Using Student Data to Improve Teaching and Learning

Findings from an Evaluation of the Formative Assessments of Student Thinking in Reading (FAST-R) Program in Boston Elementary Schools

Janet C. Quint
Susan Sepanik
Janell K. Smith

December 2008

mdrc
BUILDING KNOWLEDGE
TO IMPROVE SOCIAL POLICY

Funding for this report was provided by the William and Flora Hewlett Foundation.

Dissemination of MDRC publications is supported by the following funders that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Ambrose Monell Foundation, Bristol-Myers Squibb Foundation, The Kresge Foundation, Sandler Foundation, and The Starr Foundation. MDRC's dissemination of its education-related work is supported by the Bill & Melinda Gates Foundation, Carnegie Corporation of New York, and Citi Foundation. In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

For information about MDRC and copies of our publications, see our Web site: www.mdrc.org.

Overview

Formative assessments — assessments that measure what students do and do not know, so that teachers can modify their instruction accordingly — have been widely hailed as a potential vehicle for improving student achievement. Yet little solid research evidence exists about their effectiveness, especially in reform-rich school districts. This study examines the effects of the Formative Assessments of Student Thinking in Reading (FAST-R) initiative in the Boston Public Schools system (BPS), where the use of data to improve instruction is a general priority of the school district. The study looks at changes in reading scores over time at 21 BPS schools that operated FAST-R during the 2005-2006 and 2006-2007 school years and changes at a group of comparison schools serving demographically similar students during the same period.

The Boston Plan for Excellence (BPE), a not-for-profit school reform organization, created and operates FAST-R. The study intervention involved administering a series of short student assessments whose items resemble those in the Massachusetts Comprehensive Assessment System (MCAS), the state's "high-stakes" assessment used to measure the performance of both schools and students, and focus on students' reading comprehension skills. BPE staff members compiled the results of the assessments into easy-to-use reports that contained information about each student. Then a BPE instructional data coach met with the teachers at each school to review the reports and to suggest how teachers could respond to students' learning needs. (One BPE coach served most of the schools, and another BPE coach served the balance.)

The MDRC evaluation includes process and impact analyses. The process analysis found that teachers at the FAST-R schools who took a survey administered as part of the study reported that the professional development they received from the BPE FAST-R coaches was helpful and contributed to their understanding of data and their ability to work with students. At the same time, while the intervention was implemented as intended (it was meant to be flexible and to provide as much or as little coaching to individual schools as administrators and teachers sought), it was not very intensive; the majority of survey respondents spent only one to five hours with the FAST-R data coach during the 2006-2007 school year. Moreover, comparison-school teachers who took the survey reported receiving at least as much professional development as their FAST-R counterparts, were as likely to find it useful, and spent as much or more time analyzing data, including data from other (non-FAST-R) formative assessments.

The impact analysis examines the effects of FAST-R on the reading test scores of third- and fourth-graders. FAST-R's impacts on student achievement — that is, the difference that FAST-R made over and above what was going on in the comparison schools — are generally positive but not statistically significant, as measured by MCAS reading scores. In other words, these differences could have arisen by chance. Effects on another measure of student reading, the Stanford Achievement Test, are more mixed but are also not statistically significant.

While FAST-R schools put in place a particular model of data utilization, other BPS schools were pursuing similar goals, and this fact, along with the intervention's lack of intensity, may have undercut the likelihood that it would generate substantial and statistically significant impacts in this evaluation. Thus, this single study in a single district is not the last word on the potential of FAST-R. Much remains to be discovered about how teachers can best learn to use data to improve their instruction and boost the achievement of their students.

Contents

List of Tables, Figures, and Boxes

Table

Figure

Box

Preface

When the No Child Left Behind Act was passed in 2001, many school districts were already looking at and testing the effectiveness of various instructional strategies for building reading skills, particularly for younger children. Among those strategies, the use of student data to inform teaching — and, in turn, to improve learning — has often been identified as a potentially powerful educational tool. One important means of generating such data is the use of "formative assessments" — tests or activities that measure student learning and provide feedback to teachers that they can use to adapt their teaching practices to meet student needs. Yet formative assessments have rarely been scrutinized to determine their effectiveness. This report describes an initiative that was designed to enhance the usefulness of formative assessments and presents an evaluation of its impact on students' reading skills in Boston public elementary schools.

The Formative Assessments of Student Thinking in Reading (FAST-R) program, as the initiative is called, was created and is operated by the Boston Plan for Excellence (BPE), a not-for-profit school reform organization. BPE created a series of formative assessments that teachers could administer to their students at various points during the school year, generating data that teachers would then use to inform their reading instruction. Recognizing, however, that merely generating data is not enough if teachers don't know how to interpret and use those data, BPE included a second component in the FAST-R program — professional development provided by an "instructional data coach," who helped teachers understand the data. Encouraged by positive feedback from a qualitative study of FAST-R's early operations, BPE contracted with MDRC to evaluate the initiative's impacts in 21 schools in the Boston Public Schools system. Conducted with support from the William and Flora Hewlett Foundation, the evaluation found that while teachers at the FAST-R schools reported, by and large, that the program had been helpful, no impacts on students' reading test scores were registered in those schools compared with a group of schools serving similar students that did not implement FAST-R. It should be noted, however, that the program treatment may not have been intensive enough to have an impact, particularly in an environment where many teachers — including those in the comparison schools — were already using data and participating in professional development aimed at improving reading instruction.

The FAST-R evaluation was one test of formative assessments in one school district, and, as such, its findings cannot be considered a definitive statement on the effectiveness of training teachers in the use of data to guide reading instruction. Given that the use of data to improve learning is such an important idea, different tests of this strategy in different settings are very much in order. It is MDRC's hope that this evaluation will help to shape future research and that such research will yield findings that stimulate educators' ongoing efforts to help children boost their reading skills — whether by proceeding along a proven, evidence-based path or by recognizing the need to forge a new one.

Gordon Berlin
President

Acknowledgments

This report is a product of collaboration among MDRC, the Boston Plan for Excellence (BPE), and the funding organization, the William and Flora Hewlett Foundation.

The Formative Assessments of Student Thinking in Reading (FAST-R) program was created by BPE, and we are grateful for the opportunity to conduct and report on the first impact analysis of the intervention. The contributions of BPE staff members Jennifer Amigone, Lisa Lineweaver, and Courtney Williams have been extremely useful in helping us understand the initiatives of the Boston Public Schools district and the FAST-R program.

We would also like to acknowledge Survey Research Management for coordinating and collecting teacher surveys in FAST-R schools and their comparison schools.

Finally, we would like to express gratitude to our MDRC colleagues for the insights and guidance they provided on the study's technical research design and for their meticulous feedback while reviewing drafts of the report. We give many thanks to James Kemple, Howard Bloom, Fred Doolittle, Corinne Herlihy, Pei Zhu, Marie-Andrée Somers, Shelley Rappaport, John Hutchins, Margaret Bald, and Laura Sztejnberg (a former colleague). We would also like to thank MDRC staff members who administered the principal survey over the phone to Boston principals. Alma Moedano helped to coordinate the production of the report and fact-checked text, tables, and figures. Vivian Mateo provided assistance with formatting tables and figures. Mario Flecha provided production assistance. Alice Tufel edited the report, Bob Weber proofread it, and Stephanie Cowell and David Sobel prepared it for publication.

The Authors

Executive Summary

Formative assessments — that is, assessments administered in order to measure what students do and do not know, so that teachers can modify their instruction accordingly — have been widely hailed as a potential vehicle for improving student achievement. To date, however, little rigorous research has been done on the impacts of formative assessment and its resulting data-driven instruction, particularly in reform-rich urban school districts. This study, funded by the William and Flora Hewlett Foundation, is a step toward filling that knowledge gap. It examines the effects of the Formative Assessments of Student Thinking in Reading (FAST-R) initiative as it operated in 21 schools in the Boston Public Schools system (BPS) during the 2005-2006 and 2006-2007 school years.

The Boston Plan for Excellence (BPE) — a not-for-profit organization that works with the BPS central office and individual schools to design, pilot-test, and implement new reforms aimed at improving teaching and learning — created and operates FAST-R. The intervention (as operated during the study period) involved a series of short assessments whose items resemble the multiple-choice questions contained in the Massachusetts Comprehensive Assessment System (MCAS), the state's "high-stakes" assessment used to measure both student and school performance, and administered annually to students in grades 3 through 8 and grade 10. FAST-R questions, like those on the MCAS, focus on two essential reading skills: the ability to *find evidence* in the text that supports an explicit point and the ability to *make inferences* from the available information to support a valid interpretation.

Schools are free to choose those FAST-R assessments that comport best with their lesson plans and to administer them on a schedule that best suits their needs. After students have taken an assessment, BPE staff score their answer sheets and compile the results in reports that are designed to be easy to use and that contain information about how each student, as well as groups of students, scored on each assessment item. The reports are meant to help teachers understand not only how many students came up with the right answers but also what mistakes in reading or reasoning led students to come up with the wrong ones. Then a BPE instructional data coach meets with the teachers at each school to review the reports, help them learn how to interpret the data, and suggest how they can respond to students' learning needs. During the period under study, one BPE coach worked with most of the FAST-R elementary schools. (Another coach worked with the rest of the schools.)

The use of data to inform instruction is a general priority in the Boston Public Schools district, not just in those schools that have elected to implement FAST-R. In fact, BPS has included the use of student data to identify student needs, improve instruction, and assess progress as one of "The Six Essentials" (guiding principles) of its "Whole-School Improvement" model, which is used throughout the Boston Public Schools system.[1] (Other "essentials" include a focus on literacy and mathematics instruction and professional development for principals and teachers that is focused on improving instructional skills.) FAST-R was intended to complement the district's own professional development supports, and, while FAST-R schools used their own unique assessment tool and put in place a particular model of data utilization, other Boston public schools were pursuing similar goals, sometimes through broadly similar means. For instance, teachers at non-FAST-R schools had access to other formative assessments, worked with literacy coaches to improve their instructional techniques, and could participate in various other types of professional development. This fact needs to be borne in mind when considering the evaluation findings, since outcomes at the FAST-R schools are compared with outcomes at these other schools.

The Evaluation Design

The evaluation includes a process analysis and an impact analysis. The process analysis was intended to provide information about how teachers in the FAST-R schools used what they had learned and, more generally, about how they regarded the initiative. Equally if not more important, it aimed to shed light on the professional development efforts that took place both in the FAST-R schools and in a group of comparison schools that served students who were similar to those in the FAST-R schools but did not put the initiative into place. In this way, the process analysis helps to establish a context for the impact analysis findings.

Surveys administered in the spring of 2007 to principals and to third- and fourth-grade teachers at both sets of schools are the key source of data for the process analysis. Unfortunately, during this period, the Boston Teachers Union was negotiating a new contract. It seems likely that many teachers were unwilling to undertake any noninstructional activities until contract issues were resolved; in any case, the response rates for the surveys were low (about 54 percent for teachers in both the FAST-R and the comparison schools, and even lower for principals). Because of the low response rates, and because some schools participating in the impact analysis did not supply any survey responses at all, the survey findings can be viewed only as suggestive rather than definitive in illuminating the similarities and differences between program and comparison schools.

The impact analysis uses a comparative interrupted time series design to examine the effects of FAST-R on the reading test scores of third- and/or fourth-graders in 21 schools that implemented the program during the 2005-2006 and 2006-2007 school years. Attention centers on these grade levels because they are the earliest grades in which the MCAS is routinely

[1]As of fall 2008, BPS was using "The Seven Essentials" in its Whole-School Improvement Model.

administered and because in these grades students face new challenges as reading comprehension replaces "decoding" (sounding out words) as the major focus of instruction. The FAST-R schools selected for the study were ones identified by BPE as having actively implemented the intervention during the study period.

The impact analysis includes two comparisons. The first comparison is between test score outcomes at the 21 FAST-R schools before and after the initiative was put in place. To draw this comparison, scores on each outcome over a five-year pre-intervention "baseline" period were used to create a trend line and to project that trend line into the post-implementation period. The difference between the actual and projected scores during this post-implementation period represents the "deviation from trend" for the FAST-R schools.

If one were to look at the program schools alone, however, it would be impossible to determine how much of the observed change was attributable to FAST-R and how much that change reflected other developments throughout BPS. A second comparison, therefore, involves measuring changes in outcomes over time at a set of BPS schools that did not implement FAST-R but whose student populations resemble those at the FAST-R schools in terms of demographic characteristics and prior achievement. Thirty-six schools were selected as comparison schools. As with the FAST-R schools, baseline scores were used to project a trend line for the comparison schools, and the difference between projected and actual scores represents the deviation from trend for these schools. The *impact* of FAST-R is the difference between the deviation from the trend for the program schools and the deviation from the trend for the comparison schools.

The impact analysis relies on individual student records obtained from BPS. Students' general reading achievement is measured using three outcomes from two standardized tests: the average reading comprehension total score on the MCAS; the percentage of students scoring at or above the "proficient" level on the MCAS; and the average reading total score on the Stanford Achievement Test, version 9 (SAT-9). (Because the test items in the FAST-R assessment are so similar to those on the MCAS, use of a second test provides reassurance that any positive impacts on the MCAS are not simply the result of increased student familiarity with that assessment.) In addition to these general measures, the analysis measures how well students in program and comparison schools performed in answering MCAS questions designed to measure students' specific reading skills in terms of finding evidence in and making inferences from text.

Findings on FAST-R and Other Professional Development Activities in the Boston Public Schools System

- **Teachers at the FAST-R schools who responded to the survey reported that the professional development they received from the BPE FAST-R coaches was helpful and contributed to their understanding of data and their ability to work with students.**

While, as noted above, the survey respondents were not necessarily typical of all teachers at the FAST-R schools, their views of FAST-R were notably positive. They reported that the initiative had helped them to understand students' thinking and to use student data in reflecting on their instructional practices.

- **The majority of FAST-R teachers reported spending a limited amount of time with the BPE coach.**

Just over 60 percent of the FAST-R survey respondents reported spending one to five hours with their FAST-R data coach over the course of the 2006-2007 school year. Only 13 percent reported spending 11 or more hours with the coach. This is not surprising, given the many demands on the time of the FAST-R coach who worked with elementary schools.

