the ECLS-K or TOWRE2 sample.

The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

The number of students with a nonmissing outcome from the full sample treatment group are: 3,407 for Grade 1, 3,419 for Grade 2, and 2,745 for Grade 3.

The number of students with a nonmissing outcome from the full sample control group are: 5,712 for Grade 1; 6,230 for Grade 2, and 5,934 for Grade 3.

The number of students with a nonmissing outcome from the bandwidth sample are: 6,655 for Grade 1 ECLS-K Reading Assessment, 5,812 for Grade 1 TOWRE, 4,511 for Grade 2, and 7,035 for Grade 3.

[a]This classification does not distinguish between reading Individual Education Programs (IEPs) and other IEPs.

Appendix E

Estimation Methods

This page has been left blank for double-sided copying.

Appendix E provides details about the estimation methods used for the primary impact estimation used in this report on the Response to Intervention (RtI) evaluation. It starts with a brief discussion about the concept of the Regression Discontinuity (RD) design, especially when there is not full compliance between intended and actual treatment status. The appendix then describes the specific estimation models used for this study.

Regression Discontinuity Design

In the context of an evaluation study, the RD design is characterized by a treatment assignment that is based on whether an applicant falls above or below a "cut point" on a "rating variable," generating a discontinuity in the probability of treatment receipt at that point. The rating variable may be any continuous variable measured before treatment, such as a pretest on the outcome variable or a rating of the quality of an application. It may be determined objectively or subjectively or in both ways. For example, students might need to meet a minimum score on an objective assessment of cognitive ability to be eligible for a college scholarship. Students who score above the minimum will receive the scholarship, and those who score below the minimum will not receive the scholarship.

RD analysis can be characterized as "discontinuity at a cut point."[19] This characterization focuses on the jump in outcome at the cut point. The direction and magnitude of the jump is a direct measure of the causal effect of the treatment on the outcome for candidates who are near the cut point.

Even though the RD design could identify a treatment effect in much the same way that a randomized controlled trial (RCT) does, in order to produce unbiased impact estimates and to approach the rigor of an RCT, it has to meet the following set of conditions:[20]

- Nothing other than treatment status is discontinuous in the analysis interval. That is, there are no other relevant ways in which students on one side of the cut point are different from or are treated differently than students on the other side of the cut point.
- The rating variable and the cut point are determined before the start of treatment and are determined independently of each other. In other words, both should be determined exogenously.

[19]Hahn, Todd, and van der Klaauw (2001).
[20]For example, Hahn, Todd, and van der Klaauw (2001) and Shadish, Cook, and Campbell (2002).

- The functional form in the estimation model that is used to adjust for the relationship between the rating variable and the outcome is continuous throughout the analysis interval, absent the treatment, and is specified correctly.

Appendix F checks the first two assumptions, while Appendix G provides evidence for the third assumption for the RD design used in this study.

There are two types of RD designs: the "sharp" design, in which all subjects receive their intended treatment or comparison condition, and the "fuzzy" design. In the *sharp design,* the difference in the probability of being in treatment is 1 between the treatment and comparison groups at the cut point. On the other hand, in the "fuzzy" design, some subjects do not receive their intended treatment or comparison condition. The fuzzy design is analogous to having, in a randomized experiment, "no-shows" (treatment group members who do not receive the treatment) and/or "crossovers" (control group members who do receive the treatment). In the *fuzzy design,* the difference in the probability of being in treatment between the treatment and the comparison groups is less than 1 at the cut point. However, this difference cannot be so small that there is essentially no meaningful treatment contrast between the two groups.

The impact analysis for this study has the characteristics of a fuzzy RD design. As discussed above, not all schools in the sample adhered to the predetermined decision rules to assign students to Tier 2 or Tier 3 interventions. This deviation from the rules created no-shows and crossovers in the sample. As a result, the difference in the probability of actually being assigned to Tier 2 or Tier 3 between the treatment and comparison groups is between 0 and 1. Appendix Figure E.1 presents the relationship between the actual assignment to Tiers 2 and 3 and the rating variable for Grades 1, 2, and 3, respectively. It shows that even though the actual assignment to treatment deviated from what should have been, based on the decision rules, there is still a sizable jump in the probability of actual assignment to treatment right at the cut point. This discontinuity in treatment status was therefore used to identify the causal impact of actual assignment to intervention. The next section discusses the analytic approach in more detail.

Analytic Approach

Following the What Works Clearinghouse (WWC) "Standards for Regression Discontinuity Designs,"[21] the study used a recommended approach that estimates the impact using a linear relationship between rating and outcomes with data points that lie within a pre-selected optimal bandwidth on either side of the cut point. Specifically, for each grade and outcome measure, data from all sample schools are pooled into one data set for analysis; both rating variable and

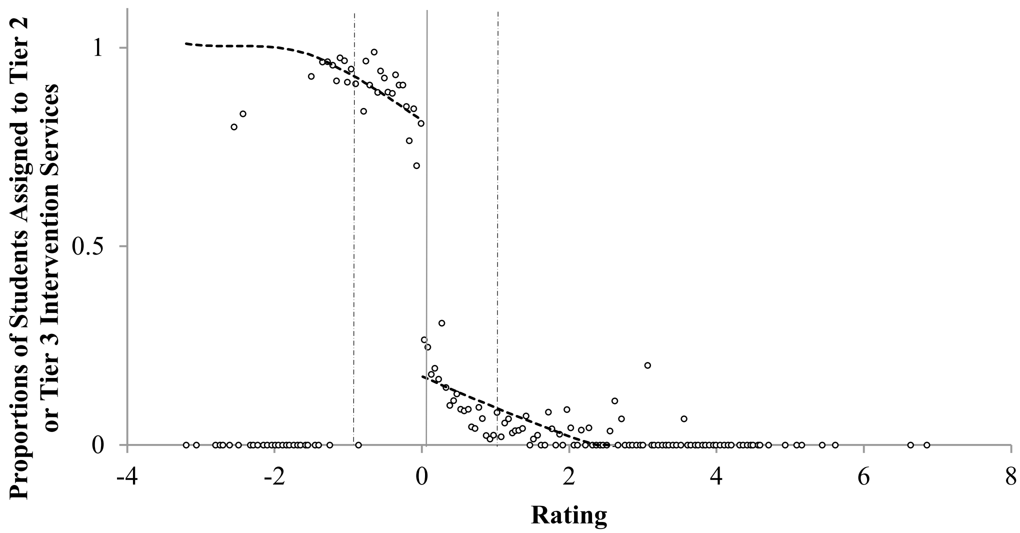
[21]Schochet et al. (2010).

The Response to Intervention (RtI) Evaluation

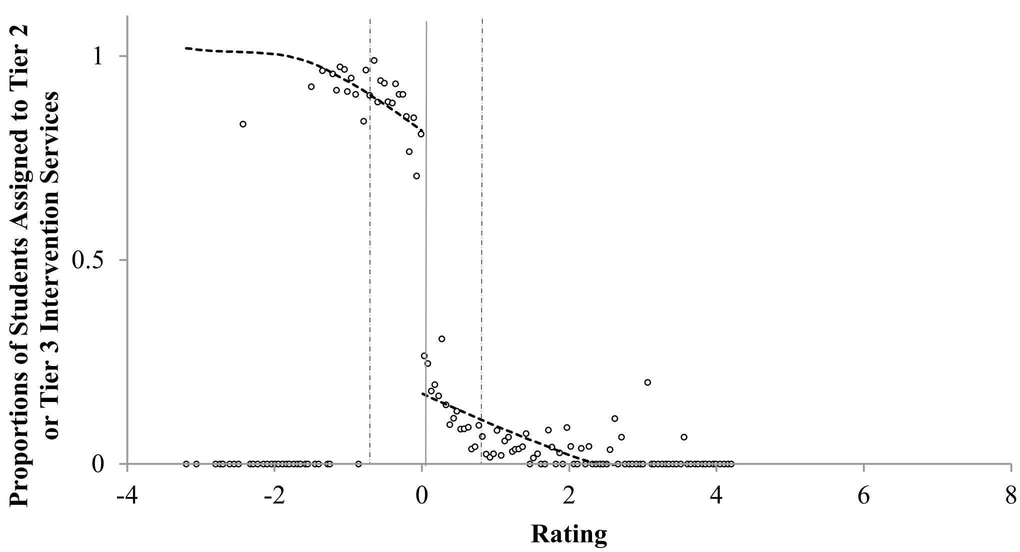
Appendix Figure E.1

Relationship Between Student Rating and Actual Assignment to Tier 2 or Tier 3 Intervention Services, by Grade and Outcome

Grade 1 ECLS-K Reading Assessment



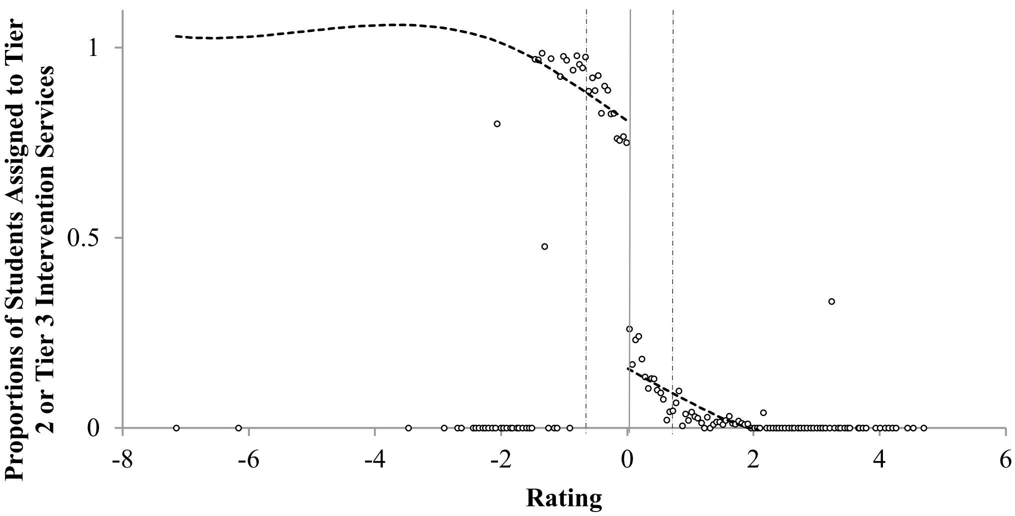
Grade 1 TOWRE2



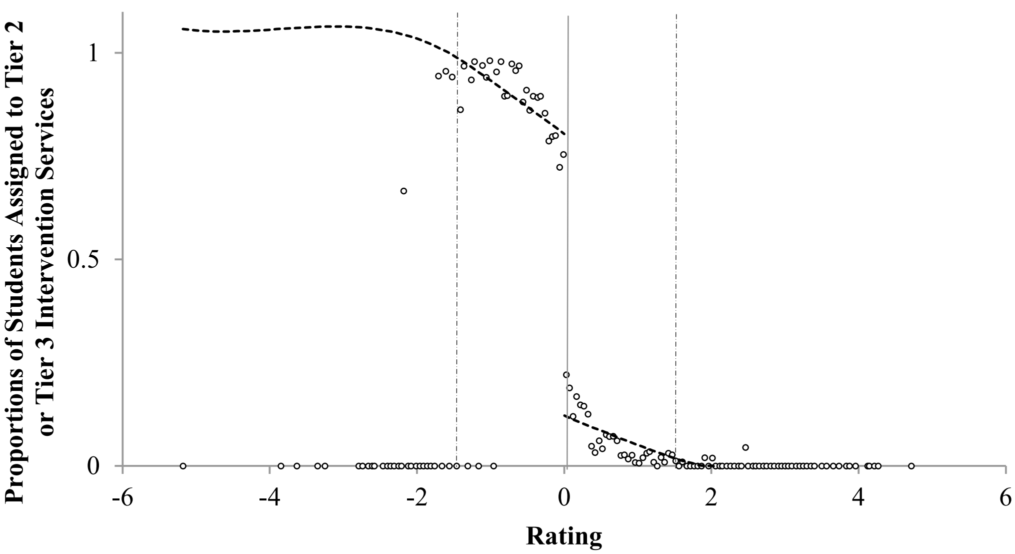
(continued)

Appendix Figure E.1 (continued)

Grade 2



Grade 3



SOURCES: Fall screening scores and student tier placement data from schools in the sample.

NOTES: Test scores have been standardized to have a mean of 0 and a standard deviation of 1.

The vertical dashed lines represent the optimal bandwidth. The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

The curved dashed line represents the smoothed Loess curve that fits the data.

Descriptions of the rating variable can be found in Appendix D.

outcomes are centered and standardized so that values from different tests are on the same effect-size metric.[22] Before conducting the impact analysis, the study research team selects an optimal bandwidth based on the pooled data set, using the algorithm described in Imbens and Kalyanaraman and recommended by the WWC.[23]

To provide context for the primary findings, the study research team first estimates the average effect of *intended* assignment to Tier 2 or Tier 3 intervention. This is done by calculating the differences in outcomes — controlling for the rating values — between (1) students whose fall screening test scores (the rating) are right below the cut point and who, therefore, should have been assigned to intervention and (2) students whose scores are right above the cut point and who should have been assigned to the comparison condition.

The study research team then fits a Two-Stage Least Squares (2SLS) linear model using only data within the optimal bandwidth on either side of the cut point to assess the effect of *actually being assigned* to Tier 2 or Tier 3 reading intervention services.[24,25] This estimate is the focus of discussion in the report.

This study modifies the WWC recommended approach in a few ways to accommodate specific features of the current analysis. First, the value of students' fall screening test scores (or the rating variable) plays an important role in determining their assignment to intervention. Given the discrete nature of certain screening tests used by sample schools, there are cases where students are clustered within a set of unique values of the rating variable, and it is those clusters of students that are assigned to treatment or comparison groups.[26] Thus, the current analysis adjusts the standard errors to account for clustering of students within unique values of the rating variable, as suggested by Lee and Card.[27]

Second, as in an RCT, the precision of the estimation for a fuzzy RD design can be improved by including students' baseline (fall) characteristics as covariates in the model. Student

[22]See Appendix D for details on the standardization process.

[23]Imbens and Kalyanaraman (2012); Schochet et al. (2010).

[24]Hahn, Todd, and van der Klaauw (2001); Lee and Lemieux (2010).

[25]Intuitively, to obtain this estimated effect of *actually being assigned* to Tier 2 or Tier 3, the estimated effect of intended assignment is rescaled by dividing the impact estimate by the difference in predicted participation rates between students just below and students just above the cut point. In the 2SLS model used here, the difference in predicted participation rates is allowed to vary by school. See a later part of this appendix for more details on the estimation model and the estimated difference in predicted participation rates. This appendix also provides more details on the preferred approach, the optimal bandwidth selection, and the 2SLS estimation model.

[26]For screening purposes, for example, 30 schools in the Grade 1 sample next used the Nonsense Word Fluency – Correct Letter Sounds (NWF-CLS) test from DIBELS. The score for this test is the number of letter sounds produced correctly in one minute. It is not surprising that multiple students end up with the same score because they pronounced the same number of letter sounds correctly within a minute.

[27]Lee and Card (2008). The standard error adjustment is discussed in more detail below in this appendix.

characteristics — such as gender, age, race/ethnicity, English Language Learner (ELL) status, Individualized Education Program (IEP) status, and low-income status — are therefore included in the preferred impact model.[28]

This following sections discuss the bandwidth selection method, the relationship between rating and outcome within the optimal bandwidth, the specific regression models, and the other analytic issues.

Optimal Bandwidth Selection

In the RD context, the most often used analytic approach (referred to as the "nonparametric" approach) involves choosing a small neighborhood (known as the "bandwidth" or "discontinuity sample") to the left and to the right of the cut point and using only data within that range to estimate the discontinuity, or "jump," in outcomes at the cut point. Choosing a bandwidth in nonparametric estimation involves finding an optimal balance between precision and bias: as the bandwidth gets larger, the estimates are more precise because more observations are used in estimation, but the potential bias could also be larger because the assumed linear specification is more likely to be wrong, given more data points. The bandwidth that best balances the trade-off between bias and precision is often referred to as the "optimal bandwidth" in the RD literature.

The optimal bandwidth selection in this study was implemented by a "plug-in" procedure proposed in Imbens and Kalyanaraman.[29] This procedure describes (using a mathematical formula) the optimal bandwidth in terms of characteristics of the actual data, with the goal of balancing the degree of bias and precision. Intuitively, this formula provides a closed-form analytic solution for the bandwidth that minimizes a particular function of bias and precision. Fan and Gijbels developed this method in the context of local linear regressions, and both Imbens and Kalyanaraman and DesJardins and McCall have adapted and modified it for the RD setting.[30]

The formula for the optimal bandwidth in an RD design is the following:[31]

$$\hat{h}_{opt} = C_K \cdot \left(\frac{2 \cdot \hat{\sigma}^2(c)/\hat{f}(c)}{(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c))^2 + (\hat{r}_+ + \hat{r}_-)}\right)^{1/5} \cdot N^{-1/5} \qquad (1)$$

[28]Appendix D reports the coding of these variables. Missing values for these variables were multiply imputed using the MIX package in R, which is based on algorithms from Schafer (1999). The RD impact estimates and standard errors were calculated separately for each of the 10 imputed data sets, and then combined using the formulas in Rubin (1996). The current appendix presents details about the multiple imputation approach used for the preferred model later.