- **Compared with their counterparts at the FAST-R schools, teachers responding to the survey at the comparison schools reported participating in at least as much professional development, were as likely to find that professional development useful, and spent as much or more time analyzing student data, including data from formative assessments.**

FAST-R may have contributed to teachers' knowledge and understanding, but that contribution was not unique. Teachers at the comparison schools spent as much time as did teachers at the FAST-R schools engaging in Collaborative Coaching and Learning — the district's school-based professional development model, which emphasizes collaboration among teachers — or working with literacy coaches, as well as observing other teachers' classrooms; comparison-school teachers actually reported spending *more* time on curriculum-specific professional development. Like the FAST-R teachers, they believed that the professional development was helpful for conducting their reading classes, leading discussions, creating assignments, and placing students in groups according to reading level. They also reported spending as many hours as the FAST-R teachers in analyzing formative assessments, and they spent substantially more hours analyzing the previous year's MCAS results.

Findings on FAST-R's Impacts on Students' Reading Skills

- **FAST-R unfolded in an environment of stable or improving reading scores.**

During the five years preceding FAST-R's implementation, average reading scores on the MCAS and SAT-9 held steady for third-graders and improved for fourth-graders at both the FAST-R and the comparison schools.

- **FAST-R's impacts on student achievement — that is, the difference that FAST-R made over and above what was going on in the comparison schools — are generally positive but not statistically significant, as measured by MCAS reading scores. In other words, these differences could have arisen by chance. Effects on the SAT-9 reading scores are more mixed, but also not statistically significant.**

The achievement gains of students at the FAST-R schools during the follow-up period were somewhat larger than those of students at the comparison schools, as measured by the MCAS, but the differences are not statistically significant. This was the case for both third- and fourth-graders. Average scores for the two groups did not differ, nor did the percentage of students whose scores placed them in the "proficient" or "advanced" categories. FAST-R was associated with both positive and negative impacts on SAT-9 scores, but none of these impacts is statistically significant.

- **The initiative did not have an impact on students' ability to find evidence in or make inferences from text, as measured on the MCAS.**

Strengthening students' reading comprehension skills as measured by their ability to find evidence in and make inferences from text was the key objective of FAST-R. The study specifically examined students' responses to MCAS questions that were designed to tap into these abilities. The results indicate that students at the FAST-R and comparison schools made similar progress in their ability to answer these two kinds of questions.

- **FAST-R did not produce consistent impacts for particular subgroups of students or schools.**

Although FAST-R had a positive and statistically significant effect on the percentage of fourth-grade boys who scored at the "proficient" or "advanced" level on the MCAS, the *difference* between the impacts for boys and for girls was not statistically significant. This means the finding must be interpreted with caution. Some impacts were registered for subgroups defined by special education status, but these were inconsistent and did not tell a clear story. The researchers sought to examine the impacts for particular subgroups of schools as well as students (for example, those schools where teachers reported receiving more FAST-R coaching) but could not do so because the subgroups were too small to yield reliable results.

Interpreting the Findings

Since the FAST-R program did not, in general, demonstrate improvements over the status quo, one possible conclusion is that the intervention is no more effective at increasing student achievement than the tools that teachers in BPS are already using. But it is also possible that the specific circumstances under which it operated in Boston undercut the likelihood that the initiative would generate significant impacts.

In order for an initiative to register positive impacts, one or both of two conditions must be in place: The initiative must be implemented in a reasonably strong way, and/or it must represent a distinct contrast from the services that are available to individuals who do not participate in the initiative.

FAST-R was implemented as intended in the study schools, but it was not very intensive. The intervention was designed to be flexible and to provide as much or as little coaching to individual schools as administrators and teachers sought. One consequence is that many teachers reported getting only a few hours of coaching over the course of the year. Although, as noted above, teachers valued the coaching they did receive, the amount may simply have been insufficient to affect teaching and learning more than would otherwise have occurred. Furthermore, FAST-R was implemented in an environment where a great deal *was* otherwise occurring. Teachers at comparison schools who responded to the survey indicated that they had received at least as much professional development and spent as much time analyzing student data as their counterparts at the FAST-R schools. Given this confluence of circumstances, it is perhaps not surprising that FAST-R did not have an effect on student achievement above and beyond what was happening in similar Boston schools.

At the same time, it is possible that FAST-R was an *efficient* way of providing professional development to teachers (assuming that such professional development makes a positive difference). Those FAST-R teachers who responded to the survey reported consistently that they had spent less time on various kinds of professional development than their counterparts at the comparison schools (although the differences were not statistically significant), yet their students achieved comparable gains.

However one interprets the impact findings, this single study in a single district — one with rich professional development opportunities — should not be taken as the last word on the potential of FAST-R. Much more remains to be discovered about how teachers can best learn to use student data to improve their instruction and boost students' achievement.

Chapter 1

Introduction

The No Child Left Behind Act (NCLB), passed in 2001, places an emphasis on improving the learning of all students and mandates annual testing of reading in the third through eighth grades. The NCLB reflects a growing national effort to learn about effective instructional practices for building reading skills — in all grades but especially for young children — that originated from recommendations made by the National Reading Panel.[1] Many districts and schools are using the data generated by annual state assessments that are required for accountability purposes to guide district- and school-level changes aimed at boosting achievement in reading. Moreover, a growing number of districts and schools have sought more targeted, detailed, and timely data that teachers can use to improve instruction in their classrooms. Districts are developing or purchasing formative assessments aligned with their state standards and administered several times during the school year to obtain useful information about the skills of individual students and the effectiveness of instruction.[2]

Formative assessments are tests or activities that measure learning and provide feedback that is then used to adapt teaching practices to meet student needs.[3] Learning is in large part driven by exchanges between teachers and their students, and formative assessments provide data to gauge these exchanges. According to one study, elementary school educators who use formative assessments as part of an English language arts (ELA) curriculum cite them as the main source of evidence about their students' reading achievement and progress.[4] An advantage of using formative assessments is the early diagnosis of reading difficulties, which, when identified in young students, can help to pinpoint opportunities for remediation in early grades.[5]

Despite their potential for improving students' achievements, little rigorous research has been done on the impacts of formative assessments and instruction informed by the data that such assessments generate (or "data-driven instruction"), particularly in reform-rich urban school districts. This report seeks to address this knowledge gap regarding the use of formative assessments in reading. It evaluates the Formative Assessments of Student Thinking in Reading (FAST-R) initiative, a program implemented in some 50 schools within the Boston Public Schools system (BPS) that trains teachers to use formative assessments of reading to inform their literacy instruction. The evaluation, funded by the William and Flora Hewlett Foundation,

[1]National Institute of Child Health and Human Development (2000).
[2]Murnane and Sharkey (2004).
[3]Black and Wiliam (1998).
[4]Meisels and Piker (2000).
[5]Snow, Burns, and Griffin (1998).

includes an impact analysis to examine the impact of FAST-R on student achievement in reading and a process analysis to examine professional development initiatives in FAST-R schools and in a group of schools designated as comparison schools for the impact analysis.

Previous external and internal evaluations of FAST-R have highlighted the promise of the intervention. These evaluations were qualitative studies based on data collected from surveys, site visits, classroom observations, and interviews with school personnel. In an evaluation of the FAST-R pilot program, researchers found that the initiative presented a "compelling picture of the ways in which authentic formative assessments can lead to teacher knowledge that informs instruction."[6] In a subsequent evaluation, researchers concluded that FAST-R was able to provide timely and valuable formative data in a way that teachers could use and that enabled them to better understand student thinking and adjust instruction accordingly.[7]

District Reform Efforts in Boston and the Role of FAST-R in Boston Public Schools

Over the years, BPS has created numerous strategies and initiatives to improve instruction and student academic performance in its schools.[8] In 1996, BPS implemented a reform plan called "Focus on Children" that emphasized a standards-based reform approach, including efforts to achieve high standards, focus on literacy and mathematics instruction, and provide targeted professional development for principals and teachers. In 2001, BPS implemented a second reform plan, "Focus on Children II," based on its "Whole-School Improvement" reform model.[9] The model is a collaborative process that aims to produce measurable improvement in student performance and is organized around six key guiding principles, called "The Six Essentials" (see Box 1.1), which BPS utilized during the school years covered by this study.[10]

The strong focus on high-quality instruction in BPS motivated the district to undertake efforts to enhance opportunities for data-driven instruction and learning from student work. One of those efforts is FAST-R, a joint initiative of BPS and the Boston Plan for Excellence (BPE). A not-for-profit organization that works to design, pilot, and implement district reform efforts by collaborating with the central office and with individual schools, BPE was responsible for

[6]Neufeld and Schwartz (2004).
[7]Chrismer (2005).
[8]Descriptions of BPS district reform efforts are based on information gathered from the BPS Web site (www.bostonpublicschools.org).
[9]Boston Public Schools (2008a).
[10]Boston Public Schools (2008a). As of fall 2008, BPS uses "The Seven Essentials."

Box 1.1

Boston Public Schools

"The Six Essentials for Whole-School Improvement"

- ONE: *Effective Instruction* — Focus on literacy and mathematics.
- TWO: *Student Work and Data* — Use student work and data to identify student needs, improve instruction, and assess progress.
- THREE: *Professional Development* — Focus on professional development to offer teachers and principals the skills they need to improve instruction.
- FOUR: *Shared Leadership* — Identify and replicate best practices for instruction.
- FIVE: *Resources* — Align all resources with the instructional focus.
- SIX: *Families and Communities* — Engage families, communities, and partners to support Whole-School Improvement.

creating, piloting, and disseminating FAST-R in BPS schools. The FAST-R program trains teachers to use instructional strategies based on formative assessment data that build students' skills in reading and understanding texts. The program is intended to complement the district's existing literacy professional development that teachers reported receiving in the 2006-2007 school year. Teachers in FAST-R schools are primarily given additional targeted training in learning how to analyze student assessment data to tailor their reading instruction. Box 1.2 outlines the process for using the FAST-R program in BPS schools.

FAST-R originated during the 2003-2004 school year as the formative assessments pilot and was used on a trial basis in 21 schools.[11] Beginning in the 2005-2006 school year, FAST-R underwent full-scale program implementation. All BPS schools were invited to submit proposals to participate in FAST-R. Schools were chosen by BPE based on their expressed need to learn how to better use data to guide reading instruction. In addition, central office administrators gave input about which schools were most suitable for the program. According to BPE, as of the 2007-2008 school year, 50 BPS elementary, middle, and high schools use FAST-R regularly.

[11]Neufeld and Schwartz (2004).

Box 1.2

The FAST-R Process in Boston Public Schools

- Teachers administer a FAST-R assessment designed by the Boston Plan for Excellence (BPE) to students.
- Teachers submit the FAST-R assessments to BPE for grading. BPE returns one detailed class report showing each individual student's performance.
- Teachers receive assistance from BPE instructional data coaches to learn how best to analyze the class reports.
- Data coaches help teachers use the assessment results to tailor their reading instruction practices in ways that will strengthen students' areas of weakness.

The FAST-R Program

The FAST-R program begins with a series of short, low-stakes student assessments that incorporate question items aligned with the Massachusetts Comprehensive Assessment System (MCAS),[12] an annual state high-stakes test for students in grades 3 through 8 and grade 10 that is used to measure the performance of both students and schools. FAST-R assessments provide information on students' strengths and weaknesses in English language arts by measuring, primarily, two skills that readers use to construct meaning from texts: *finding evidence* to determine explicit meaning and *making inferences* to determine implicit meaning. Both skills are central to reading comprehension and are frequently assessed on the MCAS. FAST-R formative assessments consist of reading passages from various literary genres, writing styles, and authors. While some content is based on former MCAS reading passages and questions, other question sets and all teacher support materials were created especially for FAST-R. Other types of formative assessments are used throughout BPS, but the FAST-R formative assessments are available only to schools participating in the FAST-R program.

Teachers are encouraged to use FAST-R as a tool to identify and target student needs. Toward this end, teachers at a given school administer the FAST-R assessments to their classes several times over the course of the school year, according to a timetable that each school determines. BPE notes that, depending on the teachers' purposes for giving the FAST-R assessments, some schools administer them every 3 to 6 weeks, while others administer them every 10 to 12 weeks over the course of the 10-month school year. FAST-R is designed to be flexible, so that teachers can choose which assessments to administer depending on their lesson plans. BPE staff grade the assessments for teachers and produce analytic reports that are designed to

[12]Descriptions of the FAST-R program are based on information gathered from the BPE Web site (www.bpe.org) and staff members.

be clear and easy to use. The reports contain detailed data for each student that are intended to help teachers notice and respond to patterns in the performance of individual students and student subgroups, and to let teachers see how students performed on the various items. The reports also explain what incorrect answers signify about each student's level of understanding, and teachers can easily identify how far off students were from selecting the correct answer.

FAST-R instructional data coaches are an integral part of the FAST-R program. The data coaches are experienced teachers, employed by BPE, who previously used FAST-R in their own classrooms. Over the course of the study, BPE typically had one elementary school coach, one middle school coach, and one high school coach on staff. Data coaches meet periodically with the principal and teachers during the school year. They meet with teachers in a group or individually to review the data, to discuss how the findings can be used to modify instruction, and to evaluate students' progress since the last FAST-R assessment administration. The aim is to build a data-driven culture in which every adult in the school shares a sense of responsibility for each student's progress in learning and to link assessment and instruction in a continuous cycle.

The FAST-R program is centered on two of the six BPS "essentials" listed in Box 1.1: (1) examination of student work and data, and (2) professional development. The focus on student work and data encourages teachers to analyze various kinds of data to inform their instructional practices and work with individual students. In addition, as already described, FAST-R provides professional development through sessions with an instructional data coach, although this is only one of numerous professional development opportunities available to teachers in BPS. Within a FAST-R school, the program may be used with specific grade levels depending on the needs of the student body and decisions by school leadership and teachers. The FAST-R program also assists schools and teachers in the use of formative assessments, but many schools not implementing the FAST-R program are also promoting the use of student data to inform instruction. In addition, professional development activities referenced in this report represent district and school-based training sessions, regularly scheduled teacher meetings to review student work, instructional strategies workshops, work with a literacy coach, work with an instructional data coach, and any related activities pursued by all teachers in the district.

The Evaluation and the Central Research Questions

This report presents the findings of an impact analysis and a process analysis to shed light on the FAST-R program in BPS during the study period.

Impact Analysis

The study uses a comparative interrupted time series design to compare the impacts of FAST-R on 21 schools that implemented the program in their third- and/or fourth-grade classes during the 2005-2006 and 2006-2007 school years with 36 control schools that did not implement FAST-R during these years.[13] This is the first evaluation of FAST-R that compares students' reading performance over time in schools that implemented the program with the reading performance of demographically similar students in schools where the program is not present.

The *impact analysis* addresses the following research questions:

- What is the effect of FAST-R on overall achievement in reading, as measured by the MCAS and the Stanford Achievement Test, version 9 (SAT-9), reading assessments?
- What is the impact of FAST-R on students' ability to draw inferences and to provide evidence for their thinking based on a specific reading passage, as indicated by items measuring those skills on the MCAS reading assessment?
- What is the impact of FAST-R on different subgroups of students (for example, by socioeconomic status, gender, prior academic achievement, and so forth)?

Process Analysis

Aspects of program implementation are examined through surveys that were administered to principals and teachers in the FAST-R and non-FAST-R schools in the spring of 2007. This process analysis examines the extent to which FAST-R schools differ from schools that have not implemented FAST-R in terms of the extent and types of professional development teachers receive and teachers' responses to this assistance.

The *process analysis* will address the following research questions:

- Where does FAST-R fit into the array of school improvement efforts being undertaken within BPS? What school improvement efforts (including other efforts to examine data) are in place in both FAST-R and non-FAST-R schools?
- To what extent do teachers in FAST-R and non-FAST-R schools use data to inform their thinking about students and to change their teaching strategies?

[13]The comparative interrupted time series design compares the change in one or more outcomes (such as reading test scores) over a period of time preceding and following the introduction of an intervention for a study group with the change over the same period of time for a similar comparison group that did not receive the treatment.

The Contents of This Report

The remainder of this report is organized into four chapters. Chapter 2 describes the comparative interrupted time series research design methodology used for the impact analysis. The chapter also discusses the sampling methodology used for the surveys that were administered to principals and teachers.

Chapter 3 discusses the professional development activities in FAST-R and non-FAST-R schools highlighted by the findings of the principal and teacher surveys. The chapter also considers how teachers perceived the utility of the FAST-R intervention for their instructional practices.

Chapter 4 describes the findings from the impact analysis of FAST-R with regard to student achievement, exploring the range of student outcomes on the MCAS and the SAT-9 reading assessments. In addition, the chapter reports on an analysis to measure the impact of FAST-R on students' ability to make inferences and find evidence while reading. Lastly, subgroup analyses to compare the effect of FAST-R on various groups of students (by, for example, gender and socioeconomic status) are discussed.

Chapter 5 presents the overall conclusions that may be drawn from the study's analyses and their implications for the use of formative assessments and data-driven instruction to improve reading skills.

Chapter 2

Evaluation Research Design

MDRC's evaluation of the Formative Assessments of Student Thinking in Reading (FAST-R) program sought to measure the impact of the FAST-R program on student achievement in reading and to look at the impact findings within the context of other activities related to professional development and data-driven instruction in the Boston Public Schools system (BPS). Accordingly, the research design includes two main components. The first section of this chapter describes the methods used to analyze the impact of the FAST-R program over and above the status quo — that is, beyond what occurred in the comparison schools, where FAST-R was not implemented. The second section of this chapter discusses the survey methodology used to collect data about teachers and principals at both FAST-R and non-FAST-R schools regarding their professional development activities and use of data to inform their reading instruction. Furthermore, the survey findings may suggest how FAST-R fits into the wider initiatives of BPS.

Impact Analysis

The evaluation focuses on identifying the impact (or ultimate effects) of FAST-R on student reading achievement (the "outcome") over and above what would have occurred in the absence of the program. The goal of this study is to determine whether FAST-R is more effective, less effective, or about as effective in improving student outcomes relative to the mix of reforms that would have occurred without it.

The Analytic Approach

The most challenging aspect of any analysis is to determine the nature of the "counterfactual" — that is, what would have happened in the absence of the program. The most reliable technique to establish the counterfactual is to randomly assign subjects (students or schools) to a treatment group that has access to the program or to a control group that does not have such access. Random assignment ensures that any differences between the two groups, before the treatment occurs, are due to chance. For this reason, the control group outcomes can be used as a reliable estimate of the outcomes that would have been observed among treatment group members without the program. In this case, however, the FAST-R program was implemented before the study was designed and schools were not selected at random for the FAST-R program, making use of an experimental design impossible.

Instead, an alternative research design known as the comparative interrupted time series is applied.[1] This approach compares the change in student outcomes over a period of time preceding and following the introduction of the intervention for schools that adopted FAST-R with the corresponding change over the same period for similar comparison schools that did not adopt it. The ideal conditions for this type of analysis include consistent student outcomes for multiple pre-intervention baseline years and multiple post-implementation follow-up years as well as the inclusion of multiple schools that implemented the FAST-R program and comparison schools that did not implement the program.[2] As can be seen from Box 2.1, the large number of both schools implementing FAST-R and comparison schools that did not implement the program, as well as consistent longitudinal test score data, led to a particularly strong execution of this analysis model.

The approach rests on two comparisons. The first compares the FAST-R schools' student outcomes before and after the program was implemented, since — absent any school or district intervention or change in the school's student body — the best predictor of educational outcomes in a given school is the history of that school's student outcomes. The second comparison controls for any effects of district-wide interventions or changes in student populations occurring simultaneously with the FAST-R program implementation by comparing the change in educational outcomes of students at FAST-R schools with the change in outcomes for similar schools that did not implement FAST-R.

Comparison 1: Deviation from the Trend

The first comparison measures differences in outcomes before and after program implementation at the group of schools where the FAST-R program was put in place. First, a pre-intervention "baseline" period is measured for each outcome in question. Since it is possible that a general growth or decline in the outcome measures could be occurring during the five baseline years, it is important to take into account these baseline "trends" when calculating program effects. For this study, a trend was used to measure the outcomes for the five baseline years (2000-2001 through 2004-2005) and predict those outcomes for the two years (2005-2006 through 2006-2007) following program implementation. The "deviation from the trend" is the difference between the predicted outcome and the actual average outcome for each program school during each follow-up year.

---

[1] Bloom (2003); Snipes (2003).

[2] The inclusion of multiple pre-intervention years allows researchers to capture the growth or decline in student test scores over time, while including more than one post-implementation year allows researchers to analyze the effects of the program beyond the first year after implementation.

Box 2.1

Ideal Elements for an Interrupted Time Series Analysis Model and Actual Components in FAST-R Analysis

- Consistent student outcomes for multiple pre-intervention baseline years:
  Five years of baseline outcomes on MCAS (spring 2001-2005) and SAT-9 (fall 2001-2005) standardized tests

- Consistent student outcomes for multiple post-implementation follow-up years:
  Two years of follow-up outcomes on MCAS (spring 2006 and 2007) and SAT-9 (fall 2006 and 2007)

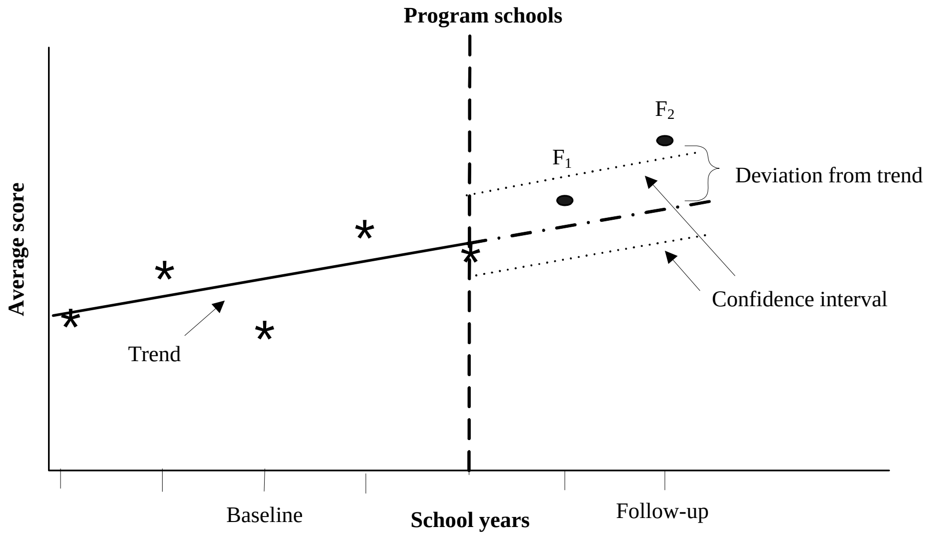
- Multiple schools that implemented the FAST-R program:
  Twenty-one FAST-R schools (20 in third-grade analysis and 19 in fourth-grade analysis)

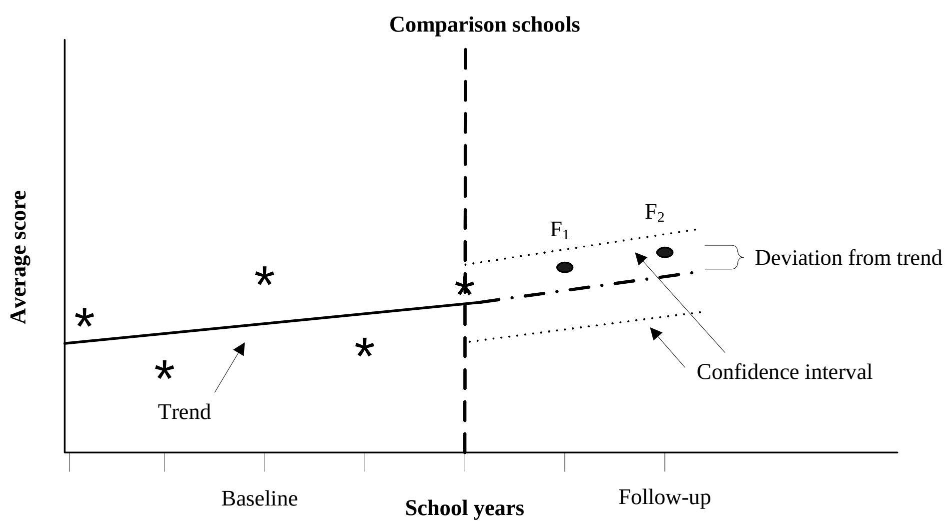
- Multiple comparison schools that did not implement the FAST-R program:
  Thirty-six comparison schools (31 in third-grade analysis and 29 in fourth-grade analysis)

NOTES: MCAS = Massachusetts Comprehensive Assessment System; SAT-9 = Stanford Achievement Test, version 9.

The top panel in Figure 2.1 presents a hypothetical example of this comparison as applied to academic achievement test scores. The diagram plots five years of student achievement before the program intervention began and two years of student achievement following the program implementation. The asterisks on the left-hand side of the graph represent average achievement scores among students in a given program school in the analysis. The solid line running through these points represents the trend in scores across all the baseline years. The dashed line on the right-hand side of Figure 2.1 projects the trend occurring in the five baseline years into the two follow-up years. This latter trend line serves as the benchmark against which post-implementation outcomes can be measured. The points marked F1 and F2 represent hypothetical outcomes in each follow-up year for the program school. The distance between each of these points and the predicted trend represents the deviation from the trend. The dotted lines above and below the projected baseline represent the 90 percent confidence interval — that is, the range of values within which there is a 90 percent probability that the projected trend actually falls. To the extent that the average outcome value in any follow-up year falls outside the confidence interval for the predicted trend, the estimated deviation from trend for that year is statistically significant.

Figure 2.1
Illustration of a Hypothetical Interrupted Time Series Analysis: Difference in the Deviation from the Trend





NOTES: The confidence interval is the range of values within which there is a 90 percent probability that the projected trend actually falls. Each notch on the x-axis represents the end of a school year in a hypothetical study. The vertical dashed line represents the break between the baseline (pre-intervention) and follow-up (post-implementation) periods. F_1 and F_2 represent hypothetical outcomes in each follow-up year.

Looking at the program schools alone, however, it would be impossible to determine how much of the observed change from the projected trend was attributable to FAST-R and how much it reflected changes in other circumstances that could have affected all Boston public schools during the period after FAST-R's implementation. If, during the same period as the FAST-R program, other major changes and/or reforms that are unrelated to FAST-R also occurred, these changes or reforms, rather than the FAST-R program, could be driving the observed effects.

Comparison 2: Difference in the Deviation from the Trend

The second step in the analysis, therefore, is to measure the change in the outcome over the same period at a set of Boston public schools that resemble each FAST-R school in terms of baseline achievement levels and demographics but that did not implement the initiative. The change over time at these comparison schools provides an estimate of how student performance would have changed at the FAST-R schools in the absence of the FAST-R reform but including other reforms that the district may have instituted. The bottom panel of Figure 2.1 illustrates the deviation from the trend at a hypothetical set of comparison schools. This deviation from the trend for the comparison schools is measured in the same manner as was done for the program schools.

The impact is the difference between the deviation from the trend for program schools and the deviation from the trend for the comparison schools.[3] To the extent that the deviations from the trend at program and comparison schools systematically differ in a positive direction, one can conclude that FAST-R had an effect on student achievement, over and above what would have occurred in the absence of the program. To the extent that the deviations do not differ, such a conclusion is not supported.

In short, the approach is to:

- Calculate the deviation from the trend at a set of schools implementing FAST-R.
- Calculate the deviation from the trend at a set of carefully chosen comparison schools.
- Compare the deviations from the trend at FAST-R schools with the deviations from the trend at comparison schools.

[3]The analytic model used to estimate impacts was structured to calculate each school's deviation from its respective trend, which was constructed using the five years of baseline data. It then compares the program schools' deviations with the deviations of the comparison schools to estimate impacts for the overall program.

Selection of FAST-R Schools and Comparison Schools

The analysis described above is conducted separately for third- and fourth-grade students.[4] The 21 FAST-R elementary and K-8 (kindergarten through eighth grade) schools included in the analyses (20 schools in the third-grade analysis and 19 schools in the fourth-grade analysis) were fully and consistently implementing the FAST-R program in third- and/or fourth-grade classrooms with the support of the Boston Plan for Excellence (BPE) during the 2005-2006 school year.[5] Schools that served special populations exclusively or did not test third- and fourth-grade students during any of the baseline or follow-up years were not included. Almost all of the schools included in the program group were also actively implementing FAST-R in the 2006-2007 school year.[6]

The comparison schools are included in the analysis in order to provide an estimate of the progress that would have occurred at the program schools without the FAST-R program. Therefore, the goal in choosing comparison schools was to find a set of BPS schools that, in the absence of the intervention, would be expected to perform similarly to the schools that implemented the FAST-R program. A set of BPS elementary schools that had not been exposed to the FAST-R program during the 2000-2001 through 2006-2007 school years was chosen as a possible comparison group. These schools were then matched to specific FAST-R schools using a set of criteria. The most accurate predictor of future performance on an outcome is usually previous performance on that same outcome. Therefore, prior academic achievement was the primary criterion by which comparison schools were selected. Still, it could be argued that, even with similar prior achievement patterns, schools that serve different student populations may evolve differently over time, particularly in response to local events or district policies. Therefore, comparison schools were also selected because they serve students who are demographically similar to those in the FAST-R schools in terms of race/ethnicity, eligibility for free or reduced-price lunch, and the percentage of special education students and students with limited English proficiency (LEP).