[29]Imbens and Kalyanaraman (2012).

[30]Fan and Gijbels (1996); Imbens and Kalyanaraman (2012); DesJardins and McCall (2008).

[31]Equation 10 in Imbens and Kalyanaraman (2012).

where:

C_K = a constant specific to the weighting function in use[32]

c = the cut-point value

$\hat{\sigma}^2(c)$ = the estimated conditional variance function of the rating variable at the cut point

$\hat{f}(c)$ = the estimated density function of the rating variable at the cut point

$\hat{m}_+^{(2)}(c)$, as well as $\hat{m}_-^{(2)}(c)$, = the second derivative of the relationship between the outcome and the rating at either side of the cut point

$\hat{r}_+ + \hat{r}_-$ = the regularization term to the denominator in the equation to adjust for the potential low precision in estimating the second derivatives[33]

N = the number of observations available

To implement this procedure, one first needs to use a starting rule to get an initial "pilot" bandwidth.[34] The conditional density function $\hat{f}(c)$ and the conditional variance $\hat{\sigma}^2(c)$ are then estimated based on data within the pilot bandwidth on both sides of the cut point c. Similarly, the second derivatives $\hat{m}_+^{(2)}(c)$, $\hat{m}_-^{(2)}(c)$, as well as the regularization term $\hat{r}_+ + \hat{r}_-$, will also be estimated based on the pilot bandwidth. Once all these pieces are estimated, one can plug them into the formula and compute the optimal bandwidth.

To accommodate the fuzzy RD design, the study research team followed the procedure suggested by Imbens and Kalyanaraman. The first step used the algorithm described for the sharp RD case separately for the treatment indicator and for the outcome, to obtain relevant parameter values. Next, the initial Silverman bandwidth that used the deviations from the means was used to estimate the conditional covariance. Then, all the estimated parameters were plugged into the expression for the optimal bandwidth.[35]

To account for the demographic covariates in the regression model, the above procedure was modified by using the conditional variance of the outcome variable, given all available covariates. Specifically, before choosing the bandwidth, the outcome variable was regressed on all demographic covariates (discussed further next). The residuals from this regression were then used as the outcome variable when determining the bandwidth.

[32]The value of the constant depends on the kernel used. Following Imbens and Kalyanaraman (2012), a triangle (or edge) kernel with C_K= 3.4375 is used in this study.

[33]For derivation of the formula, see Imbens and Kalyanaraman (2012).

[34]The rule used by Imbens and Kalyanaraman (2010) is $h = 1.84 \cdot S_X \cdot N^{-1/5}$ where the sample variance of the rating variable is equal to $S_x^2 = \sum(X_i - X)^2/(N-1)$.

[35]Imbens and Kalyanaraman (2012).

In practice, an adapted version of the "rdoptband_catch" function in the R package developed by Devin Caughey from MIT is used for the optimal bandwidth selection.[36] Procedures available from other software packages are also explored, and, based on these different procedures, the main impact findings are not sensitive to different bandwidth selections based on these different procedures. (See Appendix G for more on the sensitivity checks.)

Linear Functional Form Within Optimal Bandwidth

Theoretically, the optimal bandwidth is selected so that a linear relationship between the rating variable and the outcome best captures the data within the bandwidth. This section graphically demonstrates that this is the case for the selected optimal bandwidths for this study. Appendix Figure E.2 presents four scatter plots that describe the relationships between the rating variable and the outcome within the selected optimal bandwidth. To demonstrate the relationship between these two variables, a smoothed curve was superimposed on the scatter plot.[37] Because the IK bandwidth is intended to include only observations for which a linear functional form is a reasonable choice, it was expected that the smoothed curves shown here would be approximately linear. However, the actual impacts were estimated using a linear functional form within the IK bandwidth, not the smoothed curves shown in these figures. The plots for all four outcomes presented here clearly show that a linear functional form within the selected optimal bandwidth (with slopes differing on either side of the cut point) fits the data well.

Estimation Models

Different models were used to estimate the impacts of intended and actual assignment to intervention. Specifically, the following Ordinary Least Squares (OLS) model was used to estimate the impact of intended assignment to intervention:

$$Y_{ij} = \sum_j \alpha_{1j} S_{ij} + T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij} T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \qquad (2)$$

where:

Y_{ij} = outcome for individual *i* in school *j*

S_{ij} = 1 if individual *i* was in school *j*

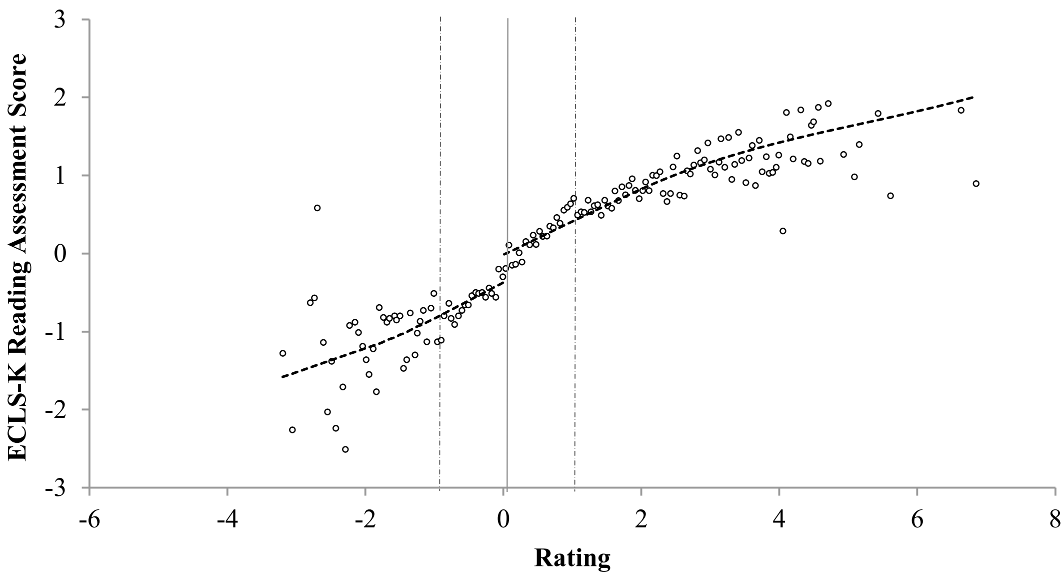
$T_{intended,ij}$ = 1 if individual *i* in school *j* should have been assigned to Tier 2 or Tier 3 intervention based on the cut point rule and 0 otherwise

[36]Website: http://web.mit.edu/caughey/www/Site/Code_files/rdoptband_catch.R.
[37]These smoothed curves are generated using the "loess.smooth" function in R.

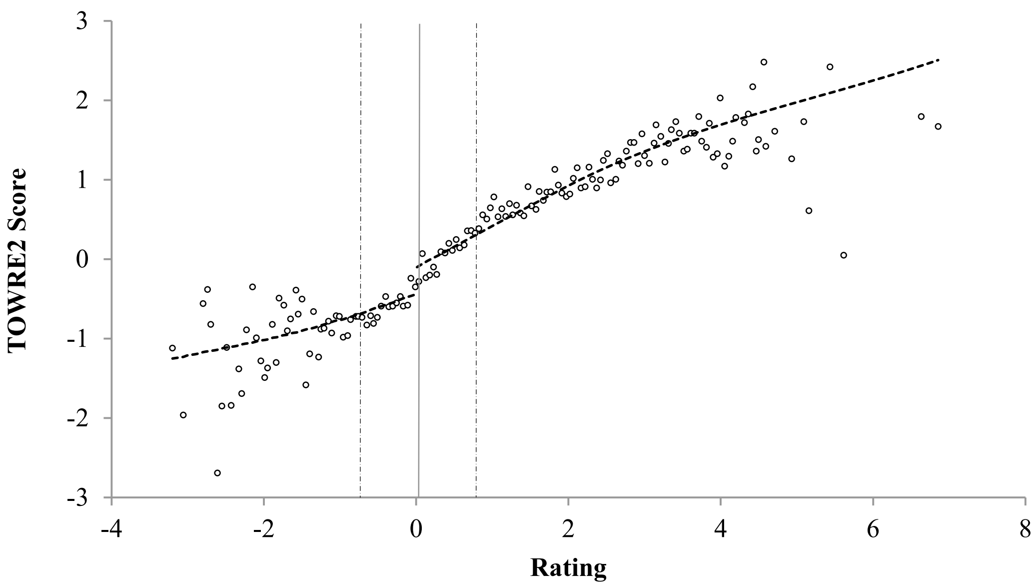
Appendix Figure E.2

Relationship Between Student Rating and Outcome Measures, by Grade and Outcome

Grade 1 ECLS-K Reading Assessment



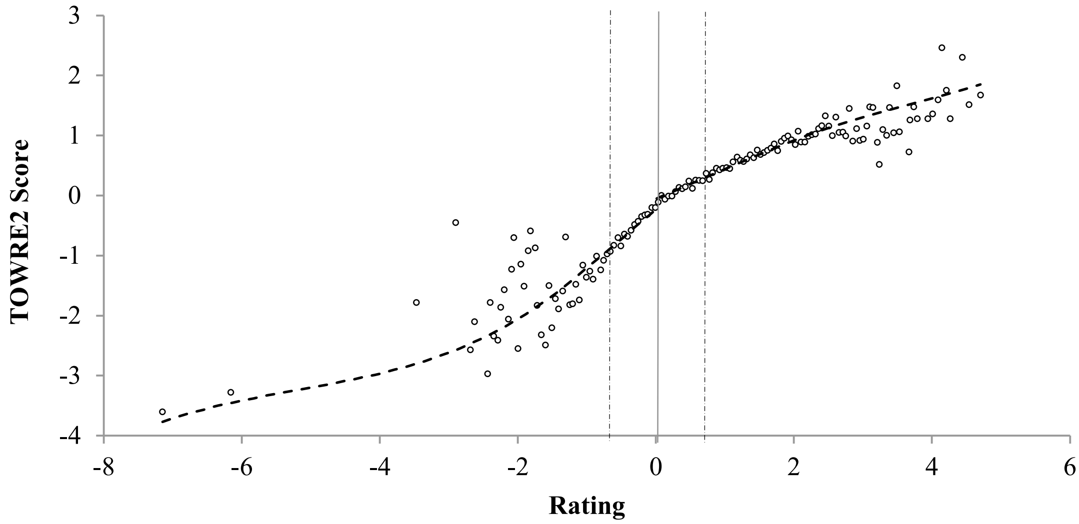
Grade 1 TOWRE2



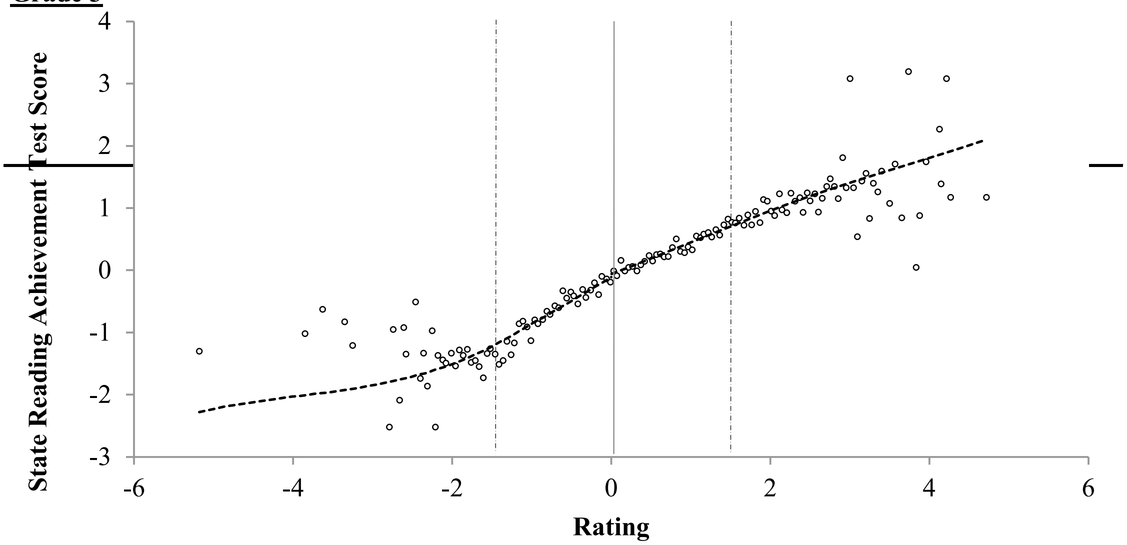
(continued)

Appendix Figure E.2 (continued)

Grade 2



Grade 3



SOURCES: Study-administered ECLS-K Reading Assessment and TOWRE2 test scores; fall benchmark test scores and student tier placement data from schools in the sample.

NOTES: Test scores were standardized to have a mean of 0 and a standard deviation of 1.

The vertical dashed lines represent the optimal bandwidth. The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

The curved dashed line represents the smoothed Loess curve that fits the data.

Descriptions of the rating variable can be found in Appendix D.

R_{ij} = rating for individual i in school j

X_{kij} = kth covariate for individual i in school j

ϵ_{ij} = random error, assumed to be identically and independently distributed

The following fixed-effect, multiple-instrument Two-Stage Least Squares (2SLS) model was used to estimate the impact of actual assignment to Tier 2 or Tier 3 intervention.

First-Stage Equation

$$T_{actual,ij} = \sum_j \alpha_{1j} S_{ij} + \sum \gamma_j S_{ij} T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij} T_{intended,ij} + \sum_k \rho_k X_{kij} + \upsilon_{ij} \quad (3)$$

Second-Stage Equation

$$Y_{ij} = \sum_j \theta_{1j} S_{ij} + \delta \widehat{T_{actual,ij}} + \theta_2 R_{ij} + \theta_3 R_{ij} T_{intended,ij} + \sum_k \varphi_k X_{kij} + \mu_{ij} \quad (4)$$

where:

$T_{actual,ij}$ = 1 if individual i in school j was actually placed in Tier 2 or Tier 3 and 0 otherwise

υ_{ij} = random error in first-stage regression, assumed to be identically and independently distributed

μ_{ij} = random error in second-stage regression, assumed to be identically and independently distributed.

And all other variables are defined as before.

There are a couple of noteworthy features in this 2SLS model:

- This model used multiple instruments ($\sum \gamma_j S_{ij} T_{intended,ij}$) instead of a single instrument ($T_{intended,ij}$). This specification was chosen to improve the precision of the estimation, as recommended by Reardon and Raudenbush.[38] Results in Appendix D show that this multiple-instrument model, indeed, had more precision than a single-instrument model and that the impact estimates were not sensitive to this specification. In addition, the first-stage F-statistics reported in the main impact table (Chapter 5, Table 5.2) indicate that the instruments used in this model were fairly strong and that bias potentially caused by weak instruments was not a concern.

[38]Reardon and Raudenbush (2013).

- Students' baseline demographic characteristics were included in the model to improve precision. Results in Appendix Table G.3 show that the impact findings were not sensitive to the inclusion of these covariates in the model.

Other Analytic Issues

These issues involve the estimation of the standard errors for the impact findings, the method used to deal with missing values, and the statistical precision of the sample.

Adjustment for the Estimated Standard Error

As discussed above, there are cases in the analysis sample in which students are clustered within unique values of the rating variable. Appendix Table E.1 presents the number of students and the number of unique rating values in the analysis sample for each grade. On average, there are about 3.4 to 4.2 students clustered within any given unique rating values.

The clustering of students was caused by the fact that the rating variable used in this study was constructed based on students' fall universal screening test scores, and many such test scores are not truly continuous. For example, a test score that counts the number of correct words that a student pronounces in a given amount of time has a finite range and is discrete within that range. As a result, it is not surprising to see multiple students with the same rating value both within a school and across schools that used the same screening tests.

The clustering of students by their rating-variable scores means that groups of students were assigned to treatment or comparison conditions. Intuitively, this is analogous to a clustered lottery in which schools rather than students are assigned to treatment and comparison groups. Thus, the estimated standard errors needed to be adjusted to reflect the clustering structure of the data. Following the suggestion by Lee and Card,[39] the primary impact analysis adjusted the estimated standard errors using the Huber-White sandwich estimators to account for clustering of students within unique values of the rating.

Multiple Imputation for Missing Covariate Values

Appendix D shows that there are missing values for some of the demographic characteristics that were used as covariates in the impact estimation model. To preserve the sample size for the impact estimation, these missing covariate values were imputed, using a multiple-imputation procedure.[40] Missing values of rating variables and reading outcomes for the impact

[39]Lee and Card (2008).
[40]Specifically, the "MIX" package in R is used for the multiple imputation.

The Response to Intervention (RtI) Evaluation

Appendix Table E.1

Counts of Students and Unique Rating Values, by Grade and Outcome

	Full Sample		Students Within Optimal Bandwidth	
Grade	Number of Students	Number of Unique Rating Values	Number of Students	Number of Unique Rating Values
Grade 1 ECLS-K Reading Assessment	8,588	2,063	6,252	1,105
Grade 1 TOWRE2	8,565	2,062	5,444	917
Grade 2 TOWRE2	9,196	2,539	4,305	1,000
Grade 3 state reading achievement tests	8,007	2,344	6,476	1,680

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample.