---

[4]Third and fourth grades were chosen for the study because of the availability of MCAS data, but also because reading for comprehension begins to be especially important at these grade levels, as reading materials become more complex.

[5]These 21 schools include 10 schools that were involved in the FAST-R pilot program conducted during the 2003-2004 and 2004-2005 school years. Since most of the schools involved in the pilot program did not fully and consistently implement the FAST-R program during these years and since the pilot version of the program did not include many of the elements existing in the later FAST-R program, the pilot years were treated as baseline for all schools in the study. A sensitivity test of the pilot program was conducted and found that there were no statistically significant positive impacts at the schools involved in the pilot during the years of the pilot program.

[6]Two schools in the third-grade analysis and one school in the fourth-grade analysis were not actively implementing FAST-R in the 2006-2007 school year.

These matching criteria were applied to the third and fourth grades separately using the data from the 2000-2001 through 2002-2003 school years.[7] Since no two schools were exactly alike on all these criteria, each FAST-R school was matched with as many comparison schools as were suitable. At the same time, each comparison school was allowed to match to more than one FAST-R school where applicable.[8] Most FAST-R schools matched closely to other BPS schools on both demographic (percentage minority, poverty, special education, and LEP) and test score variables (Stanford Achievement Test, version 9 [SAT-9], and Massachusetts Comprehensive Assessment System [MCAS]).[9] In some cases, FAST-R schools did not match to comparison schools using the strictest matching criteria, but, upon widening those parameters, every school matched at least one comparison, or non-FAST-R, school.[10]

Characteristics of FAST-R and Non-FAST-R Schools

For the comparative interrupted time series model to be effective, the comparison schools must provide a reliable indicator of how the student outcomes are likely to respond to district-wide policies or events during the years FAST-R is being implemented. To make this assumption, it is important that the FAST-R and non-FAST-R schools' student populations be similar. Table 2.1 presents characteristics of the third- and fourth-grade students in the FAST-R schools and their comparison schools in the five years prior to the full-scale implementation of

---

[7]The 2003-2004 and 2004-2005 school years were not used for matching comparison schools because some of the FAST-R schools had been exposed to the FAST-R pilot program during those years.

[8]The 36 comparison schools (31 schools in the third-grade analysis and 29 schools in the fourth-grade analysis) were weighted so that the data from all the non-FAST-R schools matching a particular FAST-R school were combined to create a "pseudo-school" that most closely matched that FAST-R school. To take into account that a comparison school could also be matched to several FAST-R schools, all the weights given to a specific non-FAST-R school for matching with several different corresponding FAST-R schools were summed.

[9]More specifically, comparison schools were selected using the strictest matching criteria: the third- and fourth-grade students' average SAT-9 and MCAS scores were no more than one-fourth of a standard deviation above or below the FAST-R school's average scores; the percentage of white students, percentage of students receiving free or reduced-price lunch, and percentage of limited English proficient or formerly limited English proficient students in third and fourth grade were no more than 10 percentage points above or below the FAST-R school's percentage; and the percentage of third- and fourth-graders in special education was no more than 15 percentage points above or below the FAST-R school's percentage.

[10]For the third grade, four FAST-R schools did not match to any comparison schools using the above parameters. To match these schools to at least one comparison school, the difference allowed for free and reduced-price lunch was increased from 10 to 15 percentage points, and the difference in schools' average MCAS and SAT-9 scores was increased from one-fourth to one-half of a standard deviation above or below the FAST-R school's average score. For fourth grade, five FAST-R schools did not match on the above parameters but matched to at least one comparison school when the difference in schools' average MCAS and SAT-9 test scores was increased from one-fourth to one-half of a standard deviation above or below the FAST-R school's average score. A sensitivity test was conducted where these schools were excluded from the analyses. The patterns of impacts were similar to those found in the analyses with these schools included.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 2.1

Characteristics of Students at FAST-R and Non-FAST-R Schools: Pre-Intervention Years, SY 2000-2001 to 2004-2005

Characteristic	FAST-R Schools	Non-FAST-R Schools[a]	Difference	P-Value	
Third-grade students					
Age (years)	8.7	8.8	0.0	0.742	
Race/ethnicity (%)					
Black	46.3	58.3	-12.0	0.043	**
White	14.7	11.9	2.8	0.356	
Hispanic	31.6	25.5	6.2	0.258	
Asian	6.8	3.5	3.3	0.137	
Other	0.2	0.3	-0.1	0.466	
Male (%)	54.6	51.6	3.0	0.004	***
Eligible for free lunch (%)	76.4	79.1	-2.7	0.268	
Limited English proficiency (%)	13.5	11.1	2.4	0.414	
Special education (%)	22.5	20.1	2.4	0.327	
Student reading achievement[b]					
MCAS ELA (distance from proficiency)	-4.5	-5.3	0.9	0.194	
MCAS ELA proficient or above (% of students)	36.8	32.6	4.2	0.210	
SAT-9 Total Reading (scaled score)	578.9	577.1	1.8	0.557	
Total number of schools	20	31			
Fourth-grade students					
Age (years)	9.8	9.8	0.0	0.335	
Race/ethnicity (%)					
Black	46.4	55.8	-9.4	0.119	
White	11.0	11.3	-0.3	0.906	
Hispanic	35.4	27.5	7.9	0.208	
Asian	6.4	4.6	1.8	0.437	
Other	0.3	0.4	0.0	0.711	
Male (%)	52.3	52.2	0.1	0.934	
Eligible for free lunch (%)	80.8	80.7	0.0	0.985	
Limited English proficiency (%)	11.0	8.6	2.3	0.314	
Special education (%)	21.4	20.9	0.4	0.871	
Student reading achievement[c]					
MCAS ELA (scaled score)	230.3	229.2	1.0	0.389	
MCAS ELA proficient or advanced (% of students)	26.4	24.0	2.4	0.437	
SAT-9 Total Reading (scaled score)	604.9	603.9	1.0	0.714	
Total number of schools	19	29			

(continued)

Table 2.1 (continued)

SOURCE: MDRC calculations from individual student school records from the Boston Public Schools' district office.

NOTES: A two-tailed t-test was applied to the difference in baseline means between FAST-R and non-FAST-R schools. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Third-grade MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3; fourth-grade MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; SY = school year.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The SAT-9 was administered in the fall of students' third-grade year, and the MCAS was administered in the spring of that year.

[c]The SAT-9 was administered in the fall of students' fourth-grade year, and the MCAS was administered in the spring of that year.

FAST-R in Boston.[11] Approximately 50 percent of the students, across both grades and in both FAST-R and non-FAST-R schools, were black, and 30 percent were Hispanic. Almost 80 percent of the students qualified for free or reduced-price lunches, and approximately 20 percent were identified as having special education needs. As can be seen from the table, the samples of FAST-R schools and non-FAST-R schools used in these analyses are similar. In third grade, there are statistically significant differences between groups in percent black and percent of students who were male. In fourth grade, there are no statistically significant differences between the FAST-R schools and the comparison schools.

Controlling for Changes in School Composition

This analysis takes into account the fact that FAST-R and non-FAST-R schools may experience a change in the composition of their student populations over the seven years included in the analysis. For example, neighborhoods may undergo demographic changes or the rules governing school assignment at the district may change. It is also possible that the implementation of FAST-R could affect the population of students directly or indirectly. In order to account for any systematic changes in the characteristics of the students at particular schools over time, the analysis incorporates individual student characteristics within a given school into the analytic model. Specifically, the analytic model includes controls for students' prior achievement, race/ethnicity, student age, gender, special education status, LEP and former LEP status, and whether or not a student is receiving free or reduced-price lunch used as a proxy for poverty.

[11]The non-FAST-R schools are weighted in the same manner as in the impact analysis (see note 8).

Selection Bias

Because the selection of schools for FAST-R was not random, schools that implemented FAST-R could differ in major (but unmeasured) ways from those that did not mount the intervention. Schools applied to participate in and then were selected for the FAST-R program by BPE staff based on a set of criteria, as described below. Later, comparison schools were chosen from a pool of schools that either applied for the program and were not accepted or never applied. The findings could be biased if the comparison schools do not represent what would have happened to the FAST-R schools in the absence of the program. The choice of a school to apply or not apply for the program may cause bias. For example, a school that was later accepted into the FAST-R program may have applied because the administration felt the school was struggling in this area and needed assistance, while a school that did not apply to be in the FAST-R program, and later became a non-FAST-R school, may have felt the school's current programs in this area were successful. This difference in the FAST-R and non-FAST-R schools may cause the estimated impacts to be smaller than the true impacts. Still, it is also possible that schools that applied for the program were more motivated to change their professional development activities and strengthen their reading programs than schools that did not apply, which might cause the estimated impacts to be larger than the true impacts.

Bias also could have occurred during BPE's selection of schools. If schools were chosen because of their capacity for successful implementation and progress, it is possible that the program effects could be overestimated. On the other hand, if schools were chosen because they had the most need for support and a low capacity for success, it is possible that the program effects could be underestimated. In the selection process, BPE took into account several factors, including both the quality of the applications (perhaps an indication of a school's higher capacity for success) and the need for assistance (perhaps an indication of a school's lower capacity for success). This mixed selection process helped to reduce systematic bias, since schools were chosen for the intervention using two somewhat opposing factors. BPE staff also noted that some schools that wrote persuasive applications did not always implement the initiative consistently, suggesting that high capacity at one point in the process did not ensure that a school's staff would be consistently and adequately able to carry out the reform. Although selection bias is an issue that cannot be fully resolved, the variety of motivations schools may have had in choosing to apply for the program and the variety of criteria that BPE used in the selection process make it unlikely that these effects would overwhelm the findings.

Data Sources and Outcome Measures

The impact analysis relies on individual student records obtained from BPS. This database includes information on individual student performance on standardized tests as well as student data regarding race/ethnicity, school enrollment status, eligibility for free or reduced-price lunch, special education status, and LEP status. The goal of this analysis is to demonstrate whether or not FAST-R has an effect on student reading achievement, which is measured using three outcomes (described below) from two standardized tests — the MCAS and the SAT-9. Two literacy skills integral to the FAST-R program, making inferences from and finding evidence in reading passages, are also measured separately.

MCAS English Language Arts Total Score

The Massachusetts Comprehensive Assessment System is the high-stakes state test in Massachusetts that is administered to all students in the third through eighth grades and tenth grade in the spring of each year. Student scoring on this test is the key impact outcome of interest because it is the state test that is used to monitor school accountability and student achievement.

The MCAS English language arts (ELA) test for third-graders — titled, "English Language Arts, Reading Comprehension, Grade 3" — is based on learning standards in the two content strands of the Massachusetts *English Language Arts Curriculum Framework* (June 2001):[12] "Language" and "Reading and Literature."[13] The third-grade test is not measured in scaled scores (that is, adjusted for comparative purposes) because it does not include material that can reliably distinguish between proficient and advanced work. The Massachusetts Department of Education does publish performance-level thresholds for third-grade test scores, including a threshold indicating proficiency. The threshold scores are measured to maintain the same performance standards across all the years included in this study.[14] To get the most accurate measure of each student's score that is comparable to other students' scores across all years included in the analysis, each student's raw (that is, unadjusted) score was measured as the distance from that year's proficiency threshold.[15]

The MCAS English language arts test for fourth-graders — titled, "English Language Arts, Grade 4," and comprising two sections, Composition and Reading Comprehension — is

---

[12]Published by the Massachusetts Department of Education. Web site: www.doe.mass.edu/frameworks/current.html.

[13]The name of the MCAS third-grade English language arts test prior to 2007 was "Reading, Grade 3."

[14]Massachusetts Department of Education (2007).

[15]There are two question formats on the MCAS reading test in both third and fourth grade, multiple-choice questions and open-ended questions. There are two open-ended questions on the third-grade tests and four on the fourth-grade tests. For third-grade students, open-ended questions were not included in the reported raw scores until the spring 2006 test, thus changing the make-up of the raw score during the follow-up years. The reported raw scores and their corresponding threshold scores for performance levels were used for all years. Although the threshold scores were created to maintain performance levels, this change in the make-up of the test scores may affect the comparability across pre- and post-implementation years. A sensitivity test was conducted where each student's score was standardized by dividing by the standard deviation for all students in that year. When the analysis was run on these standardized outcome variables, similar patterns of impacts were found.

based on learning standards in the three content strands of the Massachusetts *English Language Arts Curriculum Framework* (2001):[16] "Language," "Reading and Literature," and "Composition."[17] The scaled score was used as the outcome variable for fourth grade.

MCAS Percentage At or Above "Proficient" Level in Reading

Massachusetts defines a set of four standards for the MCAS based on students' performance on the reading test. The performance standards are defined as follows: (1) advanced (for fourth grade) and above proficient (for third grade), (2) proficient, (3) needs improvement, and (4) warning. These are the most policy-relevant thresholds with which to compare student reading performance because schools are judged based on these measures. The MCAS percentage at or above "proficient" is used to investigate changes around this policy-relevant threshold. Impact estimates are presented as the percentage of students in the school performing at or above the proficient level.

Making Inferences and Finding Evidence

To further understand students' reading skills, the FAST-R initiative focuses on two key literacy skills: making inferences and finding evidence in text. According to Boston Plan for Excellence (BPE) staff, some 60-75 percent of MCAS items tap into these two skills. To better understand whether or not the FAST-R program is affecting either of those individual literacy skills, impact estimates were calculated on the set of specific MCAS items, as identified by BPE, that measure each skill. To best measure these subsets of items over several years, where the difficulty of the questions in each subset changes from year to year, the difference between each student's score and the state average was calculated for each item in a subset.[18]

[16]Published by the Massachusetts Department of Education. Web site: www.doe.mass.edu/frameworks/current.html.

[17]The name of the Reading Comprehension section of the Grade 4 MCAS ELA prior to 2007 was "Language and Literature."

[18]The reading test consists of multiple-choice and open-ended questions. Each multiple-choice question is worth one raw score point and is counted as one if the student answered it correctly and zero if the student answered incorrectly. If the student answered it correctly, the percentage of students who answered the question correctly across the state is subtracted from one. Otherwise, the state percentage of students who answered the question correctly is subtracted from zero. For example, if 47 percent of students across the state answered a particular multiple-choice question correctly, a student answering that same question correctly would receive a scaled score of 0.53 (or 1 – 0.47); a student who did not answer that question correctly would receive a scaled score of –0.47 (or 0 – 0.47). Each open-ended question is worth four points, and students can receive a score between zero and four. For these questions, the average score statewide is subtracted from the student's score. For example, if a student had a raw score of 3 on an open-ended question and the average raw score statewide was 2.7, the student's scaled score would be 0.3.

Then the scores for all the items in each subset were averaged for each student to create a measure of each subset's average distance from the overall state mean.[19]

SAT-9 Total Reading Score

By design, the formative assessments used in FAST-R resemble MCAS items. Many FAST-R questions are either drawn directly or adapted from the MCAS. The FAST-R questions are also multiple-choice in form, as are many of the MCAS questions. The similarity between the two assessments, however, raises an interpretive problem. If impacts are found on students' MCAS scores, it would not be clear whether those impacts are connected to students' familiarity with the test or to changes in teachers' instructional practices resulting from the use of interim assessments. To address this problem, the Stanford Achievement Test, version 9 (SAT-9) was used to assess students' reading skills.

The SAT-9 is a norm-referenced test, or a test that compares each student's score with a national sample, that BPS uses as a diagnostic tool in the fall of each school year. Generally considered to be a broad test, the SAT-9 Total Reading test covers a wide range of reading-related content, including phonemic awareness, decoding, phonics, vocabulary, and comprehension. Students' scaled scores on the SAT-9 were used as a secondary measure of student reading achievement because the assessment is not similar in form to the FAST-R assessments. Still, there are some issues with the validity of this test as an outcome measure since it is given in the fall of each year and is not a high-stakes test (students have no particular incentive to do well and teachers have no particular incentive to encourage them to try hard). For the third-grade analyses, the fourth-grade fall SAT-9 Total Reading score was used as an outcome, and for the fourth-grade analyses, the fifth-grade fall SAT-9 Total Reading score was used.

Analyses of the FAST-R Program by Gender, Student Socioeconomic Status, Baseline Reading Achievement, and Special Education Status

To better understand who is most affected by the FAST-R program, a secondary set of analyses was done on the effects of FAST-R on student achievement by student socioeconomic status, gender, and students' pre-program reading achievement. The effects on student achievement for those students receiving free or reduced-price lunch were calculated to establish whether the FAST-R program was more or less successful in helping students of lower socioeconomic status. Effects on female and male students were analyzed separately to assess whether FAST-R was affecting the reading achievement of boys and girls differently. Fourth-grade students were

[19]A sensitivity test was conducted in which each student's score was standardized by dividing it by the standard deviation for all students in that year. When the analysis was run on these standardized outcome variables, similar patterns of impacts were found.

grouped by their pre-program reading achievement to explore whether FAST-R was more effective for lower- or higher-achieving students. Students' SAT-9 Total Reading scores from the fall of their third-grade year were used. Students were split into two groups: students who performed at or above the "Level 3, Proficient" [20] performance standard and those who performed below this proficiency threshold.[21] Finally, analyses were done separately on students receiving special education services and those not receiving special education services in both third and fourth grades.[22]

Process Analysis

Methodology

The process analysis is used to provide contextual information for the impact analysis findings in two ways. First, it clarifies the extent of reading instruction and professional development efforts in the non-FAST-R schools. Since the non-FAST-R schools are used to estimate what the FAST-R school student achievement would have been without the implementation of the FAST-R program, a better understanding of instructional improvement initiatives at non-FAST-R schools can aid in explaining the magnitude of any impacts. Second, the process analysis compares principal and teacher attitudes in FAST-R and non-FAST-R schools with regard to school functioning and resources.

Surveys were used to collect data for the process analysis. These surveys were administered to teachers and principals working in BPS schools that participate and do not participate in the FAST-R program.

Survey of Teachers

The survey of teachers contained both close-ended multiple-choice questions and open-ended questions. It asked teachers in both FAST-R and non-FAST-R schools about their parti-

[20]In 1995, proficiency standards were established for the SAT-9. The four performance standards include Level 4, Advanced; Level 3, Proficient; Level 2, Basic; and Level 1, Below Basic. See Harcourt Brace & Company (1997).

[21]Baseline reading achievement subgroups were not created for third-grade students because of missing data for two schools where, during one year of the study, no students took the SAT-9 test in the fall.

[22]Low sample sizes in particular schools precluded analysis of student achievement by students' race/ethnicity. Missing data made it impractical to measure student outcomes by LEP status. School-level subgroup analyses were also attempted, including comparing student-level outcomes in schools where teachers reported receiving more or less FAST-R coaching and where BPE judged that the implementation of FAST-R had been stronger or weaker. However, these school-level subgroup analyses were not possible because there were too few schools where teachers reported receiving five or more hours of coaching during the school year or where BPE judged the school to have fully implemented the program.

cipation in professional development activities that helped them to understand their students' thinking skills, such as district and school-based training sessions, teacher meetings, activities with a literacy coach, and activities pursued by teachers on their own (such as graduate school courses).

Survey items for teachers in FAST-R schools also included questions about specific components of the FAST-R program that were not available to teachers in the non-FAST-R schools, such as the extent of their work with an instructional data coach. Teachers' answers to the multiple-choice questions were computed to compare teachers in FAST-R and non-FAST-R schools on the amount of professional development they received, their attitudes toward the usefulness of that professional development, and their use of student data to inform their instructional practices. The open-ended questions were used as a source for teachers' descriptions of their experiences with professional development activities, using data to inform their teaching practices, and using the FAST-R assessments with their students.

Administration of paper surveys to third- and fourth-grade teachers, including special education and Sheltered English Instruction teachers,[23] was handled by Survey Research Management (SRM), a survey firm. The surveys were administered in spring 2007. SRM attempted to survey teachers in all FAST-R schools and all potential comparison non-FAST-R schools. Teachers were given a $15 Borders gift card as an honorarium for volunteering to complete the surveys. The final survey sample consisted of 169 teacher surveys — 84 FAST-R teachers representing 14 of the 21 FAST-R schools in the impact analysis and 85 non-FAST-R teachers representing 18 of the 34 non-FAST-R schools in the impact analysis. (FAST-R teachers are teachers who are working in participating schools. Non-FAST-R teachers work in schools that did not participate in FAST-R.)The response rates for the teacher surveys were 53 percent for FAST-R schools and 56 percent for non-FAST-R schools.[24] These low response rates can possibly be explained by events in the district. At the time of the survey administration, the Boston Teachers Union was negotiating a new contract, and teachers may have been disinclined to undertake any additional activities beyond what was required of them. Given the low response rate, the teachers who took part in the survey cannot be considered representative of all teachers in schools included in the impact analysis sample. The teacher survey findings, therefore, can be viewed as suggestive only, rather than definitive, in showing the similarities and differences between FAST-R and non-FAST-R teachers.

Table 2.2 displays the characteristics of the FAST-R and non-FAST-R teachers who are included in the survey sample. On average, FAST-R teachers had 10.7 years of experience

---

[23]Sheltered English Instruction teachers work with LEP students.

[24]Due to the low response rate for the teacher survey, researchers analyzed responses from the third- and fourth-grade teachers together in the FAST-R and non-FAST-R schools.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 2.2

Characteristics of FAST-R and Non-FAST-R Teachers Sampled in the Teacher Survey

Characteristic	FAST-R Teachers	Non-FAST-R Teachers	Difference	P-Value
Average number of years in teaching				
At current school	6.6	8.8	-2.2	0.036 **
In Boston Public Schools	9.4	13.4	-4.0	0.006 ***
Reading/English language arts	10.7	13.9	-3.2	0.030 **
Grade taught in the 2006-2007 school year (%)				
Third grade	48.2	47.6	0.6	0.482
Fourth grade	44.4	38.1	6.3	
Third and fourth grades	7.4	14.3	-6.9	
Average number of students taught English language arts	18.5	17.9	0.7	0.675
Type of classes taught (%)				
Regular education	51.8	51.8	0.0	0.585
Inclusion	10.8	1.2	9.6	
Resource room special education	6.0	5.9	0.1	
Substantially separate special education	10.8	21.2	-10.4	
Sheltered English Immersion (SEI)	12.1	11.8	0.3	
Other	8.4	8.2	0.2	
Total number of teachers	84	85		

SOURCES: MDRC calculations are from teacher surveys administered to third- and fourth-grade teachers in elementary Boston Public Schools during April 2007.

NOTES: Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of teachers who filled out a survey. Teachers for whom values are missing are not included in the calculations.

A two-tailed t-test was used to measure statistical significance. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

teaching Reading/English Language Arts compared with 13.9 years of experience for non-FAST-R teachers. Both FAST-R and non-FAST-R teachers, on average, taught English Language Arts to the same number of students in the 2006-2007 school year. On average, FAST-R teachers had 6.6 years of experience teaching at their current school compared with 8.8 years of experience for non-FAST-R teachers. Both of these differences are statistically significant. Despite the variations, both groups of schools were composed of experienced teachers.

Survey of Principals

The survey of principals consisted of questions inquiring about principals' leadership and their teachers' participation in various professional development activities, such as district- and school-based training sessions, teacher meetings, activities with a literacy coach, and activities pursued by teachers on their own. Principals in FAST-R schools were also asked about the benefits of the FAST-R program for their teachers and students.

MDRC staff members contacted the principals and conducted the surveys over the telephone beginning in spring 2007.[25] The survey was e-mailed to principals prior to their phone interview to allow them to follow along. They were given a $15 Borders gift card as an honorarium for volunteering to complete the surveys.

The final survey sample consisted of 24 surveys of principals, representing 24 out of 57 total schools in the impact analysis. Because the survey was conducted near the end of the school year, many principals could not be reached, which contributed to low response rates. The response rates for the principal surveys were 57 percent for FAST-R schools and 33 percent for non-FAST-R schools.[26] On average, FAST-R principals who answered the survey had 6.4 years of experience as principal at their current school, compared with 4.4 years of experience for non-FAST-R principals, but the difference is not statistically significant. Given the low response rates, principals who took part in the survey cannot be considered representative of all principals in schools within the impact analysis sample.

The general results of both surveys are discussed in Chapter 3. As a result of the low response rates, any conclusions drawn from the survey responses are suggestive only and not conclusive.

[25]Researchers believed it would be more efficient to administer the survey over the phone to principals, given the small number of principals in the targeted sample.

[26]Due to low responses rates, survey responses from FAST-R and non-FAST-R principals were aggregated for analysis.

Chapter 3

FAST-R and Other Professional Development Activities in Boston Public Schools

The Boston Public Schools system (BPS) is a reform-rich urban district that has undertaken multiple initiatives to improve instruction and student performance. The districtwide focus on professional development provides school leaders and teachers with a variety of training opportunities and related activities offered by local educational institutions, professional organizations such as the International Reading Association, and other BPS partners, including the Boston Plan for Excellence (BPE).[1] BPE is the district's primary partner in improving instruction and in changing central academic policies.[2] BPS's objective for professional development is to align teacher professional development with school and district goals and the specific learning needs of the student body.[3] As part of its effort to meet that objective, BPE — in partnership with BPS — created FAST-R with the goal of providing timely data to teachers to help students meet proficiency standards in reading.

FAST-R provides one kind of professional development training to teachers by offering them an opportunity to work with an instructional data coach, who helps them analyze formative assessment data, and to build on their existing work with school-based literacy coaches. Coaching is one among a plethora of professional development resources that teachers can utilize within the district, although only the FAST-R teachers work with the BPE instructional data coach.

This chapter presents the analysis of teacher and principal surveys comparing FAST-R and non-FAST-R schools' professional development opportunities and use of student data. In addition, it includes a discussion of the implementation of the FAST-R program. Although teacher and principal surveys serve as the basis for this analysis, the respondents cannot be considered representative of all teachers and principals in schools within the impact analysis sample because of the low response rate, as observed in Chapter 2. The survey findings, therefore, should be considered suggestive rather than definitive. Still, the analysis in this chapter suggests several key findings:

- FAST-R teachers who took the survey reported, in general, that their FAST-R training — that is, professional development activities, including

---

[1] Boston Public Schools (2008b).
[2] According to the BPS Web site (www.bostonpublicschools.org), BPE has a significant role in instruction reform.
[3] Boston Public Schools (2004).

coaching in data interpretation and reading instruction — was useful and contributed to their understanding of student assessment data.

- There are few differences between FAST-R and non-FAST-R teachers with respect to the *amount* of professional development they received during the 2006-2007 school year. Overall, both groups reported that professional development was helpful to some extent in informing their reading instruction practices.

- Survey data from the principals suggest that most principals in both FAST-R and non-FAST-R schools support and encourage teachers to engage in professional development activities.

- Non-FAST-R teachers who took the survey spent more time analyzing the 2006 Massachusetts Comprehensive Assessment System (MCAS) test results than did FAST-R teachers who took the survey. Otherwise, there is not much variation between the FAST-R and non-FAST-R teachers with respect to the time they spent on formative assessment data analysis and perceived usefulness of analyzing both kinds of data.

FAST-R Teacher Training

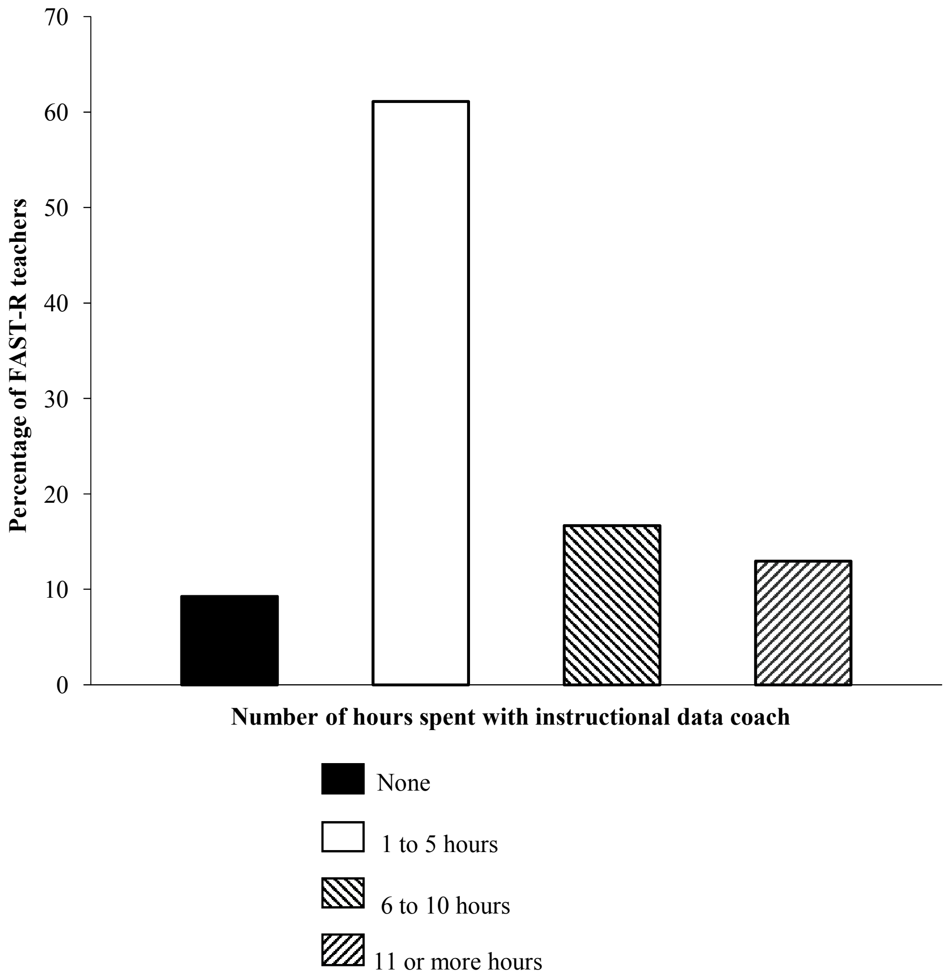
FAST-R teachers participated in targeted professional development with an instructional data coach, who taught them how to analyze formative assessment data and how to utilize data to inform their instructional practices. Generally, FAST-R teachers reported that their FAST-R training was useful and enhanced their understanding of the student assessment data.