NOTES: The data reports on the subset of students from the analytic sample.

The numbers of schools are Grade 1 = 119, Grade 2 = 127, and Grade 3 = 112.

Descriptions of the rating variable can be found in Appendix D.

The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

analysis were *not* imputed. Ten imputed data sets were created, and estimation results from these imputed data sets were then combined using the following formulas:[41]

For point estimate: $\theta = 1/M \sum_m Q_m$

For standard error: $SE = W + (1 + M^{-1})B$

where:

[41]Rubin (1987).

Point estimates for each imputed data set are $Q_1 \ldots Q_M$, and their corresponding standard errors are $s_1, \ldots, s_M$

$$W = 1/M \sum_m s_m^2$$

$$B = 1/(M-1) \sum_m (Q_m - \theta)^2$$

$$M = 1, 2, 3, \ldots M \text{ (M = number of repetitions)}$$

Statistical Precision Based on Utilized Sample

A common way to convey a study's statistical power is through the *minimum detectable effect size* (MDES). Formally, the MDES is the smallest true program impact, scaled as an effect size, that can be detected with a reasonable degree of power (in this case, 80 percent) for a given level of statistical significance (in this case, 5 percent for a two-tailed test). The number of students — and, in the case of the RD design, the number of students within the optimal bandwidth — is a crucial factor that determines the degree to which the impacts on student outcomes can be estimated with enough precision to reject with confidence the hypothesis that the program had no effect. In general, larger sample sizes provide more precise impact estimates.

Appendix Table E.2 presents the MDES for the four reading achievement outcomes across all three grade levels in this study. The minimum detectable effect sizes in this table are based on the number of students used in the actual impact estimation for each grade and the standard errors of the estimated impact of actual assignment to intervention. Hence, the values

The Response to Intervention (RtI) Evaluation

Appendix Table E.2

Minimum Detectable Effect Size for the Impact of Assignment to Tier 2 or Tier 3 Intervention Services, by Outcome

	Grade 1		Grade 2	Grade 3
	ECLS-K Reading Assessment	TOWRE2	TOWRE2	State Reading Achievement Test
MDES	0.151	0.163	0.171	0.129

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement test scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTE: The minimum detectable effect size value is expressed in effect-size units. See Appendix E for a description of how it was calculated.

in the table represent the actual precision of the analyses. The table shows that, across the grades, the study is equipped to detect impacts on reading achievement that are as small as 0.15 to 0.16 for Grade 1, 0.17 for Grade 2, and 0.13 for Grade 3. (Numbers are expressed in effect-size units.)[42] These numbers are very close to the targeted minimum detectable effect size of 0.15 at the planning stage of the study (February 2011).

Note that the study was not designed to detect subgroup or differential subgroup effects, and all analyses reported in Chapter 6 are exploratory and for hypothesis-generating purposes only.

More on the Interpretation of Impact Findings

It is important to point out that the impact findings apply *largely but not exclusively* to students being assigned to receive Tier 2 services. This is true for two reasons:

1. On average, the rating values (calculated based on students' fall screening test scores) of Tier 2 students are closer to the cut point than those of Tier 3 students. Appendix Table E.3 demonstrates this point by presenting the mean rating values for each tier, by analysis sample. Across all grades and outcomes, the average ratings for Tier 1 students are always positive; the average ratings for Tier 2 and Tier 3 students are always negative; and the average ratings for Tier 3 students are more negative and further away from the cut point (which equals zero).

2. The rating distributions of Tier 2 and Tier 3 students do overlap, however, and there are Tier 3 students whose ratings are close to the cut point. This is clear from Appendix Figures E.3, E.4, and E.5. The figure presents the histogram of rating distributions for each tier, by analysis sample. It shows that while a large proportion of Tier 2 students have ratings just to the left of the cut point (indicated by the vertical dashed line), there is also a small proportion of Tier 3 students whose ratings lie just to the right of the cut point. This pattern can be observed in all analysis samples.

Therefore, when interpreting the impact findings of this study, it is important to distinguish between the actual student tier assignment and the proximity of student ratings to the cut point. The interpretation of an RD design is based on the latter — the closeness of ratings to

[42]Note that an estimated impact smaller than the MDES can still be found to be statistically significant. This is because the calculation of the MDES incorporates not only the probability of making Type I error (that is, concluding that there is an impact when, in fact, there is not) but also the probability of making a Type II error (that is, concluding that there is no impact when, in fact, the program was effective).

The Response to Intervention (RtI) Evaluation

Appendix Table E.3

Summary Statistics of Rating Variable, by Grade and Tier

Sample	Actual Tier Assignment	Number of Observations	Mean	Standard Deviation
Grade 1				
ECLS-K Reading Assessment	Tier 1	5,259	0.87	0.89
	Tier 2	2,176	-0.24	0.64
	Tier 3	1,153	-0.78	0.68
TOWRE2	Tier 1	5,248	0.87	0.89
	Tier 2	2,171	-0.25	0.64
	Tier 3	1,146	-0.78	0.68
Grade 2				
TOWRE2	Tier 1	5,976	0.97	0.79
	Tier 2	2,024	-0.32	0.52
	Tier 3	1,196	-0.86	0.60
Grade 3				
State reading achievement test	Tier 1	5,545	0.98	0.76
	Tier 2	1,506	-0.30	0.49
	Tier 3	956	-0.98	0.63

SOURCE: Authors' calculations based on school-reported fall screening test data.

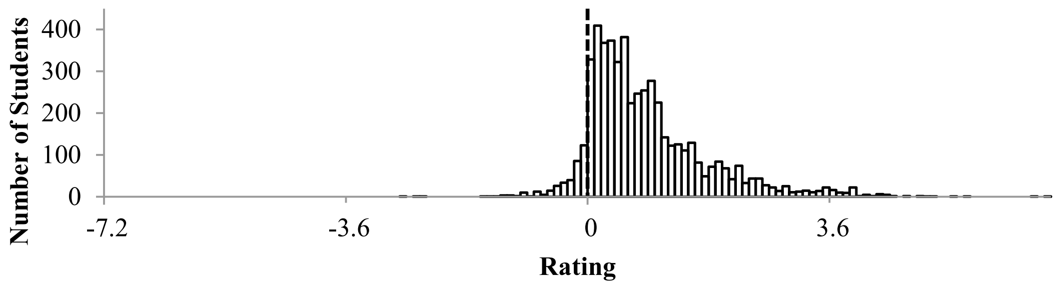
NOTE: Reported statistics represent students in the impact analysis sample.

the cut point — not on the actual tier assignment. In this specific case, even though the two are closely related (in other words, students with higher ratings were more likely to be assigned to Tier 2, and students with lower ratings were more likely to be assigned to Tier 3), this alignment is not perfect. Consequently, it can be said that the impact findings can be generalized to students with ratings close to the cut point, which most likely would include Tier 2 students and a small portion of Tier 3 students, but it is not true that the impact findings apply only to Tier 2 students.

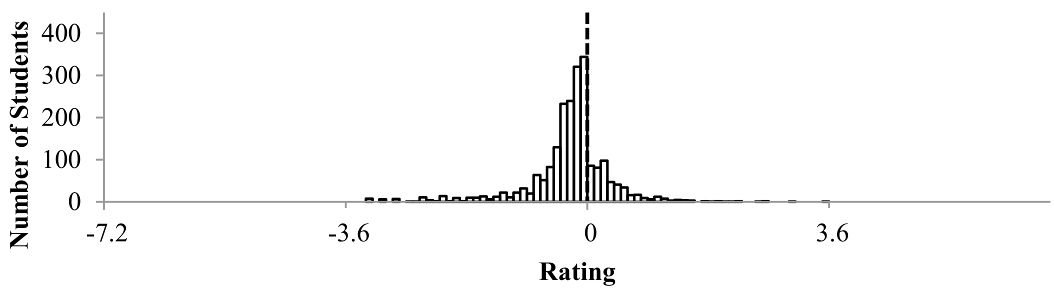
Appendix Figure E.3

Histogram of Rating Values, Grade 1

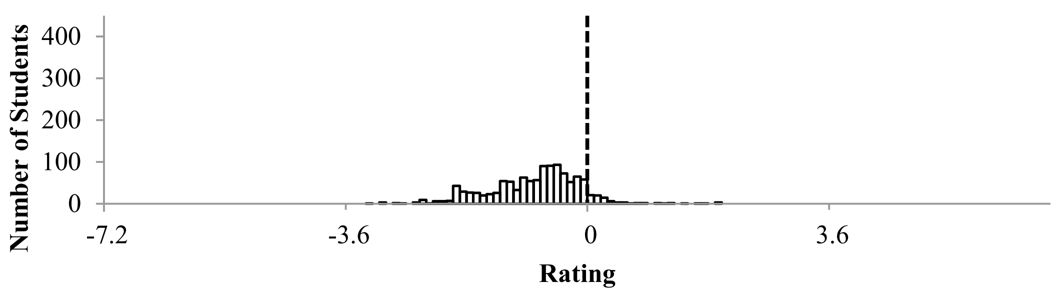
Tier 1



Tier 2



Tier 3



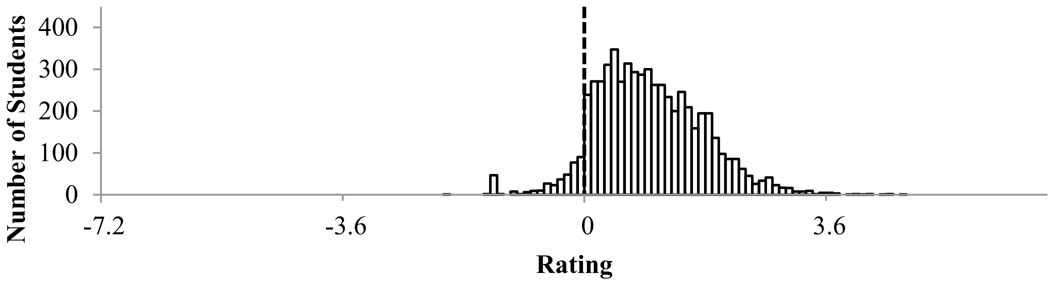
SOURCE: Authors' calculation based on school-reported fall screening test data.

NOTE: Reported statistics represent students in the impact analysis sample. Data are based on the students in the ECLS-K Reading Assessment sample.

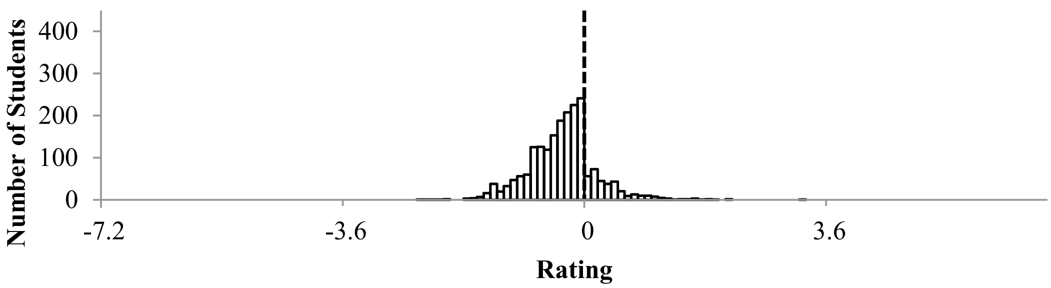
Appendix Figure E.4

Histogram of Rating Values, Grade 2

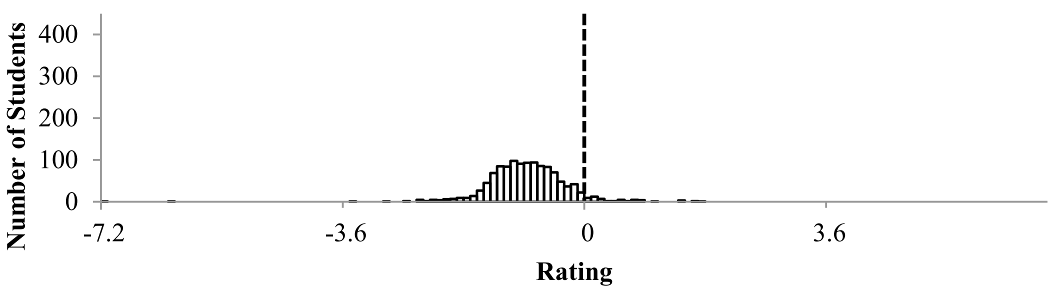
Tier 1



Tier 2



Tier 3



SOURCE: Authors' calculation based on school-reported fall screening test data.

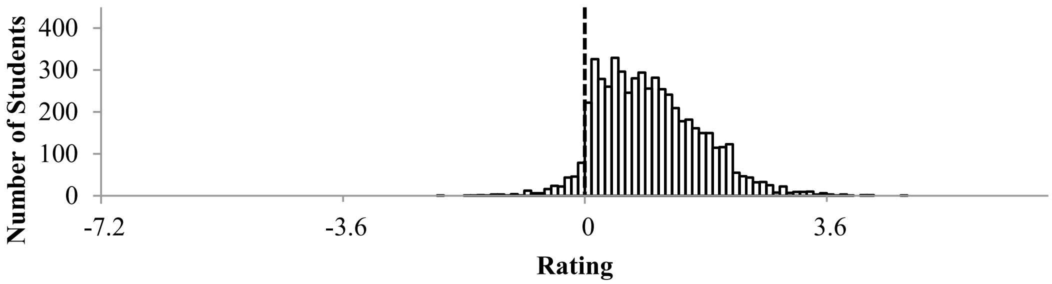
NOTE: Reported statistics represent students in the impact analysis sample.

The Response to Intervention (RtI) Evaluation

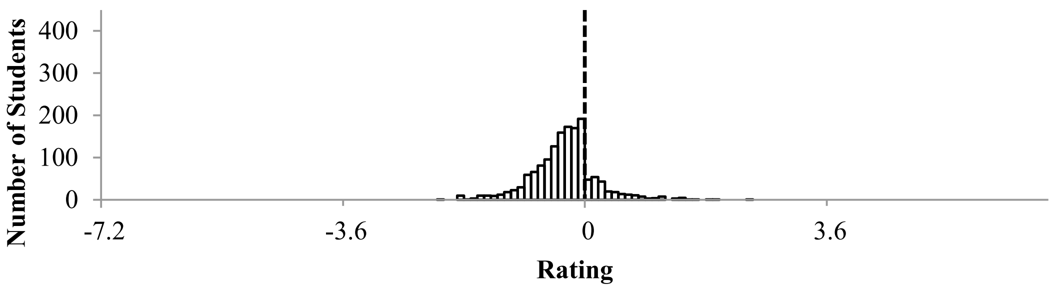
Appendix Figure E.5

Histogram of Rating Values, Grade 3

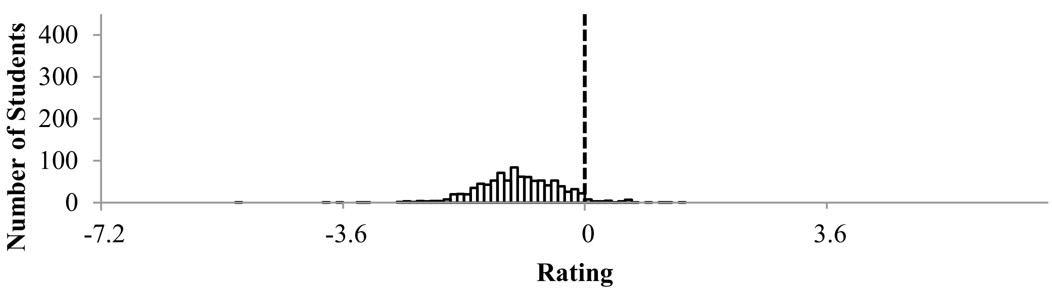
Tier 1



Tier 2



Tier 3



SOURCE: Authors' calculation based on school-reported fall screening test data.

NOTE: Reported statistics represent students in the impact analysis sample.

This page has been left blank for double-sided copying.

Appendix F

Validity of the Regression Discontinuity Design

This page has been left blank for double-sided copying.

A Regression Discontinuity (RD) design is considered to be internally valid if a valid causal inference can be made for the sample that is being observed (as opposed to the population to which these findings will be generalized).[43] Without establishing the internal validity of the RD design, no causal interpretation can be made. While a valid RD design can identify a treatment effect in much the same way that a randomized controlled trial (RCT) does, in order for an RD design to be valid, a clear discontinuity in the probability of receiving treatment must exist at the cut point, *and* candidates' ratings and the cut point must be determined independently of each other. Appendix E has demonstrated that, in the context of this Response to Intervention (RtI) study, there was a clear discontinuity in the probability of actual assignment to intervention.

Appendix F focuses on the integrity of the rating variable and cut points used in this study. The sections below investigate three potential threats to the validity of the RD design: (1) the continuity of the rating variable at the cut point, (2) manipulation of the rating variable at the cut point, and (3) data heaping.