Teachers in FAST-R schools reported spending various amounts of time reviewing FAST-R data with a FAST-R instructional data coach. Throughout the school year, one full-time data coach worked with elementary school teachers on how to interpret assessment data, draw conclusions from the data, and then use those conclusions to inform their work with students. Figure 3.1 displays the number of hours FAST-R teachers reported spending with their FAST-R data coach during the 2006-2007 school year. Sixty-one percent of FAST-R teachers spent 1 to 5 hours, while only 13 percent of FAST-R teachers spent 11 or more hours working with a FAST-R data coach. FAST-R teachers reported spending modest amounts of time utilizing this key component of the program's professional development activities. One reason for this limited time may be that only one instructional data coach, assisted sometimes by another BPE staff member, is assigned to work with all the FAST-R elementary schools, which limits

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Figure 3.1

Hours of FAST-R Training with Instructional Data Coach Reported by Teachers in FAST-R Schools, SY 2006-2007



SOURCES: MDRC calculations are based on responses to surveys administered to third- and fourth-grade teachers in elementary Boston Public Schools during April 2007.

NOTES: The sample size for FAST-R teachers was 84. Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of teachers who filled out a survey. Teachers for whom values are missing are not included in the calculations.

SY= school year.

the amount of time that the coach could spend at an individual school. Since teachers spend limited time with the data coach, they may utilize other school-based resources, such as the literacy coach, who observes and gives feedback to teachers about class lessons that are co-designed with the data coach. (The literacy coach is also available to non-FAST-R teachers.) Even though FAST-R teachers spent limited amounts of time working with data coaches, they reported that having a data coach was a valuable resource. For example, one FAST-R teacher stated: "The FAST-R coach was useful in deciphering the data and giving mini-lesson ideas for [student] areas of weakness."

FAST-R teachers described using FAST-R in several ways to identify students' English language arts skill levels. Table 3.1 presents FAST-R teachers' perceptions about the FAST-R program and how it contributed to their teaching. Eighty-eight percent of FAST-R teachers reported using FAST-R to assess students' reading comprehension. The majority of FAST-R teachers also reported that FAST-R contributed to their understanding of assessments (60.8 percent of teachers) and their use of student assessment data to "reflect on" their instructional practices (71.6 percent of teachers). In addition, teachers commented on various ways in which FAST-R data helped them explore their students' ways of thinking about reading. For example, one FAST-R teacher stated:

> One of my students did a better job at making inferences [than the] finding evidence-related questions. This was surprising. I gave him another FAST-R passage a week later (same genre), only this time he responded 100 percent correctly to finding evidence questions and 60 percent to making inferences, [which was] confusing! Yet, it was an opportunity for me to really delve into what and how he was thinking about the text.

FAST-R teachers reported that FAST-R helped them better understand students' thinking and the answers students chose during reading exercises. Figure 3.2 presents FAST-R teachers' responses regarding FAST-R's effect on their perceptions of students' abilities. Eighty-four percent of FAST-R teachers found that FAST-R data confirmed what they knew about their students. One FAST-R teacher explained:

> Using FAST-R with my class has been instrumental in providing my students with the strategies [and] skills to become better readers and thinkers. While discussing the results, students use accountable talk in explaining their reasoning.[4] This [accountable talk] is powerful because it clarifies misunderstanding and empowers them to be active learners.

---

[4]In class lessons characterized by "accountable talk," students are required to support and justify their answers with reasoning and evidence.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 3.1

FAST-R Teachers' Uses and Perceptions of the FAST-R Program: Boston Public Schools, SY 2006-2007

Use of FAST-R Program	FAST-R Teachers
Teachers reporting using FAST-R for a particular purpose (%)	
Assessment of students' reading comprehension	88.3
Assessment of students' use of particular reading skill or strategy	87.0
Way of giving students additional experience with the MCAS "testing genre"	83.1
Part of a Collaborative Coaching and Learning (CCL) cycle	47.3
Teachers reporting how FAST-R contributed to a "moderate" or "great" extent for a particular purpose (%)	
Understanding of how data can be relevant to students' work	74.3
Understanding of different kinds of assessments and their uses	60.8
Using data to reflect on practice	71.6
Sample size	84

SOURCES: MDRC calculations are from teacher surveys administered to third- and fourth-grade teachers in elementary Boston Public Schools during April 2007.

NOTES: Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of teachers who filled out a survey. Teachers for whom values are missing are not included in the calculations.

MCAS = Massachusetts Comprehensive Assessment System; SY = school year.

Professional Development in English Language Arts

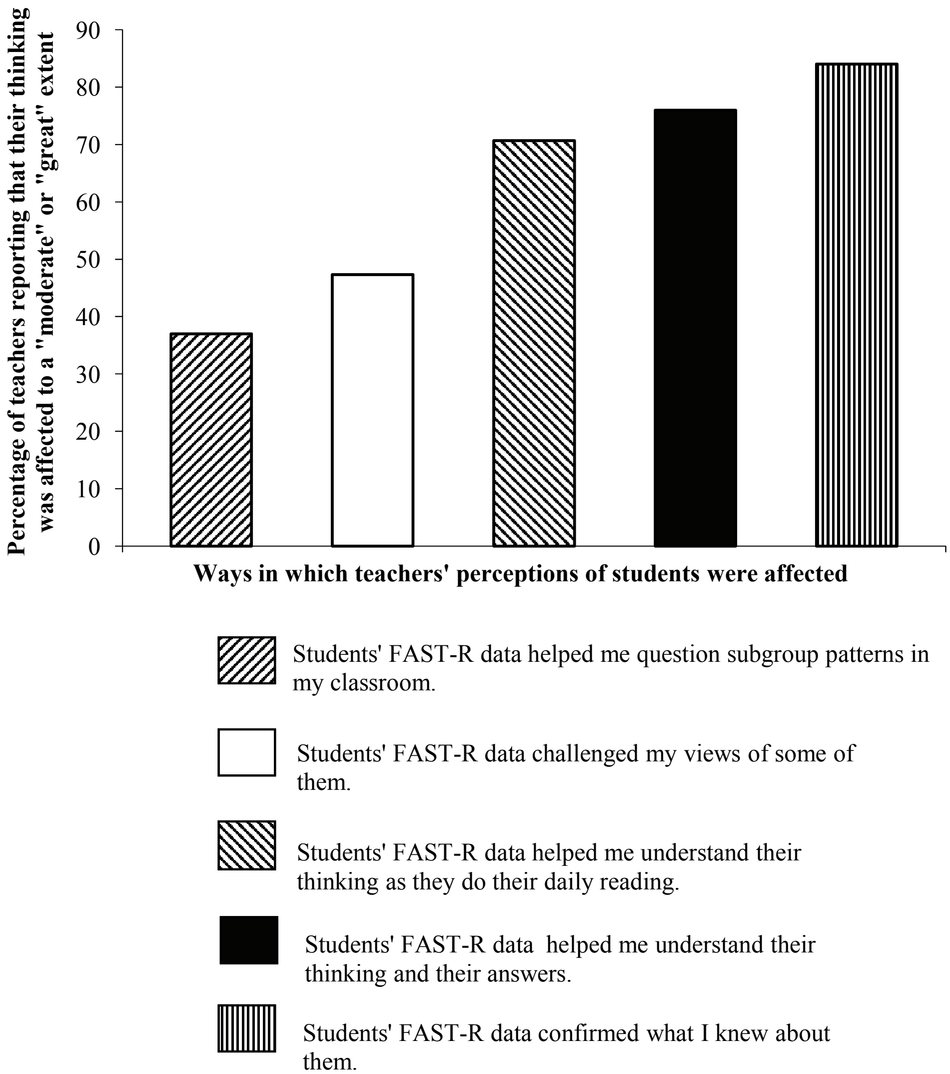
Compared with the amount of professional development that teachers in the FAST-R schools reported they had received, teachers in non-FAST-R schools reported receiving as much or more professional development. Professional development activities were available to both groups of teachers through widespread offerings provided by the Boston Public Schools district. Overall, both groups also reported that professional development was helpful to some extent in informing their reading instruction practices.

Table 3.2 compares how FAST-R and non-FAST-R teachers perceive the time spent on and usefulness of various types of professional development activities during the 2006-

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Figure 3.2

Teachers' Reports of How FAST-R Data Affected Their Perceptions of Students



SOURCES: MDRC calculations are based on responses to surveys administered to third- and fourth-grade teachers in elementary Boston Public Schools during April 2007.

NOTES: The sample size for FAST-R teachers was 84. Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of teachers who filled out a survey. Teachers for whom values are missing are not included in the calculations.

Table 3.2

FAST-R and Non-FAST-R Teachers' Perceived Usefulness of and Time Spent on Professional Development: Boston Public Schools, SY 2006-2007

Type of Professional Development	FAST-R Teachers	Non-FAST-R Teachers	Difference
Teachers reporting time spent on professional development (%)			
Curriculum-specific professional development			
None	33.3	12.0	21.3 ***
1 to 5 hours	17.3	5.3	12.0
6 to 10 hours	10.7	20.0	-9.3
11 or more hours	38.7	62.7	-24.0
Observing other teachers' classrooms			
None	36.3	30.7	5.6 *
1 to 5 hours	47.5	38.7	8.8
6 to 10 hours	7.5	6.7	0.8
11 or more hours	8.8	24.0	-15.2
Collaborative Coaching and Learning (CCL) or work with a literacy coach			
None	18.5	11.4	7.1
1 to 5 hours	18.5	19.0	-0.5
6 to 10 hours	19.8	21.5	-1.7
11 or more hours	43.2	48.1	-4.9
Teacher's own professional development			
None	33.8	24.3	9.5
1 to 5 hours	17.5	17.6	-0.1
6 to 10 hours	11.3	6.8	4.5
11 or more hours	37.5	51.4	-13.9

(continued)

Table 3.2 (continued)

Type of Professional Development	FAST-R Teachers	Non-FAST-R Teachers	Difference
Teachers reporting professional development as "somewhat" or "very" useful (%)			
Curriculum-specific professional development	60.6	88.2	-27.6 ***
Observing other teachers' classrooms	61.3	64.4	-3.1
Collaborative Coaching and Learning (CCL) or work with a literacy coach	92.4	98.5	-6.1 **
Teacher's own professional development	67.6	74.7	-7.1
Teachers reporting to a "moderate" or "great" extent that professional development contributed to understanding how data can be used (%)			
To guide instruction	65.5	71.4	-5.9
To build on students' strengths	71.1	72.6	-1.5
Sample size	84	85	

SOURCES: MDRC calculations are from teacher surveys administered to third- and fourth-grade teachers in elementary Boston Public Schools during April 2007.

NOTES: Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of teachers who filled out a survey. Teachers for whom values are missing are not included in the calculations.

SY = school year.

A chi-square test was used to measure statistical significance. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * =10 percent. The chi-square test may not be valid due to small sample sizes within the cross-tabulation distribution.

2007 school year. Generally, similar proportions of FAST-R and non-FAST-R teachers reported that the various types of professional development in which they had participated over the course of the year were "somewhat useful" or "very useful." FAST-R teachers and non-FAST-R teachers similarly reported that the professional development they had received contributed to their understanding of data to both guide instruction and build on students' strengths.

The majority of teachers in both groups found professional development provided by BPS to be useful and were especially enthusiastic about Collaborative Coaching and Learning (CCL), the district's school-based professional development delivery model. CCL is used in all BPS schools to encourage and create opportunities for collaboration among teachers. Teachers work in teams during eight-week cycles to learn about instructional strategies for reading and writing. CCL coaches share these strategies with teachers, who then take turns demonstrating them to their colleagues.

Table 3.3 displays the percentage of BPS principals who reported that they encouraged their teachers to participate in professional development activities and to analyze student data during the 2005-2006 and 2006-2007 school years. Generally, BPS principals encouraged their third- and fourth-grade teachers "to a moderate extent" or "to a great extent" to participate in these activities. BPS principals in both FAST-R and non-FAST-R schools supported professional development activities as part of the instructional reforms taking place throughout the district.

Table 3.4 compares FAST-R and non-FAST-R teachers' responses regarding the specific ways in which professional development helped their instructional practices. Both FAST-R and non-FAST-R teachers reported that professional development strengthened their instructional practices by helping them conduct classroom discussions and create assignments focused on reading skills like finding evidence in and making inferences from text. The FAST-R intervention stresses these skills, so it is not surprising that FAST-R teachers receive this guidance. Non-FAST-R teachers may receive similar guidance focused on these skills from other professional development resources such as their school literacy coaches. Literacy coaches are a common resource among FAST-R and non-FAST-R teachers because they are members of the school staff in BPS.

Examining the Use of Data at FAST-R and Non-FAST-R Schools

The MCAS test and formative assessments provide the two main forms of student performance data available to BPS teachers. The MCAS, which is an annual high-stakes state test used for accountability, measures students' achievement on state standards. Formative assessments are low-stakes assessments administered frequently during the year. Through the FAST-R program, FAST-R teachers receive formative assessments designed by BPE to distribute to

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 3.3

Proportion of Principals Who Encouraged Professional Development Activities and Data Analysis: Boston Public Schools, SY 2005-2006 and SY 2006-2007

Type of Professional Development and Data Analysis Activity	BPS Principals
Principals reporting that they encouraged professional development to a "moderate" or "great" extent (%)	
Curriculum-specific professional development	69.6
Observing other teachers' classrooms	78.3
Collaborative Coaching and Learning (CCL) or work with a literacy coach	78.3
Work with an instructional data coach	65.2
Teacher's own professional development	69.6
Other professional development	79.0
Principals reporting that they encouraged data analysis to a "moderate" or "great" extent (%)	
2006 MCAS ELA results	100.0
Formative assessments	95.2
Sample size	24

SOURCES: MDRC calculations are from surveys administered to principals in FAST-R and non-FAST-R schools in elementary Boston Public Schools during April 2007.