Continuity of the Rating Variable at the Cut Point

To obtain an impact estimate with valid causal inference under an RD design, there must be evidence that, in the absence of the intervention, there would be a smooth relationship between the outcome and the rating variable at the cut point. This condition is needed to ensure that any observed discontinuity in the outcomes of the treatment group and the comparison group at the cut point can be attributable only to the intervention. However, this *smoothness condition* cannot be checked directly. Instead, it can be assessed indirectly by calculating impact estimates on student characteristics before intervention. This is similar to a baseline equivalence analysis for an RCT.

Appendix Table F.1 presents results for this baseline equivalence test for the analysis sample, by grade.[44] Tests for 11 baseline demographic characteristics of students are reported for each grade.

By chance, one would expect to see one statistically significant finding for each grade. For Grade 1, 2 of the 11 tests are statistically significant. However, the differences for both variables are less than 0.25 in effect size. There are no statistically significant differences in baseline characteristics for Grade 2 or Grade 3.

[43] Shadish, Cook, and Campbell (2002).

[44] The tests for Grade 1 are based on the sample used for the ECLS-K Reading Assessment analysis. The sample used for Grade 1 TOWRE2 analysis is very similar and, therefore, is not used here.

Appendix Table F.1

Difference in Background Characteristics, by Intended Assignment to Tier 2 or Tier 3 Intervention Services, for Students Within Optimal Bandwidth, by Grade

Grade/Characteristic	Estimated Difference at Cut Point	Effect Size of Estimated Difference	P-Value
Grade 1			
Age (years)	-0.04	-0.10	0.019
Low-income students (%)	6.48	0.13	0.002
Race/ethnicity (%)			
White, non-Hispanic	1.79	0.04	0.316
Black, non-Hispanic	0.78	0.03	0.343
Hispanic	-2.17	-0.05	0.128
Asian/Pacific Islander	-1.51	-0.07	0.212
Other	0.76	0.04	0.417
Male (%)	4.48	0.09	0.051
English Language Learners (%)	1.42	0.04	0.415
Students with Individualized Education Programs (IEP)[a] (%)	1.72	0.06	0.152
Overage for grade[b] (%)	-0.44	-0.02	0.690
Number of observations	6,049		
Grade 2			
Age (years)	0.03	0.08	0.211
Low-income students (%)	-1.07	-0.02	0.698
Race/ethnicity (%)			
White, non-Hispanic	0.69	0.01	0.766
Black, non-Hispanic	-1.58	-0.06	0.255
Hispanic	-1.37	-0.04	0.443
Asian/Pacific Islander	2.17	0.11	0.063
Other	0.44	0.02	0.706
Male (%)	2.45	0.05	0.472
English Language Learners (%)	0.22	0.01	0.888
Students with IEPs[a] (%)	3.19	0.11	0.062
Overage for grade[b] (%)	2.24	0.10	0.123
Number of observations	4,195		

(continued)

Appendix Table F.1 (continued)

Grade/Characteristic	Estimated Difference at Cut Point	Effect Size of Estimated Difference	P-Value
Grade 3			
Age (years)	0.01	0.04	0.454
Low-income students (%)	-4.00	-0.08	0.084
Race/ethnicity (%)			
White, non-Hispanic	0.69	0.01	0.685
Black, non-Hispanic	-1.17	-0.05	0.249
Hispanic	-0.07	0.00	0.959
Asian/Pacific Islander	0.34	0.02	0.703
Other	0.22	0.01	0.797
Male (%)	1.18	0.02	0.614
English Language Learners (%)	-0.65	-0.03	0.579
Students with IEPs[a] (%)	-0.47	-0.01	0.752
Overage for grade[b] (%)	1.40	0.06	0.251
Number of observations	6,360		

SOURCES: Fall screening scores from schools in the sample; student demographic data from district records.

NOTES: The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

Grade 1 data are based on the students in the sample who completed the ECLS-K Reading Assessment and in the optimal bandwidth.

The number of observations in the table represent the total number of students with data for at least one baseline characteristic. The numbers of students for specific baseline characteristics, within the optimal bandwidth, vary as follows: 5,083 - 6,004 for Grade 1; 3,565 - 4,165 for Grade 2; and 5,771 - 6,311 for Grade 3.

[a]This classification does not distinguish between reading IEPs and other IEPs.

[b]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

Manipulation of the Rating Variable at the Cut Point[45]

A key condition for an RD design to produce unbiased estimates of effects of an intervention is that there be no systematic manipulation of the rating variable. This situation is analogous to the nonrandom manipulation of treatment and comparison group assignments under an RCT. In an

[45]Note that all tests for Grade 1 are based on the ECLS-K Reading Assessment analysis sample; the analysis sample for TOWRE2 in Grade 1 largely overlaps with the ECLS-K Reading Assessment sample, and so the tables and figures that follow below do not report the results based on this sample.

RD design, "manipulation" means that rating scores for some units are systematically changed from their true values to influence treatment assignments. With nonrandom manipulation, the true relationship between the outcome and the assignment variable can no longer be identified, which could lead to biased impact estimates.

McCrary suggests a test of potential rating manipulation by testing whether the density of the rating variable is continuous over the range of values covered by the sample.[46] The McCrary test is based on an estimator for the discontinuity in the density function of the rating variable at the cutoff. The null hypothesis is that the discontinuity is zero.

Appendix Figure F.1 shows the unrestrictive frequency distribution of the rating variable for each grade in the study. These figures are unrestrictive in the sense that ratings are not grouped into equal-size bins. Rather, each bar in the figure represents the number of observations with a given unique rating value. The figures show that, for all grades, there is an observable discontinuity in rating density at around the cut point.

As shown in the first three columns of Appendix Table F.2, formal testing results confirmed this observation.[47] The estimated log differences in rating density at the cut point are statistically significant for all three grades.

The issue of manipulation was explored further by conducting the same test for each district in the sample (the four rightmost columns of Appendix Table F.2). For each grade, there were a handful of districts for which the McCrary test could not be conducted, possibly due to small sample sizes, and another handful of districts that had significant test results. These results imply that the significant McCrary test finding could have been driven by a small number of districts.

One potential explanation for the McCrary test results is that the rating variable lacked continuity across its full distribution, since it was constructed mostly using discrete test scores. As shown in Appendix Table E.1, there were three to five times as many students as unique values of the rating variable within the optimal bandwidth. This clustering of observations at unique rating values might have caused natural discontinuity in the density of rating that was not unique to the cut point.

To test for the validity of this explanation, the study research team conducted the McCrary test at a series of "fake" cut points. By construction, there should not be any rating manipulation at these cut points, since they are not the actual cut point used by the schools for tier placement purpose but, rather, are some random rating values picked by the team. If the test

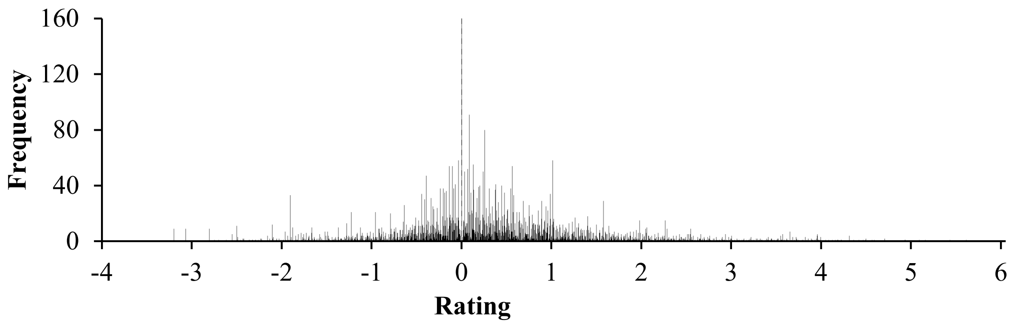
[46]McCrary (2008).
[47]The McCrary test was carried out using the "DC_Density" function in the R package.

The Response to Intervention (RtI) Evaluation

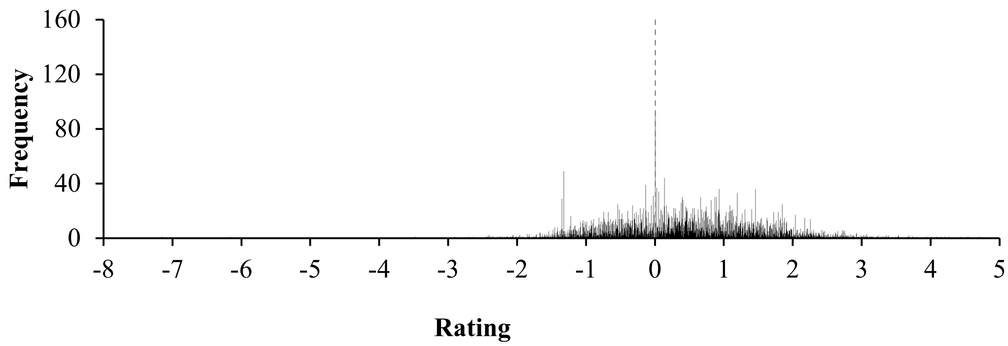
Appendix Figure F.1

Distribution of Rating Values, by Grade

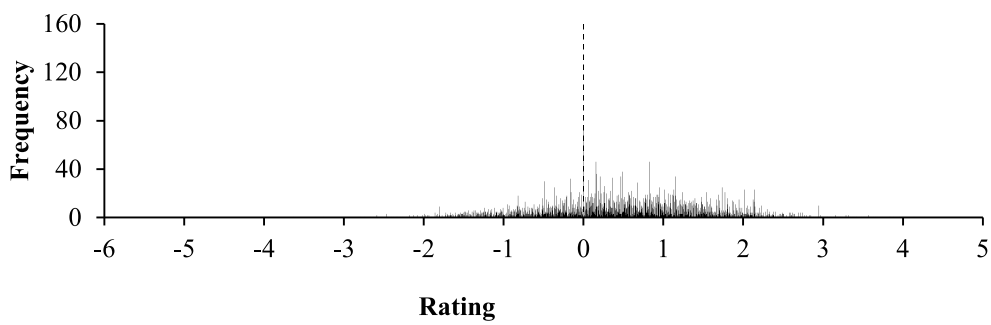
Grade 1



Grade 2



Grade 3



SOURCE: Fall screening test scores from schools in the sample.

NOTES: Grade 1 data represent the students in the sample who completed the ECLS-K Reading Assessment.

Descriptions of the rating variable can be found in Appendix D.

Appendix Table F.2

McCrary Test Results, by Grade

	Log Difference in Heights	Standard Error	P-Value	Total Number of Districts	Number of Districts That Could Not Be Tested	Number of Districts Significant at $p \leq 0.05$ Level	Number of Districts Significant at $p \leq 0.10$ Level
Grade 1	0.27	0.044	0.000	40	11	--	5
Grade 2	0.19	0.050	0.000	41	5	4	7
Grade 3	0.34	0.053	0.000	38	--	--	9

SOURCE: Fall screening scores from schools in the sample.

NOTES: Grade 1 data represent the students in the sample who completed the ECLS-K Reading Assessment.

Descriptions of the rating variable can be found in Appendix D. Descriptions of the McCrary test can be found in Appendix F.

The total number of districts is less than 43 because eligibility requirements excluded some schools in each grade.

The districts significant at the $p \leq 0.05$ level are a subset of the districts significant at the $p \leq 0.10$ level.

"--" indicates that a value has been suppressed due to small cell size.

shows significant results at these “cut points,” it is an indication that the discreteness of the rating variable could have led the McCrary test to show false positive findings. Appendix Table F.3 shows the number of fake cut points tested for each grade and the proportion of significant findings from the McCrary test. Across all grades, there are significant McCrary test results for at least 20 percent of the tests conducted, indicating significant jumps in rating density at these randomly picked rating values. This could be a result of the discreteness, or clustering of observations, in the rating variable, as discussed above. This finding indicates that the significant McCrary test results found at the real cut point may not necessarily indicate manipulation of the rating value at that point.

Data Heaping

“Data heaping” refers to the phenomenon that multiple observations share a unique rating value due to data rounding or data discretization. As Barreca et al. point out, nonrandom heaping in the rating variable — especially heaping that can be predicted by attributes related to the

The Response to Intervention (RtI) Evaluation

Appendix Table F.3

McCrary Test Results Using Artificial Cut Points, by Grade

Rating by Grade	Total Number of Artificial Cut Points Tested	Number of Cut Points Significant at $p \leq 0.05$ Level	Number of Cut Points Significant at $p \leq 0.10$ Level
Grade 1	40	26	29
Grade 2	37	13	15
Grade 3	40	8	11

SOURCE: Fall screening scores from schools in the sample.

NOTES: Grade 1 data represent the students in the sample who completed the ECLS-K Reading Assessment.

Descriptions of the rating variable can be found in Appendix D. Descriptions of the McCrary test can be found in Appendix F.

The cut points significant at the $p \leq 0.05$ level are a subset of the cut points significant at the $p \leq 0.10$ level.

outcome of interest — can lead to composition bias in the RD estimate.[48] This section assesses whether the heaping in the rating variable in the analysis sample poses any threat to the validity of the RD design for this study.

Appendix Figure F.2 clearly demonstrates the existence of data heaping in the study samples. There is observable data heaping in the analysis samples for all three grades, and, in some cases, the size of the cluster is larger than 100 observations.

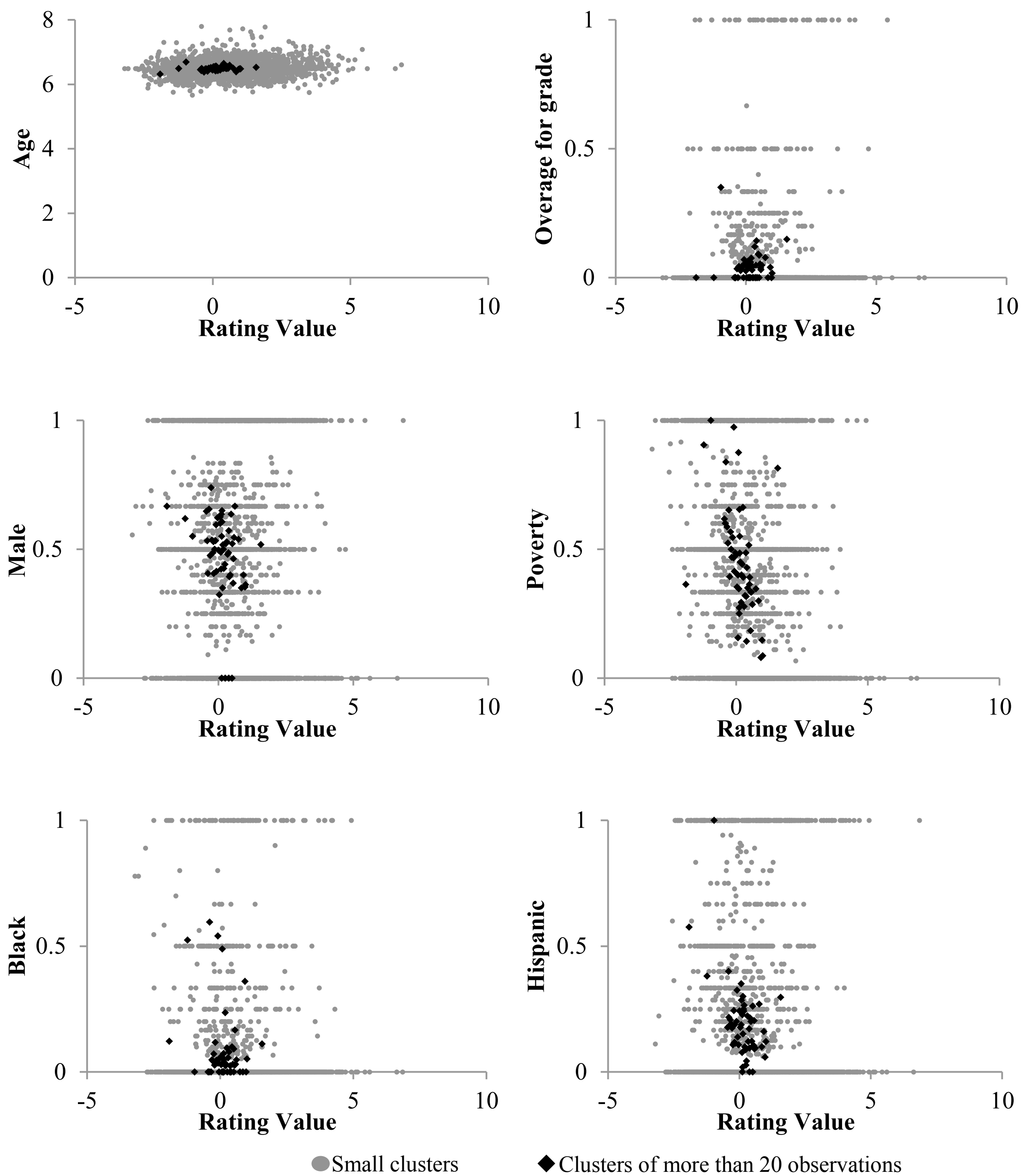
Several conclusions are supported by these plots:

- First, for all three grades, there is a big cluster of observations right at the cut point (that is, rating = 0). This is a result of the way that the rating variable was standardized. As described in Appendix D, all ratings were centered on the cut point for a given school and then were scaled by the in-sample standard deviation of the specific screening or benchmark test. As a result, all students whose fall screening test score was equal to the cut point would have a standardized rating of zero.

[48]Barreca, Lindo, and Waddell (2011).

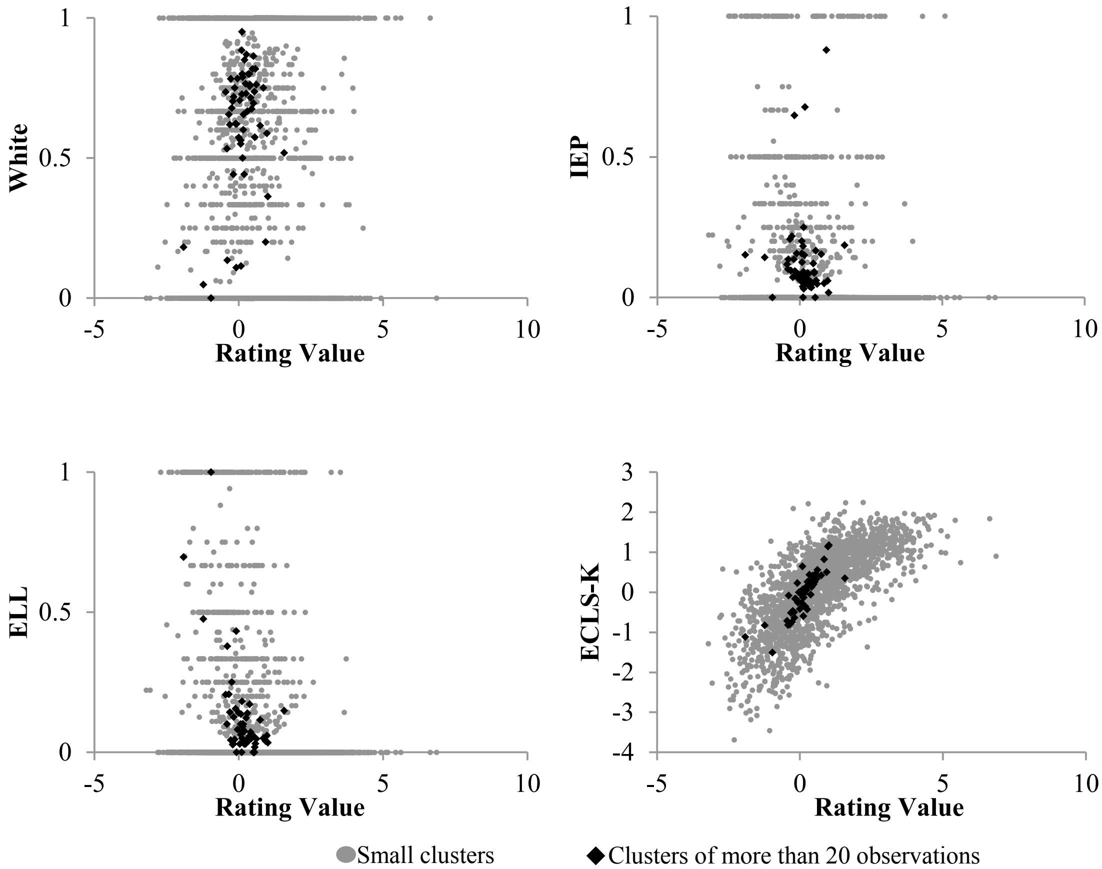
Appendix Figure F.2

Relationship Between Rating Value and Demographic Covariates and Outcomes, by Rating-Value Cluster for Grade 1 Students Who Completed the ECLS-K Reading Assessment



(continued)

Appendix Figure F.2 (continued)



SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; fall screening scores from schools in the sample; student demographic data from district records.

NOTE: Descriptions of the rating variable can be found in Appendix D.

- Further checks of the distribution of these observations with zero ratings showed that they are fairly evenly distributed across many schools. Specifically, there were 157 observations with zero rating in the Grade 1 analysis sample, obtained from 56 different schools in the sample. On average, each of these schools had about three observations in which screening test scores were right on the cut point set by the school. For Grade 2, there were 97 such observations across 37 schools; for Grade 3, there were 72 such observations from 37 schools. Therefore, it is unlikely that the heaping of observations at the cut point was caused by any one school.

- For all three grades, there are other large clusters at other rating values. Most of these clusters consist of students from schools that used certain fall screening or benchmark tests (such as those identified as F&P, DRA2, FAIR_WR, and DRLA in Appendix Table F.4). One thing that these tests have in common is that their scores are fairly discrete. For example, for one school that used F&P for first grade, the score has only six unique values. What's more, these schools tended to be in Texas, Florida, Utah, and Minnesota, where the sample sizes tend to be large and there can be more than 100 students in one grade. These two factors worked together to create some of the large clusters observed in the figures, including the large clusters at the cut point. The study research team identified 10 schools in the Grade 1 sample, 10 schools in the Grade 2 sample, and 5 schools in the Grade 3 sample that used fall benchmark tests with fairly discrete scores. Appendix Table F.4 presents a list of such schools, by grade.

The study research team identified two sources of heaping in the sample data: observations with rating values of zero and observations from schools that used a fall benchmark test with fairly discrete score scales. Sensitivity tests presented in Appendix G show that the impact estimates are not sensitive to the exclusion of these observations from the analysis. There is also evidence indicating that the data heaping observed in the analysis sample was not systematic.

To assess whether the heaping was systematic, the study research team identified heaping points (that is, clusters with more than 20 observations) and then plotted the mean values of student outcome and mean values of student baseline characteristics for each cluster against the rating (Appendix Figure F.2). The plots distinguish clusters with more than 20 observations (dark dots) and those with fewer than or equal to 20 observations (gray dots). By and large, the patterns for the large clusters do not seem to differ from the other clusters, indicating that the clustering in the sample was not systematic. These plots provide supporting evidence that the big clusters did not differ systematically from other, smaller clusters in both measured baseline variables and the outcome; therefore, they were unlikely to cause bias in the impact estimates. Results from the sensitivity checks are presented in Appendix G.

The Response to Intervention (RtI) Evaluation

Appendix Table F.4

Schools with Fall Screening Tests That Have Discrete Scales, by Grade

Screening Test	State	School	Number of Unique Score Values	Total Students
Grade 1				
BRLA	CA	A	13	106
DRA2	TX	B	12	158
DRA2	TX	C	12	146
F&P	MN	D	13	71
F&P	TX	E	6	72
F&P	UT	F	11	103
FAIR_WR	FL	G	11	57
FAIR_WR	FL	H	7	58
FAIR_WR	FL	I	11	134
FAIR_WR	FL	J	10	68
Number of observations				973
Grade 2				
DRA2	TX	B	14	158
DRA2	TX	C	15	174
F&P	MN	K	16	54
F&P	MN	D	14	64
F&P	PA	L	18	111
F&P	TX	E	8	76
F&P	UT	M	14	117
F&P	UT	F	16	82
FAIR_WR	FL	N	54	138
FAIR_WR	FL	J	39	63
Number of observations				1,037
Grade 3				
DRA2	TX	B	13	124
F&P	MN	D	14	57
F&P	TX	E	6	73
F&P	UT	M	17	117
F&P	UT	F	15	98
Number of observations				469

SOURCE: Fall screening test data from schools in the sample.

NOTE: Anonymous letter names for schools are consistent across grades.

This page has been left blank for double-sided copying.

Appendix G

Sensitivity of the Main Impact Findings

This page has been left blank for double-sided copying.

Appendix G summarizes the results from several sensitivity analyses for the main impact findings presented in this report on the Response to Intervention (RtI) evaluation. As noted, the main impact findings are based on the preferred analytic approach described in Appendix E, which involves a variety of methodological choices that could potentially influence the findings. Results presented here assess whether the main findings are sensitive to these methodological choices. In addition, Appendix F identifies potential threats to the validity of the Regression Discontinuity (RD) design caused by data heaping in certain schools. This appendix examines whether the main impact findings are sensitive to the exclusion of these observations.

Sensitivity to Alternative Bandwidth Selections

The main impact findings were estimated using the preferred approach described in Imbens and Kalyanaraman for optimal bandwidth selection.[49] Details of this approach are described in Appendix E. In carrying out this method, the study research team used a function in R that is adopted from the "rdoptband_catch" function developed by Devin Caughey from MIT.[50] This function differs from other available software routines in that it allows for fuzziness of the RD design and covariates in the model and, therefore, is most suitable for the current design. However, other available software routines provide viable alternatives to this benchmark approach. The study research team tests the sensitivity of the main estimation results using these alternatives for optimal bandwidth selection. Specifically, the following alternatives are used:

- Alternative I: using the original "rdoptband_catch" function in R as developed by Caughey. This function assumes a sharp RD design.
- Alternative II: using the -rdbwselect- command in STATA developed by Calonico, Cattaneo, and Titiunik.[51] This approach assumes a sharp design.
- Alternative III: using the -rdob- command in STATA developed by Imbens and Kalyanaraman. This approach assumes a sharp design. Note that the -rd- command in STATA developed by Nichols generates the same optimal bandwidths.[52]

Even though these alternative methods are all based on the "plug-in" approach proposed in Imbens and Kalyanaraman in principle, they differ in the specific way that they optimize the procedure and, therefore, offer different optimal bandwidth choices. Appendix

---

[49] Imbens and Kalyanaraman (2012).

[50] Caughey (2009 to 2012). Website: http://web.mit.edu/caughey/www/Site/Code_files/rdoptband_catch.R.

[51] Calonico, Cattaneo, and Titiunik (2014).

[52] Imbens and Kalyanaraman (2012); Nichols (2007).

Table G.1 presents the alternative optimal bandwidth choices based on these three approaches; the first column reports the benchmark bandwidth. Appendix Table G.2 presents the impact estimation results based on these alternative optimal bandwidths; for comparison purposes, the main results are reported in the first set of columns.

Results in Appendix Table G.2 show that the point estimates are fairly consistent and robust across all these bandwidth choices. The significant levels of the findings vary a bit, especially for the results on Grade 1 and Grade 2 TOWRE2. This is most likely caused by the fact that narrower bandwidth led to smaller sample sizes for the impact estimation, and smaller sample size results in less statistical precision even if the point estimate remains the same.

The Response to Intervention (RtI) Evaluation

Appendix Table G.1

Alternative Optimal Bandwidth Selection for Impact Analysis Generated by Different Software Packages, by Grade

	Using R Packages		Using Stata Packages	
Grade/Outcome	Bandwidths Used in Preferred Model[a]	Alternative I Bandwidths	Alternative II Bandwidths	Alternative III Bandwidths
Grade 1				
ECLS-K Reading Assessment	0.99	0.94	0.89	0.62
TOWRE2	0.79	0.77	0.78	0.63
Grade 2				
TOWRE2	0.70	0.70	0.68	0.80
Grade 3				
State reading achievement test	1.50	1.53	1.22	1.00

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure in the preferred model was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details. A complete description of the alternative bandwidth selection methods can be found in Appendix G.

[a]These bandwidths are the ones used for the primary impact findings reported in Table 5.2.

Appendix Table G.2

Impact of Assignment to Tier 2 or Tier 3 Intervention Services, Based on Alternative Optimal Bandwidths, by Grade

	Using R Packages				Using Stata Packages			
	Bandwidth Used in Preferred Model[a]		Alternative I Bandwidth		Alternative II Bandwidth		Alternative III Bandwidth	
Grade/Outcome	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	
Grade 1								
ECLS-K Reading Assessment	-0.17	0.002	-0.15	0.007	-0.14	0.010	-0.15	0.015
	(0.054)		(0.056)		(0.056)		(0.061)	
TOWRE2	-0.11	0.057	-0.10	0.080	-0.10	0.069	-0.10	0.107
	(0.058)		(0.057)		(0.057)		(0.061)	
Grade 2								
TOWRE2	0.10	0.084	0.10	0.084	0.10	0.100	0.12	0.048
	(0.061)		(0.060)		(0.061)		(0.059)	
Grade 3								
State reading achievement test	-0.01	0.823	-0.01	0.841	-0.03	0.490	-0.02	0.761
	(0.046)		(0.046)		(0.049)		(0.055)	

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure in the preferred model was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details. A complete description of the alternative bandwidth selection methods can be found in Appendix G.

(continued)

Appendix Table G.2 (continued)

All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect-size units. The impact of actual assignment to Tier 2 or Tier 3 intervention services for the preferred model is estimated using a 2SLS regression of the outcome on the indicator of students receiving intervention in at least the fall semester, using treatment status as determined by the school decision rule interacted with school indicators as the instrument variables. A complete description of the estimation model can be found in Appendix E.

A two-tailed t-test was applied to the estimated effect.

Using the bandwidths from the preferred model, first-stage F-statistics are 92.0 for Grade 1 ECLS-K; 79.3 for Grade 1 TOWRE2; 60.4 for Grade 2; 121 for Grade 3. Using Alternative I Bandwidth, first-stage F-statistics are 88.1 for Grade 1 ECLS-K; 78.0 for Grade 1 TOWRE2; 60.5 for Grade 2; 123 for Grade 3. Using Alternative II Bandwidth, first-stage F-statistics are 85.2 for Grade 1 ECLS-K; 78.7 for Grade 1 TOWRE2; 58.7 for Grade 2; 103 for Grade 3. Using Alternative III Bandwidth, first-stage F-statistics are 66.5 for Grade 1 ECLS-K; 67.4 for Grade 1 TOWRE2; 70.6 for Grade 2; 85.2 for Grade 3.

The numbers of students vary by bandwidth selection method used and by outcome. Using the bandwidths from the preferred model: 6,224 for Grade 1 ECLS-K; 5,448 for Grade 1 TOWRE2; 4,305 for Grade 2; 6,478 for Grade 3. Using Alternative I Bandwidth: 6,005 for Grade 1 ECLS-K; 5,364 for Grade 1 TOWRE2; 4,308 for Grade 2; 6,548 for Grade 3. Using Alternative II Bandwidth: 5,834 for Grade 1 ECLS-K; 5,407 for Grade 1 TOWRE2; 4,224 for Grade 2; 5,716 for Grade 3. Using Alternative III Bandwidth: 4,810 for Grade 1 ECLS-K; 4,850 for Grade 1 TOWRE2; 4,815 for Grade 2; 4,943 for Grade 3.

[a]These bandwidths are the ones used for the primary impact findings reported in Table 5.2.

Sensitivity to Alternative Model Specifications

The study research team assessed the robustness of the main impact findings with regard to different model specifications. Appendix Table G.3 presents impact estimates based on the following three alternative model specifications.

1. Model with no covariates

Students' demographic characteristics at baseline were included as covariates in the preferred impact model (the first pair of columns in Appendix Table G.3) to improve the precision of the estimation and to control for accidental imbalance in certain variables between the treatment and comparison groups at the cut point. Results reported in the table's second pair of columns show that the impact estimates remain essentially the same when these demographic covariates are dropped from the model.

2. Model that constrains the relationship between rating and outcome to be the same on both sides of the cut point within the optimal bandwidth

The primary impact model allowed the linear relationship between rating and outcome to have different slopes on the left side and the right side of the cut point. This specification gave the model more flexibility in accounting for the true relationship between these two variables. However, it would be of interest to see whether the results are sensitive to this specification, because if the linear relationships on both sides of the cut point are the same — in other words, if the regression lines to the left and to the right of the cut point are parallel — then it is potentially possible to generalize the impact findings to observations away from the cut point.[53] Results in the third pair of columns in Appendix Table G.3 suggest that the primary impact estimates are not sensitive to this alternative model specification.

3. Model that uses a single instrumental variable rather than multiple instrumental variables

The Two-Stage Least Squares (2SLS) model described in Appendix B used multiple instrumental variables to account for the variation in compliance across schools in the sample (as recommended by Reardon and Raudenbush)[54] and, as a result, to improve the precision of

[53]To be able to do so convincingly, however, other tests are required. For example, Wing and Cook (2013) discuss using baseline outcome measures to demonstrate that the relationship remains unchanged before and after the program is implemented, therefore allowing the generalization of the impact findings to a broader population. Nomi and Raudenbush (2012) demonstrate this approach using the "double dose" example. However, this study lacks the baseline outcome data to test this assumption.

[54]Reardon and Raudenbush (2013).

Appendix Table G.3

Impact of Actual Assignment to Tier 2 or Tier 3 Intervention Services with Alternative Model Specifications, by Grade

	Preferred Impact Model		Model Without Demographic Covariates		Model With Constrained Relationship Between Rating and Outcome		Model With a Single Instrumental Variable	
Grade/Outcome	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value
Grade 1								
ECLS-K Reading	-0.17	0.002	-0.19	0.001	-0.17	0.002	-0.24	0.000
Assessment	(0.054)		(0.057)		(0.054)		(0.064)	
TOWRE2	-0.11	0.057	-0.13	0.028	-0.11	0.055	-0.24	0.002
	(0.058)		(0.059)		(0.057)		(0.077)	
Grade 2								
TOWRE2	0.10	0.084	0.09	0.133	0.11	0.082	0.08	0.305
	(0.061)		(0.058)		(0.061)		(0.077)	
Grade 3								
State reading	-0.01	0.823	0.01	0.895	-0.03	0.508	-0.02	0.721
achievement test	(0.046)		(0.048)		(0.046)		(0.051)	

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect-size units. The impact of actual assignment to Tier 2 or Tier 3 intervention services for the preferred model is estimated using a 2SLS regression of the outcome on the indicator of students receiving intervention in at least the fall semester, using treatment status as determined by the school decision rule interacted with school indicators as the instrument variables. A complete description of the estimation model can be found in Appendix E. A complete description of the alternative model specifications can be found in Appendix G.