NOTES: Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of principals who filled out a survey. Principals for whom values are missing are not included in the calculations.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts; SY= school year.

students. Non-FAST-R teachers may have access to formative assessments through supporting materials that accompany their reading curriculum guides.

Table 3.5 shows how FAST-R and non-FAST-R teachers compare on the time spent and perceived usefulness of analyzing data. Non-FAST-R teachers reported spending more time than FAST-R teachers analyzing the 2006 MCAS results. This difference is statistically significant. On the other hand, there is not a statistically significant difference in the time non-FAST-R and FAST-R teachers reported analyzing formative assessments. Although not statistically significant, non-FAST-R teachers were more likely than FAST-R teachers (87 percent compared with 81 percent) to report that analyzing the MCAS results was “somewhat” or “very”

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 3.4

Ways in Which Professional Development Helped FAST-R and Non-FAST-R Teachers: Boston Public Schools, SY 2006-2007

Type of Professional Development	FAST-R Teachers	Non-FAST-R Teachers	Difference
Teachers reporting professional development helped to a "moderate" or "great" extent (%)			
Conduct classroom discussions focused on reading skills like finding evidence and making inferences	76.2	75.0	1.2
Create assignments focused on reading skills like finding evidence and making inferences	73.8	75.9	-2.1
Group students strategically for guided reading or instruction in particular reading strategies	65.8	66.7	-0.9
Sample size	84	85	

SOURCES: MDRC calculations are from teacher surveys administered to third- and fourth-grade teachers in elementary Boston Public Schools during April 2007.

NOTES: Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of teachers who filled out a survey. Teachers for whom values are missing are not included in the calculations.

A chi-square test was used to measure statistical significance. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * =10 percent. The chi-square test may not be valid due to small sample sizes within the cross-tabulation distribution.

useful. FAST-R teachers were slightly more likely than non-FAST-R teachers (95 percent compared with 92 percent) to report that analyzing formative assessments was “somewhat” or “very” useful for teaching, although, again, the difference was not statistically significant. One FAST-R teacher elaborated, stating,

> Analyzing student data has informed me about the areas in [the] students’ knowledge that need to be reinforced or focused on. It doesn’t just inform me on what/how to teach to get students ready for assessments but also [what is needed] for them to be academically competent and well rounded.

It is possible that FAST-R teachers spend less time analyzing data because the student performance reports generated from the FAST-R assessment are designed to be user-friendly. These FAST-R reports were created to provide student data in a more efficient manner than other standard reading curriculum assessments. They contain color-coded graphics to allow

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 3.5

FAST-R and Non-FAST-R Teachers' Reported Usefulness of and Time Spent Analyzing Data: Boston Public Schools, SY 2006-2007

Type of Data	FAST-R Teachers	Non-FAST-R Teachers	Difference
Number of hours teachers reported analyzing data (%)			
2006 MCAS ELA results			
None	10.1	5.0	5.1 ***
1 to 5 hours	73.4	53.8	19.6
6 to 10 hours	8.9	23.8	-14.9
11 or more hours	7.6	17.5	-9.9
Formative assessments			
None	2.5	3.7	-1.2
1 to 5 hours	48.2	46.9	1.3
6 to 10 hours	18.5	18.5	0.0
11 or more hours	30.9	30.9	0.0
Teachers reporting analyzing data as "somewhat" or "very" useful (%)			
2006 MCAS ELA results	80.5	87.2	-6.7
Formative assessments	95.0	92.3	2.7
Sample size	84	85	

SOURCES: MDRC calculations are from teacher surveys administered to third- and fourth-grade teachers in elementary Boston Public Schools during April 2007.

NOTES: Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of teachers who filled out a survey. Teachers for whom values are missing are not included in the calculations.

A chi-square test was used to measure statistical significance. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * =10 percent. The chi-square test may not be valid due to small sample sizes within the cross-tabulation distribution.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts; SY= school year.

teachers to identify quickly and easily which questions students answered incorrectly — individually, within subgroups, and overall as a class.

Table 3.6 displays the various school personnel with whom teachers were likely to review student data. Overall, both FAST-R and non-FAST-R teachers reported they were most likely to review data with other teachers, followed by their principals and then their literacy coaches. None of these differences is statistically significant. Both FAST-R and non-FAST-R

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 3.6

FAST-R and Non-FAST-R Teachers, by School Personnel with Whom They Reviewed Data: Boston Public Schools, SY 2006-2007

School Personnel with Whom Teachers Reviewed Data (%)	FAST-R Teachers	Non-FAST-R Teachers	Difference
Principal			
Yes	61.9	60.7	1.2
No	38.1	39.3	-1.2
Assistant principal			
Yes	14.3	15.5	-1.2
No	85.7	84.5	1.2
Literacy coach			
Yes	57.1	59.5	-2.4
No	42.9	40.5	2.4
Other teachers			
Yes	88.1	86.9	1.2
No	11.9	13.1	-1.2
Sample size	84	85	

SOURCES: MDRC calculations are from teacher surveys administered to third- and fourth-grade teachers in elementary Boston Public Schools during April 2007.

NOTES: Rounding may cause slight discrepancies in calculating sums and differences. The sample size reported represents the number of teachers who filled out a survey. Teachers for whom values are missing are not included in the calculations.

A chi-square test was used to measure statistical significance. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * =10 percent. The chi-square test may not be valid due to small sample sizes within the cross-tabulation distribution.

SY = school year.

teachers stated in their open-ended survey responses that many opportunities exist for reviewing data with other teachers, including CCL meetings and grade-level meetings.

A variety of initiatives emphasizing the use of data to improve instruction are diffused throughout BPS. Overall, surveyed teachers represented in both FAST-R and non-FAST-R schools have reported taking advantage of district-wide professional development opportunities. The FAST-R program is one of many supplemental tools for instructional support available to teachers working in the BPS district.

Chapter 4

The Effects of FAST-R on Reading Achievement for Boston Public Schools' Third- and Fourth-Grade Students

The main goal of the FAST-R impact analysis is to assess whether FAST-R improved achievement in reading for students in the Boston Public Schools system (BPS) over and above what would have been observed in the absence of the program. The analysis compares changes in student achievement at a set of FAST-R schools with changes in achievement at a set of comparison schools that did not implement the FAST-R program. As discussed in the first chapter, BPS has incorporated a wide array of educational and teacher professional development programs and has made strengthening the use of assessment data a district-wide goal. Therefore, FAST-R schools are being compared with schools that have programs that may include some similar components, making it harder for the FAST-R program to attain positive impacts beyond the status quo.

This chapter presents the impacts of FAST-R on reading achievement for third- and fourth-grade students during the 2005-2006 and 2006-2007 school years. The analysis in this chapter suggests several key findings:

- Although the estimated impacts of FAST-R on third- and fourth-grade student reading achievement as measured by the Massachusetts Comprehensive Assessment System (MCAS) are generally positive, these findings are not statistically significant. Therefore, the effect of FAST-R on student reading achievement in both third and fourth grade over and above what occurred in similar schools throughout the district is indeterminate.

- As a secondary measure of student reading achievement, the effect of FAST-R on the Stanford Achievement Test, version 9 (SAT-9) given in the fall following a full year of the FAST-R intervention was also measured. The estimated impacts on student reading achievement on the SAT-9 were mixed and were not statistically significant, suggesting that FAST-R did not improve these outcomes beyond what would have happened without the program.

- The FAST-R formative assessment focuses on two types of questions that make up a large percentage of the questions on the MCAS English Language Arts (ELA) test, questions that ask students both to make inferences from and to find evidence in a two-page reading passage (see Appendix D). The

FAST-R program did not have a statistically significant impact on students' ability to answer either type of question.

- Subgroup analyses were conducted on boys and girls, students of differing socioeconomic status, students performing above or below the "proficient" level on a baseline reading test, and students receiving and not receiving special education services. There are very few statically significant impacts across all of these analyses. Since these few impacts do not create a coherent story and the number of impacts is less than might be expected as a result of chance, they should be interpreted cautiously.

Effects of the FAST-R Program on Third-Grade Students' Reading Achievement

The effects of FAST-R on third-graders' reading achievement are measured primarily by looking at students' performance on the MCAS. In third grade, the MCAS ELA total score is measured as the deviation from the proficiency threshold. This is because MCAS scores for third-graders are not measured in scaled scores, which are conversions of students' raw scores into scores that are comparable across different versions of the test. The raw scores, or actual number of items a student answered correctly on the test, that are provided by the Massachusetts Department of Education are not comparable over the five years before the intervention and the two years after implementation that are included in the study. The threshold between scores in the "needs improvement" and "proficient" categories, published by the Massachusetts Department of Education each year, offers a method for comparing students' test scores across the years of the study. For instance, the threshold score between "needs improvement" and "proficient" for the third-grade reading test in spring 2006 was 37 raw score points. If a student answered 40 of the questions correctly, her distance from the proficiency threshold would be +3. If a student scored 35 on the test, his distance from the proficiency threshold would be –2.

Figure 4.1 (page 44) illustrates the average change in third-grade students' MCAS ELA scores at a "representative" FAST-R and non-FAST-R school over the five pre-implementation years and then projects those trends into the post-implementation years.[1] As can be seen from

[1]To create Figure 4.1, the average baseline scores and the trends over time for "representative" FAST-R and non-FAST-R schools were calculated. ("Representative" reflects what is being portrayed based on the steps described here.) To do this, the student-level baseline characteristics were first centered by subtracting the school-level mean from each characteristic for each student. For example, if the student is male, the value for whether or not a student is male equals one (it equals zero if the student is female). If 52 percent of the students in the school this student attends were male, the student's centered value for male would equal 0.48 (1 – 0.52). The model was run using every student's centered baseline characteristics to find the average baseline score and trend for the average student in each school. The average baseline score and the trend for a representative FAST-R school are the
(continued)

this figure, the trends in student test scores for both the representative FAST-R and non-FAST-R schools during the five years before implementation were nearly flat (that is, 0.0 for the representative FAST-R school and 0.03 for the representative non-FAST-R school), suggesting that third-grade student achievement on the MCAS, during the five years before FAST-R was implemented, was steady in both FAST-R and non-FAST-R schools.

Impact findings of the effects of FAST-R on third-grade students' reading achievement for both the 2005-2006 and 2006-2007 school years are presented in Table 4.1 (page 46). In the first year after implementation, the average distance from proficiency for the students in the representative FAST-R school was –7.8. This means that students in FAST-R schools scored, on average, 7.8 points below the proficiency threshold in the spring of 2006. The *deviation from the trend* for FAST-R students (–2.1) in 2006 is the amount that the distance from proficiency (–7.8) for students in FAST-R schools deviates from the expected outcome (–5.7). It suggests that in the spring of 2006, third-grade students in FAST-R schools scored 2.1 points below what would have been expected given the trend in test scores for the five years before the FAST-R program was implemented. Students in the non-FAST-R schools scored an average of 8.1 points below the proficiency threshold in the spring of 2006. The *deviation from the trend* for non-FAST-R students was –2.2, suggesting that students in non-FAST-R schools scored 2.2 points below the expected outcome. Therefore, in both FAST-R and non-FAST-R schools, average reading achievement on the MCAS for third-grade students was lower than expected in the first year after FAST-R was implemented, given the trends in test scores during the five years before the program began for each group.

Since the decline in test scores from pre- to post-implementation is similar for students in both the FAST-R and non-FAST-R schools (that is, the difference is not statistically significant), the decline in student achievement cannot be attributed to the FAST-R program.[2] The estimated impact of FAST-R is the difference between the deviation from the trend for students in FAST-R and non-FAST-R schools. Since the main difference between these two groups is that one participated in the FAST-R program and the other did not, the estimated impact can be con-

---

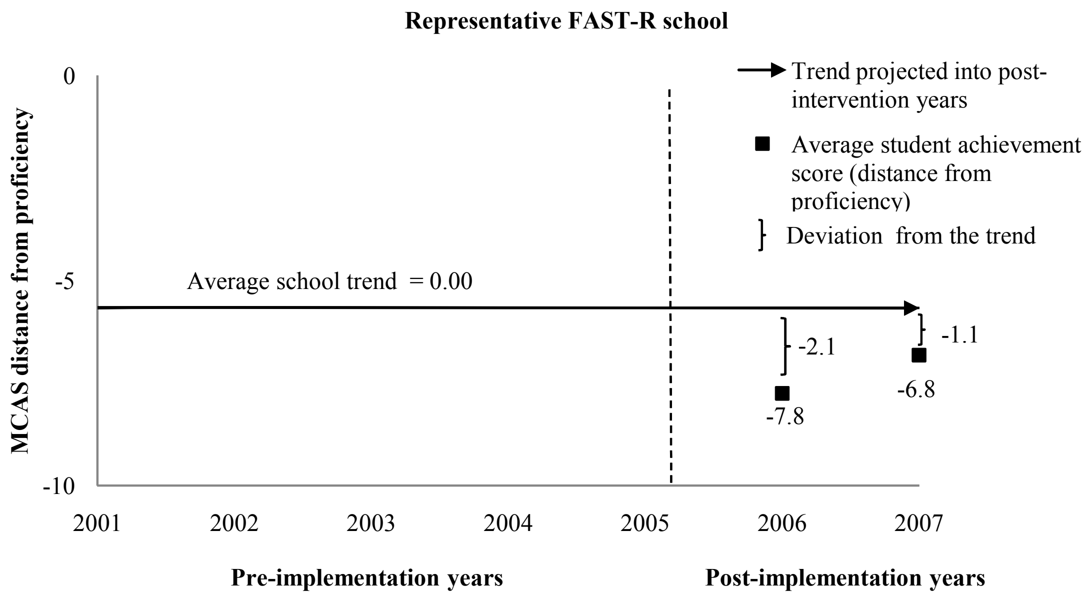
weighted averages of the results for all FAST-R schools. The average baseline score and the trend for a representative non-FAST-R school are the weighted averages of the results for all non-FAST-R schools.