A two-tailed t-test was applied to the estimated effect.

(continued)

Appendix Table G.3 (continued)

In the preferred impact model, first-stage F-statistics are 92.0 for Grade 1 ECLS-K; 79.3 for Grade 1 TOWRE2; 60.4 for Grade 2; 121 for Grade 3. In the model without demographic covariates, first-stage F-statistics are 93.8 for Grade 1 ECLS-K; 76.8 for Grade 1 TOWRE2; 64.8 for Grade 2; 127 for Grade 3. In the model with a constrained relationship between the rating and outcome, first-stage F-statistics are 92.4 for Grade 1 ECLS-K; 79.6 for Grade 1 TOWRE2; 60.6 for Grade 2; 122 for Grade 3. In the model with a single instrumental variable, first-stage F-statistics are 159 for Grade 1 ECLS-K; 136 for Grade 1 TOWRE2; 104 for Grade 2; 215 for Grade 3.

In the preferred impact model, the model with a constrained relationship between rating and outcome, and the model with a single instrumental variable, the numbers of students are 6,224 for Grade 1 ECLS-K; 5,448 for Grade 1 TOWRE2; 4,305 for Grade 2; 6,478 for Grade 3. In the model without demographic covariates, the numbers of students are 6,136 for Grade 1 ECLS-K; 5,211 for Grade 1 TOWRE2; 4,443 for Grade 2; 6,547 for Grade 3.

the impact estimation. However, it is of interest to see whether the estimation results are sensitive to this specification. The rightmost pair of columns in Appendix Table G.3 present results based on a 2SLS model that uses students' intended assignment to intervention as the single instrumental variable. These point estimates are fairly consistent with the primary findings, even though, as expected, the standard errors for them are larger than those for the primary estimates, which lead to less significant findings. These results confirm the choice of using multiple instrumental variables in the preferred model.

Sensitivity to Alternative Sample Specification

The primary impact findings were estimated using all available observations in the analysis samples. However, some observations in these samples have extreme rating values. For example, 29 observations in Grade 1 have rating values that are 4 standard deviations away from zero. In addition, Appendix F shows that there might be observations and schools in the sample that caused data heaping, which might have led to bias in the impact findings. To assess the sensitivity of the main impact findings regarding these issues, the study research team reestimated the impacts of actual assignment to intervention using the following alternative samples:[55]

- **Excluding outliers.** Grades 1, 2, and 3 had, respectively, 29, 10, and 6 observations that were excluded from the samples because their rating variable values are 4 standard deviations away from zero.

- **Excluding observations with zero rating values.** As discussed in Appendix F, there is a large cluster of observations exactly at the cut point. To see whether this kind of clustering affects the sensitivity of the impact results, these observations (157, 93, and 72 for Grades 1, 2, and 3, respectively) were dropped from the sample.

- **Excluding observations from schools with discrete fall screening test scores.** As discussed in Appendix F, these schools used F&P, FAIR, DRA2, or DRLA tests as fall screening tests. In this category are 10 schools in the Grade 1 analysis sample, 10 schools in the Grade 2 sample, and 5 schools in the Grade 3 sample.

---

[55]Note that the exclusion of certain observations from the sample is done before the optimal bandwidth selection. Therefore, even though certain observations are excluded, the actual number of observations used in the regression could be larger than that of the benchmark approach because the new optimal bandwidth could be wider than what it was before.

Appendix Table G.4 presents the impact estimates based on these alternative samples. The alternative point estimates are fairly consistent with the primary findings, while their significance levels vary slightly from the benchmark. This was caused primarily by the change in sample sizes used for the estimation.[56]

[56]Note that once the outliers are dropped, a new optimal bandwidth is selected for the new sample, which could be wider or narrower than the original bandwidth, causing changes in the sample sizes used in these sensitivity checks.

The Response to Intervention (RtI) Evaluation

Appendix Table G.4

Impact of Actual Assignment to Tier 2 or Tier 3 Intervention Services with Alternative Sample Specifications, by Grade

	Sample Used in Preferred Impact Model		Sample Excluding Outliers		Sample Excluding Observations with Rating Value of Zero		Sample Excluding Sites with Discrete Benchmark Scores and Students with Rating Value of Zero	
Grade/Outcome	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value	Estimated Impact (Standard Error)	P-Value
Grade 1								
ECLS-K Reading Assessment	-0.17 (0.054)	0.002	-0.15 (0.056)	0.009	-0.16 (0.062)	0.012	-0.12 (0.060)	0.050
TOWRE2	-0.11 (0.058)	0.057	-0.12 (0.057)	0.028	-0.13 (0.063)	0.042	-0.08 (0.062)	0.173
Grade 2								
TOWRE2	0.10 (0.061)	0.084	0.10 (0.061)	0.088	0.14 (0.063)	0.030	0.15 (0.062)	0.013
Grade 3								
State reading achievement test	-0.01 (0.046)	0.823	-0.01 (0.046)	0.852	-0.03 (0.049)	0.533	-0.01 (0.047)	0.895

SOURCES: Study-administered ECLS-K Reading Assessment scores for Grade 1; study-administered TOWRE2 test scores for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records.

NOTES: All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect-size units. The impact of actual assignment to Tier 2 or Tier 3 intervention services for the preferred model is estimated using a 2SLS regression of the outcome on the indicator of students receiving intervention in at least the fall semester, using treatment status as determined by the school decision rule interacted

(continued)

Appendix Table G.4 (continued)

with school indicators as the instrument variables. A complete description of the estimation model can be found in Appendix E. A complete description of the alternative sample specifications can be found in Appendix G.

A two-tailed t-test was applied to the estimated effect.

In the sample from the preferred impact model, first-stage F-statistics are 92.0 for Grade 1 ECLS-K; 79.3 for Grade 1 TOWRE2; 60.4 for Grade 2; 121 for Grade 3. In the sample excluding outliers, first-stage F-statistics are 84.5 for Grade 1 ECLS-K; 80.0 for Grade 1 TOWRE2; 58.7 for Grade 2; 124 for Grade 3. In the sample excluding students with a rating value of zero, first-stage F-statistics are 86.2 for Grade 1 ECLS-K; 74.5 for Grade 1 TOWRE2; 55.1 for Grade 2; 115 for Grade 3. In the sample excluding these students as well as sites with discrete benchmark scores, first-stage F-statistics are 85.6 for Grade 1 ECLS-K; 72.0 for Grade 1 TOWRE2; 67.4 or Grade 2; 127 for Grade 3.

In the sample from the preferred impact model, the numbers of students are 6,224 for Grade 1 ECLS-K; 5,448 for Grade 1 TOWRE2; 4,305 for Grade 2; 6,478 for Grade 3. In the sample excluding outliers, the numbers of students are 5,803 for Grade 1 ECLS-K; 5,529 for Grade 1 TOWRE2; 4,217 for Grade 2; 6,595 for Grade 3. In the sample excluding students with a rating value of zero, the numbers of students are 5,994 for Grade 1 ECLS-K; 5,321 for Grade 1 TOWRE2; 4,062 for Grade 2; 6,184 for Grade 3. In the sample excluding these students as well as sites with discrete benchmark scores, the numbers of students are 5,206 for Grade 1 ECLS-K; 4,462 for Grade 1 TOWRE2; 3,987 for Grade 2; 6,121 for Grade 3.

This page has been left blank for double-sided copying.

Appendix H

Exploratory Analyses

This page has been left blank for double-sided copying.

Appendix H discusses the analytic approach used for the exploratory analyses presented in Chapter 6 of this report on the Response to Intervention (RtI) evaluation.[57] Specifically, it describes the method used to describe and measure the variation in impact estimates across schools, how the school-level features associate with the school-level impact estimates, and how the impact estimates vary with certain characteristics. For each of these topics, it presents the methods first and then reports supplementary results using these methods.

Variation in Impact Estimates Across Schools

A two-step approach was used to assess the variation in impact estimates across schools. This approach closely follows the recommendation by Bloom, Raudenbush, Weiss, and Porter.[58]

First, the Two-Stage Least Squares (2SLS) model used for the primary impact estimation was modified to estimate a separate mean effect of reading intervention in an RtI school ($\hat{\delta}_j$) for each school (j). Specifically, the following models were used for the estimation.

First-Stage Equations (one for each school-by-actual-assignment interaction)

$$S_{i1}T_{actual,ij} = \sum_j \alpha_{1j}S_{ij} + \sum \gamma_j S_{ij}T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij}T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \quad (1.1)$$

$$S_{i2}T_{actual,ij} = \sum_j \alpha_{1j}S_{ij} + \sum \gamma_j S_{ij}T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij}T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \quad (1.2)$$

…

$$S_{iJ}T_{actual,ij} = \sum_j \alpha_{1j}S_{ij} + \sum \gamma_j S_{ij}T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij}T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \quad (1.J)$$

Second-Stage Equation

$$Y_{ij} = \sum_j \theta_{1j}S_{ij} + \sum_j \delta_j \widehat{S_{ij}T_{actual,ij}} + \theta_2 R_{ij} + \theta_3 R_{ij}T_{intended,ij} + \sum_k \varphi_k X_{kij} + \mu_{ij} \quad (2)$$

where:

Y_{ij} = outcome for individual *i* in school *j*

S_{ij} = 1 if individual *i* was in school *j*

[57] Note that all analyses discussed in this appendix are based on the estimated impact of actual assignment to Tier 2 and Tier 3 intervention services.

[58] Bloom, Raudenbush, Weiss, and Porter (under review).

$T_{actual,ij}$ = 1 if individual i in school j was actually placed in Tiers 2 or 3 and 0 otherwise

$T_{intended,ij}$ = 1 if individual i in school j should have been assigned to Tiers 2 or 3 based on the decision rule and 0 otherwise

R_{ij} = rating for individual i in school j

X_{kij} = kth student-level covariate for individual i in school j, including students' gender, age, race/ethnicity, ELL, IEP, and low-income status

ϵ_{ij} = random error in first-stage regression, assumed to be identically and independently distributed

μ_{ij} = random error in second-stage regression, assumed to be identically and independently distributed

This step yielded a separate estimate of the actual assignment impact and its corresponding error variance ($\hat{\delta}_j$ and $\hat{V}_j$) for each RtI school in the sample.

Second, a V-Known random-effects meta-analysis approach was used to detect and quantify the true variation in the estimated impacts across RtI schools.[59] Specifically, this step estimated the following unconditional or "empty" model of variation in the impact estimates across schools:

$$\delta_j = \delta + r_j \tag{3}$$

where:

$\delta =$ the grand mean effect of receiving intervention

$r_j =$ an error term that is distributed independently and identically across schools with a mean of zero and a variance of $\tau_r^2 = \tau_\delta^2$

This step produced an estimate, $\hat{\tau}_\delta^2$, of the variance of the estimated impacts across schools. In addition, by computing a Q-statistic based on the estimated values of $\hat{\delta}_j$ and $\hat{V}_j$, this step assessed the statistical significance of $\hat{\tau}_\delta^2$.

The magnitude of $\hat{\tau}_\delta^2$ indicates the extent to which the LATE effects vary across schools, which is important substantive information. The statistical significance of this estimate indicates how likely it is to represent a real cross-school difference in the effects of receiving

[59]For a description of this approach, see Raudenbush and Bryk (2002) and Konstantopoulos and Hedges (2004).

intervention (rather than an observed difference that could happen by chance when, in fact, there was no real difference).[60]

The output from the V-Known random-effects meta-analysis was then used to obtain an Empirical Bayes estimate ($\hat{\delta}_j^*$) of the impact for each RtI school. These estimates, as well as their corresponding standard errors, were then used to make the caterpillar plots presented in Chapter 6. These plots provide helpful illustration for the spread of these estimates and the confidence intervals around them.

School-Level Correlational Analysis

The correlational analyses at school level used a V-known random-effects meta-analysis to estimate the following model, which used school-level characteristics or moderators to predict variation in the impact estimates across schools.

$$\delta_j = \delta_0^* + \sum_{f=1}^{F} \lambda_f \cdot Z_{f,j} + r_j' \qquad (4)$$

where:

$Z_{f,j}$ = the value of the f[th] school-level characteristic for school j

Estimated values for the λ_f reflect the extent to which RtI intervention effects vary as a function of exogenous school characteristics, with or without controlling statistically for all other predictors in the model.

Based on this framework, the specific analyses were carried out in two ways:

1. A *bivariate regression model* was estimated to examine the relationship between each of the features and the school-level impacts on assignment to intervention. This allowed for an assessment of this relationship for one feature at a time, independent of other factors. For dichotomous factors, this approach is equivalent to a subgroup analysis whereby subgroups are defined by this factor.

2. A *multivariate regression model* was estimated to examine the relationship between groups of features (as categorized above) and school-level impact estimates. By including multiple features in the model simultaneously, the study research team allowed for an assessment of the relationships between each factor and the estimated

[60]However the statistical significance of this estimate should not be used as a "gateway" test of whether to attempt to predict variation in the LATE effects. This is because an omnibus test of whether estimated effects vary across schools (like the Q statistic) can have less power (sometimes far less power) than a focused test of the relationship between the effects and a specific school-level characteristic or moderator. (See Appendix C of Bloom, Raudenbush, Weiss, and Porter, under review.)

impacts while controlling for the values of other factors. Joint statistical significance of these variables was also tested to see whether these features are associated with the impact estimates as a group.

In general, a positive and statistically significant estimate for a given feature from the model implies that an RtI school with that feature (for dichotomous measures), or with a higher value of that feature (for continuous measures), tended to have less negative or more positive impacts on receiving intervention than schools without, or with a lower value of, that factor. In contrast, a negative and statistically significant estimate indicates that an RtI school with (or with a higher value of) that feature tended to have more negative or less positive impacts than schools without (or with a lower value of) it. A nonsignificant estimate indicates that the given feature is not likely to be associated with the impact estimates.

Some schools in the sample had missing values for some of the school features studied in this analysis. A "dummy variable adjustment" approach was used to address this issue. This approach set missing cases to a constant and added "missing data flags" to the analysis model. By doing so, the model controlled for the variable when its value was available and did not control for it when its value was missing. This approach allowed the sample of schools to be intact.[61]

Table 6.2 in the report provides descriptive statistics of all the school-level features used in the analysis. Appendix Table H.1 presents more detailed description of these variables, while Appendix Tables H.2 to H.4 provide the correlation coefficients across all these features. These tables show that, overall, these features are not highly correlated with each other. The only exceptions are moderate correlations between the proportion of low-income students in the school and the school's prior reading performance and Title I eligibility status.

Appendix Tables H.5 to H.8 present detailed results for this school-level correlational analysis for each of the four outcomes, respectively. The first pair of columns show results based on the bivariate regression approach. For each school feature, the estimated intercept and the estimated coefficient for the given feature, as well as their respective standard errors and p-values, are reported. The *estimated intercept* represents the average impact for schools without a certain feature (if the feature was measured by a dichotomous variable) or the average impact for schools with the mean level of the feature (if the feature was measured by a continuous variable).[62]

The second and third pairs of columns in Appendix Tables H.5 to H.8 provide detailed estimation results for the two models discussed in Chapter 6.

[61]Puma, Olsen, Bell, and Price (2009) discuss this approach in detail.
[62]All continuous measures are centered on the sample mean in this analysis.

Appendix Table H.1

Descriptive Statistics of School Features Examined in the Exploratory Analysis, by Grade

Grade/Characteristic	Mean (%)	Standard Deviation	Number of Schools
Grade 1			
RtI Reading Practices			
Grade uses a single screening test to assign students to intervention	79.0	40.91	119
Grade provides intervention to some students at all reading levels	47.7	50.18	109
Percentage of intervention groups meeting outside the core	59.0	35.38	109
Percentage of students identified for intervention	37.6	17.32	119
School Context			
School's prior Grade 3 reading performance relative to state mean[a]	3.5	14.35	117
Title I eligible school	69.7	46.13	119
School uses behavioral RtI in Grade 1	30.7	46.33	114
Student Body Composition			
English Language Learners	12.5	16.15	110
Students with an Individualized Education Program[b]	9.4	10.13	118
Male students	50.7	6.84	116
Low-income students	42.7	27.14	113
Students overage for grade[c]	5.4	5.49	103
Grade 2			
RtI Reading Practices			
Grade uses a single screening test to assign students to intervention	74.8	43.59	127
Grade provides intervention to some students at all reading levels	33.6	47.45	113
Percentage of intervention groups meeting outside the core	61.0	35.52	113
Percentage of students identified for intervention	35.6	15.66	127
School Context			
School's prior Grade 3 reading performance relative to state mean[a]	3.5	13.40	125
Title I eligible school	67.7	46.94	127
School uses behavioral RtI in Grade 1	27.5	44.84	120
Student Body Composition			
English Language Learners	9.8	13.56	118
Students with an Individualized Education Program[b]	10.3	11.03	126
Male students	50.9	5.86	124
Low-income students	39.4	24.26	120
Students overage for grade[c]	5.6	4.50	112

(continued)

Appendix Table H.1 (continued)

Grade/Characteristic	Mean (%)	Standard Deviation	Number of Schools
Grade 3			
RtI Reading Practices			
Grade uses a single screening test to assign students to intervention	79.5	40.58	112
Grade provides intervention to some students at all reading levels	30.2	46.16	96
Percentage of intervention groups meeting outside the core	60.1	36.27	96
Percentage of students identified for intervention	31.4	13.86	112
School Context			
School's prior Grade 3 reading performance relative to state mean[a]	3.8	12.98	112
Title I eligible school	65.2	47.85	112
School uses behavioral RtI in Grade 1	28.8	45.52	104
Student Body Composition			
English Language Learners	7.0	10.14	105
Students with an Individualized Education Program[b]	11.2	10.60	111
Male students	50.9	6.44	110
Low-income students	37.8	23.89	105
Students overage for grade[c]	6.5	5.75	101

SOURCES: Fall screening test information from schools in the sample; interventionist and teacher survey responses about reading groups; state achievement data downloaded from 13 state websites, links are provided in Appendix D.

NOTES: [a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and was measured as the deviation from the state mean.

[b]This classification does not distinguish between reading Individualized Education Programs (IEPs) and other IEPs.

[c]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

Appendix Table H.2

Correlations Between School-Level Features, Grade 1

School Feature	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
RtI Reading Practices												
[1] Grade used a single screening test to assign students to intervention	1.00											
[2] Grade provided intervention to some students at all reading levels	0.04	1.00										
[3] Percentage of intervention groups that met outside the core	0.13	-0.21	1.00									
[4] Percentage of students identified for intervention	-0.19	0.14	-0.02	1.00								
School Context												
[5] School's prior Grade 3 reading performance relative to state mean[a]	-0.05	-0.09	0.19	-0.15	1.00							
[6] Title I eligible school	-0.07	0.15	-0.14	0.29	-0.37	1.00						
[7] School used Behavioral RtI in Grade 1	-0.15	-0.02	0.04	0.10	0.06	-0.02	1.00					
Student Body Composition												
[8] Percentage of English Language Learners (ELL)	0.15	0.15	-0.04	0.34	-0.17	0.22	0.11	1.00				
[9] Percentage of students with IEPs[b]	0.08	0.09	-0.05	-0.06	-0.13	0.11	-0.22	-0.17	1.00			
[10] Percentage of male students	0.07	-0.04	0.00	0.02	-0.03	0.00	-0.01	-0.04	-0.00	1.00		
[11] Percentage of low-income students	-0.08	0.14	-0.12	0.28	-0.62	0.62	0.05	0.31	-0.04	-0.04	1.00	
[12] Percentage of students overage for grade	-0.04	-0.21	-0.03	-0.10	0.05	-0.06	0.05	-0.16	-0.03	0.05	0.01	1.00

SOURCES: Fall screening information from schools in the sample; school characteristics information from the 2010-2011 Common Core of Data (CCD); interventionist and teacher survey responses; state achievement data downloaded from state websites.

NOTES: The numbers of schools vary by feature and range from 95 to 119 based on data availability.

[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and was measured as the deviation from the state mean.

[b]This classification does not distinguish between reading Individualized Education Programs (IEPs) and other IEPs.

Appendix Table H.3

Correlations Between School-Level Features, Grade 2

School Feature	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
RtI Reading Practices												
[1] Grade used a single screening test to assign students to intervention	1.00											
[2] Grade provided intervention to some students at all reading levels	-0.01	1.00										
[3] Percentage of intervention groups that met outside the core	0.09	-0.12	1.00									
[4] Percentage of students identified for intervention	-0.14	0.11	0.02	1.00								
School Context												
[5] School's prior Grade 3 reading performance relative to state mean[a]	-0.16	0.09	0.20	-0.31	1.00							
[6] Title I eligible school	-0.01	0.25	-0.05	0.42	-0.35	1.00						
[7] School used Behavioral RtI in Grade 1	-0.22	-0.02	0.01	0.01	0.14	0.01	1.00					
Student Body Composition												
[8] Percentage of English Language Learners (ELL)	0.14	0.21	0.01	0.26	-0.35	0.27	0.00	1.00				
[9] Percentage of students with IEPs[b]	0.06	-0.09	0.11	-0.08	-0.05	-0.06	-0.17	-0.16	1.00			
[10] Percentage of male students	0.10	-0.07	0.09	0.01	0.04	-0.16	-0.10	-0.15	0.16	1.00		
[11] Percentage of low-income students	0.07	0.09	-0.07	0.53	-0.59	0.57	-0.03	0.29	-0.16	0.06	1.00	
[12] Percentage of students overage for grade	0.04	-0.21	-0.11	-0.10	-0.03	0.00	-0.01	-0.35	-0.05	-0.03	0.03	1.00

SOURCES: Fall screening information from schools in the sample; school characteristics information from the 2010-2011 Common Core of Data (CCD); interventionist and teacher survey responses; state achievement data downloaded from state websites.

NOTES: The numbers of schools vary by feature and range from 100 to 127 based on data availability.

[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and was measured as the deviation from the state mean.

[b]This classification does not distinguish between reading Individualized Education Programs (IEPs) and other IEPs.

The Response to Intervention (RtI) Evaluation

Appendix Table H.4

Correlations Between School-Level Features, Grade 3

School Feature	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
RtI Reading Practices												
[1] Grade used a single screening test to assign students to intervention	1.00											
[2] Grade provided intervention to some students at all reading levels	0.13	1.00										
[3] Percentage of intervention groups that met outside the core	-0.04	-0.13	1.00									
[4] Percentage of students identified for intervention	-0.06	0.05	-0.10	1.00								
School Context												
[5] School's prior Grade 3 reading performance relative to state mean[a]	-0.18	0.11	0.09	-0.39	1.00							
[6] Title I eligible school	0.05	-0.03	-0.02	0.46	-0.36	1.00						
[7] School used Behavioral RtI in Grade 1	-0.19	0.01	0.13	-0.03	0.13	-0.05	1.00					
Student Body Composition												
[8] Percentage of English Language Learners (ELL)	0.24	0.12	0.02	0.34	-0.33	0.28	0.09	1.00				
[9] Percentage of students with IEPs[b]	0.02	0.01	-0.16	0.19	-0.16	0.01	-0.19	-0.03	1.00			
[10] Percentage of male students	-0.06	0.09	0.09	-0.08	0.03	0.01	0.10	0.15	-0.15	1.00		
[11] Percentage of low-income students	0.10	-0.05	-0.04	0.60	-0.54	0.64	-0.03	0.27	-0.06	-0.01	1.00	
[12] Percentage of students overage for grade	0.09	-0.17	-0.07	0.05	-0.08	0.15	0.06	-0.27	-0.00	-0.09	0.22	1.00

SOURCES: Fall screening information from schools in the sample; school characteristics information from the 2010-2011 Common Core of Data (CCD); interventionist and teacher survey responses; state achievement data downloaded from state websites.

NOTES: The numbers of schools vary by feature and range from 87 to 112 based on data availability.

[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and was measured as the deviation from the state mean.

[b]This classification does not distinguish between reading Individualized Education Programs (IEPs) and other IEPs.

Appendix Table H.5

Association Between School Features and School-Level Impact Estimates, Grade 1 ECLS-K Reading Assessment

School Feature	Bivariate Regression: Intercept (Std. Err.)	Bivariate Regression: P-Value	Bivariate Regression: Estimate (Std. Err.)	Bivariate Regression: P-Value	Multivariate Regression 1: Estimate (Std. Err.)	Multivariate Regression 1: P-Value	Multivariate Regression 2: Estimate (Std. Err.)	Multivariate Regression 2: P-Value
Reading RtI Practices								
Single benchmark used by grade to assign students to intervention	-0.12 (0.109)	0.260	0.00 (0.001)	0.750	0.00 (0.001)	0.896	0.00 (0.001)	0.707
Grade provided intervention to some students at all reading levels	-0.18 (0.065)	0.005	0.00 (0.001)	0.574	0.00 (0.001)	0.444	0.00 (0.001)	0.782
Percentage of intervention groups that met outside the core (%)	-0.16 (0.045)	0.001	0.00 (0.001)	0.160	0.00 (0.001)	0.201	0.00 (0.001)	0.417
Percentage of students identified for intervention (%)	-0.16 (0.044)	0.000	0.00 (0.003)	0.106	0.01 (0.003)	0.031	0.00 (0.003)	0.211
School Characteristics								
School's prior Grade 3 reading performance relative to state mean[a]	-0.18 (0.046)	0.000	0.00 (0.003)	0.217	0.00 (0.004)	0.539	0.00 (0.004)	0.977
School had Title I status	-0.03 (0.078)	0.706	0.00 (0.001)	0.057	0.00 (0.001)	0.049	0.00 (0.001)	0.680
School used Behavioral RtI in Grade 1	-0.19 (0.054)	0.001	0.00 (0.001)	0.307	0.00 (0.001)	0.423	0.00 (0.001)	0.571
Student Body Composition								
Percentage of English Language Learner (ELL) students (%)	-0.15 (0.045)	0.002	0.00 (0.003)	0.362			0.00 (0.003)	0.713
Percentage of Individualized Education Program (IEP) students[b] (%)	-0.15 (0.044)	0.001	-0.01 (0.004)	0.160			0.00 (0.004)	0.247

(continued)

Appendix Table H.5 (continued)

School Feature	Bivariate Regression: Intercept (Std. Err.)	P-Value	Estimate (Std. Err.)	P-Value	Multivariate Regression 1: Estimate (Std. Err.)	P-Value	Multivariate Regression 2: Estimate (Std. Err.)	P-Value
Percentage of male students (%)	-0.13 (0.043)	0.003	-0.01 (0.006)	0.264			0.00 (0.006)	0.509
Percentage of students with low-income status (%)	-0.14 (0.044)	0.001	0.00 (0.002)	0.109			0.00 (0.003)	0.355
Percentage of students overage for grade (%)[c]	-0.11 (0.047)	0.017	-0.01 (0.008)	0.444			0.00 (0.009)	0.706
Intercept					-0.11 (0.154)	0.457	-0.17 (0.163)	0.297
Joint Significance Test					F-Stat 1.845	P-Value 0.086	F-Stat 0.749	P-Value 0.701

SOURCES: Study-administered ECLS-K Reading Assessment scores; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records; school characteristics information from the 2010-2011 Common Core of Data (CCD); interventionist and teacher survey responses; state achievement data downloaded from state websites.

NOTES: All coefficients are in effect-size units. A two-step procedure was used for the estimation. First, the fixed-effect impact for each school was estimated using a 2SLS regression of the outcome on the indicator of actual assignment to intervention interacted with school indicators, using the intended treatment status interacted with school indicators as the instrument variables. The estimated impact for each school was then used in a V-known random-effects meta-analysis model whereby school characteristics, as well as their corresponding missing indicators, were used as explanatory variables in the model. A complete description of the estimation model can be found in Appendix H.

A two-tailed t-test was applied to the estimated effect.

The number of schools is 119.

[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and was measured as the deviation from the state mean.

[b]This classification does not distinguish between reading IEPs and other IEPs.

[c]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

Appendix Table H.6

Association Between School Features and School-Level Impact Estimates, Grade 1 TOWRE2

	Bivariate Regression				Multivariate Regression 1		Multivariate Regression 2	
School Feature	Intercept (Std. Err.)	P-Value	Estimate (Std. Err.)	P-Value	Estimate (Std. Err.)	P-Value	Estimate (Std. Err.)	P-Value
Reading RtI Practices								
Single benchmark used by grade to assign students to intervention	-0.18 (0.112)	0.116	0.00 (0.001)	0.997	0.00 (0.001)	0.843	0.00 (0.001)	0.967
Grade provided intervention to some students at all reading levels	-0.21 (0.063)	0.002	0.00 (0.001)	0.768	0.00 (0.001)	0.550	0.00 (0.001)	0.861
Percentage of intervention groups that met outside the core (%)	-0.19 (0.045)	0.000	0.00 (0.001)	0.251	0.00 (0.001)	0.382	0.00 (0.001)	0.594
Percentage of students identified for intervention (%)	-0.18 (0.044)	0.000	0.00 (0.003)	0.927	0.00 (0.003)	0.698	0.00 (0.003)	0.282
School Characteristics								
School's prior Grade 3 reading performance relative to state mean[a]	-0.20 (0.045)	0.000	0.01 (0.003)	0.121	0.00 (0.004)	0.285	0.00 (0.004)	0.695
School had Title I status	-0.06 (0.079)	0.469	0.00 (0.001)	0.074	0.00 (0.001)	0.231	0.00 (0.001)	0.615
School used Behavioral RtI in Grade 1	-0.21 (0.054)	0.000	0.00 (0.001)	0.352	0.00 (0.001)	0.279	0.00 (0.001)	0.621
Student Body Composition								
Percentage of English Language Learner (ELL) students (%)	-0.18 (0.044)	0.000	0.01 (0.003)	0.030			0.01 (0.003)	0.012
Percentage of Individualized Education Program (IEP) students[b] (%)	-0.18 (0.044)	0.000	-0.01 (0.004)	0.127			-0.01 (0.004)	0.205

(continued)

Appendix Table H.6 (continued)

School Feature	Bivariate Regression Intercept (Std. Err.)	Bivariate Regression P-Value	Bivariate Regression Estimate (Std. Err.)	Bivariate Regression P-Value	Multivariate Regression 1 Estimate (Std. Err.)	Multivariate Regression 1 P-Value	Multivariate Regression 2 Estimate (Std. Err.)	Multivariate Regression 2 P-Value
Percentage of male students (%)	-0.15 (0.043)	0.000	-0.01 (0.006)	0.066			-0.01 (0.006)	0.178
Percentage of students with low-income status (%)	-0.18 (0.045)	0.000	0.00 (0.002)	0.144			0.00 (0.003)	0.579
Percentage of students overage for grade (%)[c]	-0.12 (0.046)	0.008	0.00 (0.008)	0.777			0.00 (0.008)	0.997
Intercept					-0.21 (0.161)	0.193	-0.14 (0.160)	0.378
Joint Significance Test					F-Stat	P-Value	F-Stat	P-Value
					1.022	0.420	1.283	0.240

SOURCES: Study-administered TOWRE2 test scores; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records; school characteristics information from the 2010-2011 Common Core of Data (CCD); interventionist and teacher survey responses; state achievement data downloaded from state websites.

NOTES: All coefficients are in effect-size units. A two-step procedure was used for the estimation. First, the fixed-effect impact for each school was estimated using a 2SLS regression of the outcome on the indicator of actual assignment to intervention interacted with school indicators, using the intended treatment status interacted with school indicators as the instrument variables. The estimated impact for each school was then used in a V-known random-effects meta-analysis model whereby school characteristics, as well as their corresponding missing indicators, were used as explanatory variables in the model. A complete description of the estimation model can be found in Appendix H.

A two-tailed t-test was applied to the estimated effect.

The number of schools is 119.

[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and was measured as the deviation from the state mean.

[b]This classification does not distinguish between reading IEPs and other IEPs.

[c]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

The Response to Intervention (RtI) Evaluation

Appendix Table H.7

Association Between School Features and School-Level Impact Estimates, Grade 2

School Feature	Bivariate Regression Intercept (Std. Err.)	Bivariate Regression P-Value	Bivariate Regression Estimate (Std. Err.)	Bivariate Regression P-Value	Multivariate Regression 1 Estimate (Std. Err.)	Multivariate Regression 1 P-Value	Multivariate Regression 2 Estimate (Std. Err.)	Multivariate Regression 2 P-Value
Reading RtI Practices								
Single benchmark used by grade to assign students to intervention	0.14 (0.076)	0.069	0.00 (0.001)	0.334	0.00 (0.001)	0.666	0.00 (0.001)	0.441
Grade provided intervention to some students at all reading levels	0.02 (0.045)	0.633	0.00 (0.001)	0.039	0.00 (0.001)	0.096	0.00 (0.001)	0.236
Percentage of intervention groups that met outside the core (%)	0.08 (0.037)	0.035	0.00 (0.001)	0.906	0.00 (0.001)	0.877	0.00 (0.001)	0.710
Percentage of students identified for intervention (%)	0.07 (0.034)	0.038	0.01 (0.002)	0.025	0.01 (0.003)	0.012	0.01 (0.003)	0.008
School Characteristics								
School's prior Grade 3 reading performance relative to state mean[a]	0.06 (0.036)	0.109	0.00 (0.003)	0.323	0.00 (0.003)	0.227	0.00 (0.004)	0.934
School had Title I status	0.02 (0.065)	0.733	0.00 (0.001)	0.350	0.00 (0.001)	0.762	0.00 (0.001)	0.991
School used Behavioral RtI in Grade 1	0.06 (0.042)	0.136	0.00 (0.001)	0.441	0.00 (0.001)	0.427	0.00 (0.001)	0.447
Student Body Composition								
Percentage of English Language Learner (ELL) students (%)	0.08 (0.035)	0.031	0.00 (0.003)	0.142			0.00 (0.003)	0.817
Percentage of Individualized Education Program (IEP) students[b] (%)	0.07 (0.035)	0.042	0.00 (0.003)	0.589			0.00 (0.003)	0.976

(continued)

Appendix Table H.7 (continued)

School Feature	Bivariate Regression: Intercept (Std. Err.)	P-Value	Estimate (Std. Err.)	P-Value	Multivariate Regression 1: Estimate (Std. Err.)	P-Value	Multivariate Regression 2: Estimate (Std. Err.)	P-Value
Percentage of male students (%)	0.07 (0.035)	0.050	-0.01 (0.006)	0.032			-0.01 (0.007)	0.040
Percentage of students with low-income status (%)	0.08 (0.035)	0.033	0.00 (0.001)	0.464			0.00 (0.002)	0.293
Percentage of students overage for grade (%)[c]	0.08 (0.037)	0.041	-0.02 (0.008)	0.027			-0.01 (0.010)	0.144
Intercept					-0.02 (0.117)	0.835	-0.03 (0.123)	0.789
Joint Significance Test					F-Stat 1.782	P-Value 0.097	F-Stat 1.760	P-Value 0.064

SOURCES: Study-administered TOWRE2 test scores; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records; school characteristics information from the 2010-2011 Common Core of Data (CCD); interventionist and teacher survey responses; state achievement data downloaded from state websites.

NOTES: All coefficients are in effect-size units. A two-step procedure was used for the estimation. First, the fixed-effect impact for each school was estimated using a 2SLS regression of the outcome on the indicator of actual assignment to intervention interacted with school indicators, using the intended treatment status interacted with school indicators as the instrument variables. The estimated impact for each school was then used in a V-known random-effects meta-analysis model whereby school characteristics, as well as their corresponding missing indicators, were used as explanatory variables in the model. A complete description of the estimation model can be found in Appendix H.

A two-tailed t-test was applied to the estimated effect.

The number of schools is 127.

[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and was measured as the deviation from the state mean.

[b]This classification does not distinguish between reading IEPs and other IEPs.

[c]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

The Response to Intervention (RtI) Evaluation

Appendix Table H.8

Association Between School Features and School-Level Impact Estimates, Grade 3

	Bivariate Regression				Multivariate Regression 1		Multivariate Regression 2	
School Feature	Intercept (Std. Err.)	P-Value	Estimate (Std. Err.)	P-Value	Estimate (Std. Err.)	P-Value	Estimate (Std. Err.)	P-Value
Reading RtI Practices								
Single benchmark used by grade to assign students to intervention	-0.07 (0.069)	0.286	0.00 (0.001)	0.442	0.00 (0.001)	0.381	0.00 (0.001)	0.800
Grade provided intervention to some students at all reading levels	-0.08 (0.039)	0.043	0.00 (0.001)	0.069	0.00 (0.001)	0.236	0.00 (0.001)	0.288
Percentage of intervention groups that met outside the core (%)	-0.04 (0.033)	0.226	0.00 (0.001)	0.664	0.00 (0.001)	0.912	0.00 (0.001)	0.802
Percentage of students identified for intervention (%)	-0.03 (0.030)	0.396	0.00 (0.002)	0.086	0.00 (0.003)	0.168	0.00 (0.003)	0.369
School Characteristics								
School's prior Grade 3 reading performance relative to state mean[a]	-0.03 (0.032)	0.295	0.00 (0.002)	0.415	0.00 (0.003)	0.069	0.01 (0.003)	0.030
School had Title I status	-0.07 (0.052)	0.151	0.00 (0.001)	0.249	0.00 (0.001)	0.206	0.00 (0.001)	0.207
School used Behavioral RtI in Grade 1	-0.04 (0.038)	0.319	0.00 (0.001)	0.747	0.00 (0.001)	0.801	0.00 (0.001)	0.674
Student Body Composition								
Percentage of English Language Learner (ELL) students (%)	-0.03 (0.031)	0.400	0.01 (0.003)	0.041			0.01 (0.004)	0.040
Percentage of Individualized Education Program (IEP) students[b] (%)	-0.02 (0.031)	0.498	0.00 (0.003)	0.848			0.00 (0.004)	0.810

(continued)

Appendix Table H.8 (continued)

School Feature	Bivariate Regression Intercept (Std. Err.)	Bivariate Regression P-Value	Bivariate Regression Estimate (Std. Err.)	Bivariate Regression P-Value	Multivariate Regression 1 Estimate (Std. Err.)	Multivariate Regression 1 P-Value	Multivariate Regression 2 Estimate (Std. Err.)	Multivariate Regression 2 P-Value
Percentage of male students (%)	-0.02 (0.031)	0.496	0.00 (0.005)	0.809			-0.01 (0.005)	0.205
Percentage of students with low-income status (%)	-0.02 (0.032)	0.465	0.00 (0.001)	0.768			0.00 (0.002)	0.718
Percentage of students overage for grade[c] (%)	-0.02 (0.032)	0.499	-0.01 (0.006)	0.070			-0.01 (0.007)	0.369
Intercept					-0.21 (0.094)	0.029	-0.17 (0.101)	0.100
Joint Significance Test					F-Stat	P-Value	F-Stat	P-Value
					1.435	0.199	1.719	0.075

SOURCES: State reading achievement scores from district records; fall screening scores and student tier placement data from schools in the sample; student demographic data from district records; school characteristics information from the 2010-2011 Common Core of Data (CCD); interventionist and teacher survey responses; state achievement data downloaded from state websites.

NOTES: All coefficients are in effect-size units. A two-step procedure was used for the estimation. First, the fixed-effect impact for each school was estimated using a 2SLS regression of the outcome on the indicator of actual assignment to intervention interacted with school indicators, using the intended treatment status interacted with school indicators as the instrument variables. The estimated impact for each school was then used in a V-known random-effects meta-analysis model whereby school characteristics, as well as their corresponding missing indicators, were used as explanatory variables in the model. A complete description of the estimation model can be found in Appendix H.

A two-tailed t-test was applied to the estimated effect.

The number of schools is 112.

[a]"School's prior Grade 3 reading performance" refers to the percentage of students at or above reading proficiency on state tests and was measured as the deviation from the state mean.

[b]This classification does not distinguish between reading IEPs and other IEPs.

[c]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 are classified as overage.

Student-Level Correlational Analysis

The basic approach used to explore the relationship between the estimated impact of actual assignment to Tier 2 or Tier 3 intervention and students' demographic characteristics is an expanded 2SLS model with interactions between the actual assignment status and student characteristics. Specifically, the following regression models were used.

First-Stage Equations (J+p equations, one for each term that has $\boldsymbol{T_{actual,ij}}$)

$$S_{i1}T_{actual,ij} = \sum_j \alpha_{1j}S_{ij} + \sum \gamma_j S_{ij}T_{intended,ij} + \sum \gamma_p X_{pij}T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij}T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \quad (5.1s)$$

$$S_{i2}T_{actual,ij} = \sum_j \alpha_{1j}S_{ij} + \sum \gamma_j S_{ij}T_{intended,ij} + \sum \gamma_p X_{pij}T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij}T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \quad (5.2s)$$

$$S_{iJ}T_{actual,ij} = \sum_j \alpha_{1j}S_{ij} + \sum \gamma_j S_{ij}T_{intended,ij} + \sum \gamma_p X_{pij}T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij}T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \quad (5.Js)$$

$$X_{1ij}T_{actual,ij} = \sum_j \alpha_{1j}S_{ij} + \sum \gamma_j S_{ij}T_{intended,ij} + \sum \gamma_p X_{pij}T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij}T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \quad (5.1x)$$

…

$$X_{pij}T_{actual,ij} = \sum_j \alpha_{1j}S_{ij} + \sum \gamma_j S_{ij}T_{intended,ij} + \sum \gamma_p X_{pij}T_{intended,ij} + \alpha_2 R_{ij} + \alpha_3 R_{ij}T_{intended,ij} + \sum_k \rho_k X_{kij} + \epsilon_{ij} \quad (5.px)$$

Second-Stage Equation

$$Y_{ij} = \sum_j \theta_{1j}S_{ij} + \sum_j \delta_j S_{ij}\widehat{T_{actual,ij}} + \sum_p \lambda_p X_{pij}\widehat{T_{actual,ij}} + \theta_2 R_{ij} + \theta_3 R_{ij}T_{intended,ij} + \sum_k \varphi_k X_{kij} + \mu_{ij} \quad (6)$$

where:

k = the number of student demographic covariates

p = the number of student characteristics under investigation

$k \geq p$

The joint significance of λ_p was also tested to inform readers whether the estimated impacts of reading intervention on students marginally below grade level varies significantly by these student characteristics. Missing values in the student characteristics were dealt with using the dummy variables for observations with missing values and imputation of missing values to equal the mean.

Similar to the school-level analyses, these student characteristics were examined separately and with each other as a group. In general, a positive and statistically significant estimate for a given characteristic from the model implies that students with that characteristic tended, on average, to have less negative or more positive impacts on receiving intervention than students without that characteristic. In contrast, a negative and statistically significant estimate indicates that students with that feature tended, on average, to have more negative or less positive impacts than students without it. A nonsignificant estimate indicates that the given characteristic is not likely to be associated with the magnitude of the impact estimate.

Appendix Table H.9 provides the correlation coefficients among the student characteristics studied in this analysis. Appendix Table H.10 provides detailed estimation results for each outcome. The first five columns in the table (Models 1 to 5) present the results for each characteristic when the other characteristics are not included in the model. The rightmost pair of columns report results based on a model that includes all five characteristics jointly (Model 6). The findings based on these two different approaches are fairly consistent.

The Response to Intervention (RtI) Evaluation

Appendix Table H.9

Correlations Between Student Characteristics Within Optimal Bandwidth, by Grade and Outcome

Student Characteristic	[1]	[2]	[3]	[4]	[5]
Grade 1 ECLS-K Reading Assessment					
[1] English Language Learners	1.00				
[2] Students with IEPs[a]	-0.03	1.00			
[3] Male students	0.04	0.09	1.00		
[4] Low-income students	0.21	0.01	-0.01	1.00	
[5] Overage for grade[b]	0.01	0.03	0.07	-0.01	1.00
Grade 1 TOWRE2					
[1] English Language Learners	1.00				
[2] Students with IEPs[a]	-0.03	1.00			
[3] Male students	0.04	0.09	1.00		
[4] Low-income students	0.21	0.01	-0.01	1.00	
[5] Overage for grade[b]	-0.00	0.04	0.08	-0.02	1.00
Grade 2					
[1] English Language Learners	1.00				
[2] Students with IEPs[a]	-0.00	1.00			
[3] Male students	0.04	0.10	1.00		
[4] Low-income students	0.16	-0.01	-0.02	1.00	
[5] Overage for grade[b]	-0.04	0.06	0.08	0.04	1.00
Grade 3					
[1] English Language Learners	1.00				
[2] Students with IEPs[a]	0.01	1.00			
[3] Male students	0.02	0.09	1.00		
[4] Low-income students	0.14	-0.01	-0.01	1.00	
[5] Overage for grade[b]	-0.01	0.09	0.07	0.02	1.00

SOURCES: Fall screening scores from schools in the sample; student demographic data from district records.

NOTES: The optimal bandwidth defines the sample of students to be used in the impact regression to best balance the trade-off between bias and precision. The optimal bandwidth for each grade and outcome measure was pre-selected using the algorithm described in Imbens and Kalyanaraman (2012). See Appendix E for more details.

The numbers of schools are Grade 1 = 119, Grade 2 = 127, and Grade 3 = 112. The numbers of students are 6,236 for Grade 1 ECLS-K reading assessment; 5,398 for Grade 1 TOWRE2; 4,301 for Grade 2; 6,549 for Grade 3. Test scores for Grade 1 and 2 were standardized and centered on zero. Test scores for Grade 3 were standardized relative to state means.

[a]"IEP" represents Individualized Education Programs. This classification does not distinguish between reading IEPs and other IEPs.

[b]Overage for grade was calculated based on student age as of August 15, 2011. Grade 1 students over the age of 7, Grade 2 students over the age of 8, and Grade 3 students over the age of 9 were classified as overage.

Appendix Table H.10

Association Between Individual Student Characteristics and the Impact of Actual Assignment to Tier 2 or Tier 3 Intervention Services, by Outcome

Student Characteristic	Model Specification: Model 1	Model 2	Model 3	Model 4	Model 5	Model 6: Estimate	Model 6: P-Value
Grade 1 ECLS-K Reading Assessment							
Student is male	-0.035					-0.012	0.811
	(0.051)					(0.052)	
Student had low-income status		-0.089				-0.082	0.209
		(0.066)				(0.065)	
Student had English Language Learner (ELL) status			0.011			0.022	0.841
			(0.108)			(0.109)	
Student had an Individualized Education Program (IEP)[a]				-0.337		-0.323	0.006
				(0.115)		(0.117)	
Student was overage for grade[b]					-0.224	-0.207	0.018
					(0.088)	(0.087)	
Significance Test[c]	P-Value	P-Value	P-Value	P-Value	P-Value	F-Statistic	P-Value
	0.491	0.180	0.922	0.003	0.011	3.270	0.006
Grade 1 TOWRE2							
Student is male	-0.041					-0.034	0.569
	(0.059)					(0.059)	
Student had low-income status		-0.085				-0.084	0.278
		(0.078)				(0.077)	
Student had English Language Learner (ELL) status			-0.002			0.017	0.875
			(0.111)			(0.110)	
Student had an Individualized Education Program (IEP)[a]				-0.095		-0.078	0.458
				(0.105)		(0.105)	
Student was overage for grade[b]					-0.163	-0.153	0.115
					(0.096)	(0.097)	
Significance Test[c]	P-Value	P-Value	P-Value	P-Value	P-Value	F-Statistic	P-Value
	0.487	0.276	0.988	0.369	0.091	0.971	0.434

(continued)

Appendix Table H.10 (continued)

Student Characteristic	Model Specification Model 1	Model 2	Model 3	Model 4	Model 5	Model 6 Estimate	P-Value
Grade 2 TOWRE2							
Student is male	-0.010					-0.006	0.918
	(0.054)					(0.054)	
Student had low-income status		-0.081				-0.083	0.189
		(0.063)				(0.063)	
Student had English Language Learner (ELL) status			0.130			0.147	0.233
			(0.126)			(0.123)	
Student had an Individualized Education Program (IEP)[a]				-0.160		-0.152	0.217
				(0.121)		(0.123)	
Student was overage for grade[b]					-0.091	-0.076	0.393
					(0.089)	(0.089)	
Significance Test[c]	P-Value	P-Value	P-Value	P-Value	P-Value	F-Statistic	P-Value
	0.850	0.201	0.301	0.187	0.303	1.213	0.301
Grade 3 State Reading Achievement Test							
Student is male	-0.075					-0.065	0.148
	(0.045)					(0.045)	
Student had low-income status		-0.007				-0.016	0.756
		(0.052)				(0.053)	
Student had English Language Learner (ELL) status			0.196			0.192	0.048
			(0.097)			(0.097)	
Student had an Individualized Education Program (IEP)[a]				-0.203		-0.185	0.024
				(0.081)		(0.082)	
Student was overage for grade[b]					-0.109	-0.089	0.181
					(0.066)	(0.067)	
Significance Test[c]	P-Value	P-Value	P-Value	P-Value	P-Value	F-Statistic	P-Value
	0.093	0.887	0.044	0.012	0.098	2.749	0.018

SOURCES: Study-administered test scores: ECLS-K Reading Assessment for Grade 1, TOWRE2 for Grades 1 and 2; state reading achievement scores from district records for Grade 3; fall screening scores and student tier placement data from sample schools; student demographic data from district records.

NOTES: All coefficients are in effect-size units. The model used a 2SLS regression of the outcome on the following variables: interactions between student's actual assignment to Tier 2 or Tier 3 intervention and school indicators, interactions between student's actual assignment and the four listed student characteristics, using as instrumental variables the interactions between student's intended assignment to Tier 2 or Tier 3 intervention and school indicators, and the interactions between student's intended assignment and the four listed student characteristics. School indicators, rating variable and its interaction with student’s intended assignment, and other student level covariates are also included in the model. A full description is found in Appendix H.

(continued)

Appendix Table H.10 (continued)

The numbers of students are 6,236 for Grade 1 ECLS-K; 5,398 for Grade 1 TOWRE2; 4,301 for Grade 2; 6,549 for Grade 3.

[a]This classification does not distinguish between reading IEPs and other IEPs.

[b]Overage for grade was calculated based on student age as of August 15, 2011: Grade 1 over the age of 7, Grade 2 over the age of 8, and Grade 3 over the age of 9 are classified as overage.

[c]For Models 1-5, this is a two-tailed t-test. For Model 6, this is a joint F-test.

This page has been left blank for double-sided copying.