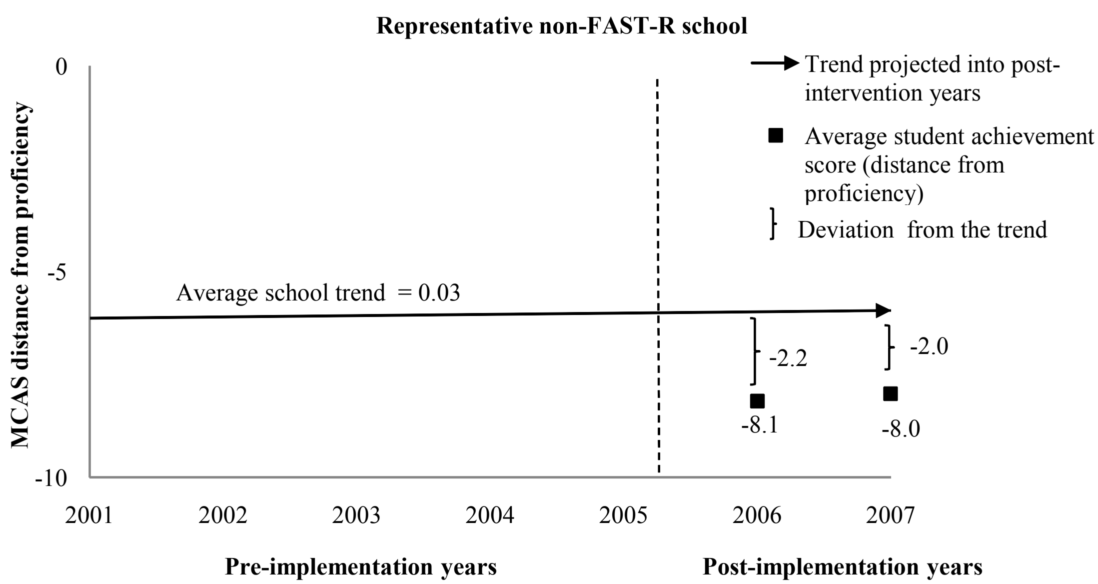
[2]As suggested in Chapter 2, the scoring for the third-grade MCAS ELA raw score was different for the pre-implementation and post-implementation years. In all seven years, students answered two open-ended questions but these questions were not included in the calculations of the students' raw scores until 2006. Although this change in the test make-up is accounted for in the MCAS performance standards used to create the "distance from proficiency" measure, it may have affected the deviations from the trends for both follow-up years, causing some of the drop in test scores. But, since the impacts are measured as the comparison between FAST-R and non-FAST-R students and the scoring methods for these two groups are the same, the change does not affect the impact analysis. As a sensitivity test, the scores were standardized by dividing each student's score by the standard deviation of all student scores during that year. When the analysis was run on these standardized scores, the patterns in outcomes were similar.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Figure 4.1

Trends in Third-Grade Students' MCAS ELA Scores





(continued)

Figure 4.1 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: To calculate the trends and intercepts for a representative FAST-R and non-FAST-R school, the student-level baseline characteristics were first centered by subtracting the school-level mean from the characteristic for each student. The model was run using these centered variables to find the trend and intercept for the average student in each school. The trend and intercept for each school was then multiplied by the percentage of the total students in the school. The results for all FAST-R schools were summed to find the trend and intercept for a "representative" FAST-R school, and the results for all non-FAST-R schools were summed to find the average trend and intercept for a "representative" non-FAST-R school.

The deviation from trend for each post-implementation year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each post-implementation year for the FAST-R schools is the regression-adjusted impact added to the non-FAST-R schools' deviation from trend. The average student achievement scores are calculated by subtracting the deviation from trend from the predicted average for a given follow-up year, calculated using the average school trend and intercept.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3.

sidered the effect of the FAST-R program. As shown in Table 4.1, the estimated impact of FAST-R is 0.1 raw score point during the first year after implementation, which corresponds to an effect size of 0.01 standard deviation (or 1 percent of the standard deviation, a measure of the variation in the distribution that is used to compare outcomes across variables, of both FAST-R and non-FAST-R students' MCAS scores during the five years before FAST-R was implemented).[3] The difference between the students in the FAST-R and non-FAST-R schools is negligible and is statistically indistinguishable, as indicated by the lack of asterisks. Thus, it cannot be concluded that the actual impact is different from zero. In the second year after implementation, there is a notably larger estimated impact of FAST-R on the third-grade MCAS ELA score of 0.9 raw score point, with an effect size of 0.11 standard deviation. Still, this finding is not statistically significant and, again, it cannot be concluded that this impact is actually different from zero.

Effect sizes are used to make outcomes across different studies and variables comparable. Although no absolute standard exists to define whether a specific effect size is large or small, there are some guidelines. One way to think about the findings in this report is to compare them with a typical year of growth for a similar population of students. An analysis of the students outcomes on seven standardized reading tests given to students across the country found that the average growth in student test scores from second to third grade was an effect size of 0.60 standard deviation. This means that students' general overall growth in reading from the spring of second grade to the spring of third grade was approximately 60 percent of a

---

[3]The effect size is a measurement used to compare outcomes across variables and studies. It was calculated by dividing the impact by the standard deviation of all students' scores during the five baseline years.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 4.1

Impacts on Third-Grade Students' Reading Test Scores

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (distance from proficiency)	-7.8	-8.1			
Deviation from trend	-2.1	-2.2	0.1	0.01	0.888
MCAS ELA proficient or above (% of students)	27.4	27.6			
Deviation from trend	-5.1	-3.5	-1.7	NA	0.595
SAT-9 Total Reading, fall fourth grade (scaled score)	604.3	605.8			
Deviation from trend	0.9	3.7	-2.8	-0.07	0.280
Second year after implementation					
MCAS ELA (distance from proficiency)	-6.8	-8.0			
Deviation from trend	-1.1	-2.0	0.9	0.11	0.243
MCAS ELA proficient or above (% of students)	28.6	25.9			
Deviation from trend	-4.5	-5.6	1.1	NA	0.770
SAT-9 Total Reading, fall fourth grade (scaled score)	604.4	604.3			
Deviation from trend	0.8	1.9	-1.1	-0.03	0.722
Sample size[c]			20,127		

(continued)

Table 4.1 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "first year after implementation," all outcomes are for the group of students in third grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fourth-grade year. In the bottom panel, "second year after implementation," all outcomes are for the group of students in third grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fourth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students in the entire sample (by grade) during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or above outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all third-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 16,792 students and includes all third-grade students with SAT-9 scores reported in the fall of their fourth-grade year.

standard deviation. From third to fourth grade, the effect size was 0.36 standard deviation.[4] Effect sizes of the impacts of the FAST-R program can be compared with these measures of overall growth during these grades. For instance, the effect size on the MCAS English Language Arts test for third-grade students in the first year after implementation of 0.01 standard deviation is quite small considering that student reading achievement, on average, increases by an effect size of 0.60 standard deviation during the third-grade year. Of course, it should be remembered that the above impact estimates, and most of the impact estimates found in this study, are not statistically significant, and so it is not possible to know whether the true impacts are actually different from zero.

Another way to think about the outcomes for this study is to compare them with the outcomes for other similar types of interventions in similar studies. A meta-analysis (that is, a statistical analysis of data combined from multiple studies) of 76 meta-analyses of educational interventions found that the mean effect size for elementary school students in these studies was 0.23 standard deviation for lower grades (first through third) and 0.22 standard deviation for higher grades (fourth through fifth). Another meta-analysis of 61 random assignment studies (the most rigorous impact design available) of educational interventions found that the mean effect size for all the elementary school findings in these studies was 0.33 standard deviation. This meta-analysis also found that the mean effect size for random assignment studies conducted using a broad standardized test (similar to MCAS) was 0.07.[5] Again, an effect size, like the one seen above, of 0.01 standard deviation is small considering that the average effect size for studies using similar measures is at least seven times as large.

The second measure in Table 4.1 is the percentage of third-grade students who scored at a level of proficient or above on the Grade 3 MCAS ELA test. A positive impact on this measure would suggest that the program was particularly successful in helping lower-achieving students score at or above their grade level. As can be seen from the table, 27.4 percent of third-grade students at a representative FAST-R school scored at the proficient level or above in the spring of 2006 — 5.1 percentage points lower than expected, given the trend in the percentage of students scoring at that level in the five years before implementation.[6] In a representative non-FAST-R school, 27.6 percent of third-grade students scored at or above proficient — 3.5 percentage points lower than anticipated by the trend in test scores from the five years before FAST-R implementation. Although the FAST-R schools show a greater negative deviation

[4]Hill, Bloom, Black, and Lipsey (2007).
[5]Hill, Bloom, Black, and Lipsey (2007).
[6]The average yearly baseline trends for students in the representative FAST-R school (0.52 percentage point per year) and those in the representative non-FAST-R school (0.50 percentage point per year) are positive, suggesting that the percentage of students scoring at or above proficient during the five years before program implementation was growing slightly in both FAST-R and non-FAST-R schools.

from the trend, the difference in the deviations is not statistically significant, and thus it cannot be concluded that the actual impact is different from zero. During the second year after implementation, FAST-R schools show a smaller negative deviation from the trend than non-FAST-R schools (–4.5 percentage points compared with –5.6 percentage points, respectively). Still, the difference in this deviation, or the estimated impact, of 1.1 percentage points is not statistically significant.

The final measure in Table 4.1 is the effect of the FAST-R program on the SAT-9 Total Reading scaled score given in the fall of the students' fourth-grade year. The deviation from the trend is positive for both FAST-R and non-FAST-R students, suggesting that both groups had higher average test scores than expected from the trend in SAT-9 scores during the five years before the implementation of FAST-R.[7] The deviation from the trend for students in FAST-R schools is smaller than the deviation from the trend for students in non-FAST-R schools, making the impact –2.8 scaled score points, with an effect size of –0.07 standard deviation. This negative impact is not statistically significant. In the second year after implementation, the estimated impact on the SAT-9 Total Reading test is –1.1 scaled score points, and is also not statistically significant. Overall, there are no statistically significant impacts for the FAST-R program on the Grade 3 MCAS ELA or SAT-9 Total Reading scores.

Effects of the FAST-R Program on Fourth-Grade Students' Reading Achievement

Table 4.2 presents the impacts for both the 2005-2006 and the 2006-2007 school year on Grade 4 MCAS ELA and SAT-9 Total Reading scores. In fourth grade, the MCAS ELA test is measured in scaled scores. This metric is comparable over the seven years of the study and is used in the analysis. The findings for the Grade 4 MCAS ELA scaled score tell a similar story as those for third grade. Figure 4.2 (page 52) illustrates the average change in fourth-grade MCAS ELA scores for students in a representative FAST-R and non-FAST-R school over the five pre-implementation years and then projects those trends into the post-implementation years.[8] For

[7]The average yearly trends for the SAT-9 scaled scores for both the representative FAST-R schools (0.20 scaled score point per year) and the representative non-FAST-R schools (0.29 scaled score point per year) during the five baseline years are positive, suggesting that all schools were showing some progress on this test in the years before the FAST-R program was implemented.

[8]To calculate the trends and intercepts for a representative FAST-R and non-FAST-R school, the student-level baseline characteristics were first centered by subtracting the school-level mean from each characteristic for each student. The model was run using these centered variables to find the trend and intercept for the average student in each school. The trend and intercept for each school were multiplied by the percentage of the total students in the school. The results for all FAST-R schools were summed to find the trend and intercept for a representative FAST-R school, and the results for all non-FAST-R schools were summed to find the average trend and intercept for a representative non-FAST-R school.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 4.2

Impacts on Fourth-Grade Students' Reading Test Scores

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (scaled score)	230.6	229.5			
Deviation from trend	-0.5	-1.3	0.8	0.06	0.499
MCAS ELA proficient or advanced (% of students)	25.2	21.6			
Deviation from trend	-0.6	-5.4	4.8	NA	0.190
SAT-9 Total Reading, fall fifth grade (scaled score)	634.1	632.7			
Deviation from trend	4.0	2.5	1.5	0.04	0.538
Second year after implemtation					
MCAS ELA (scaled score)	230.7	228.8			
Deviation from trend	-1.0	-2.8	1.8	0.14	0.200
MCAS ELA proficient or advanced (% of students)	27.3	24.6			
Deviation from trend	0.9	-3.8	4.7	NA	0.262
SAT-9 Total Reading, fall fifth grade (scaled score)	630.5	631.9			
Deviation from trend	-0.3	0.7	-1.0	-0.03	0.722
Sample size[c]			18,540		

(continued)

Table 4.2 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "first year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

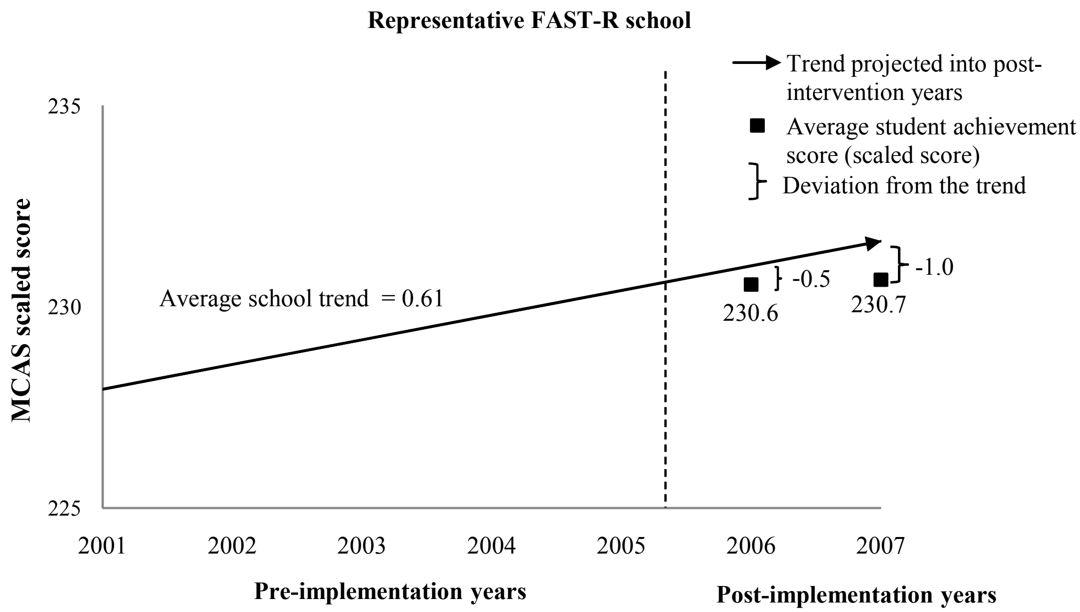
[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

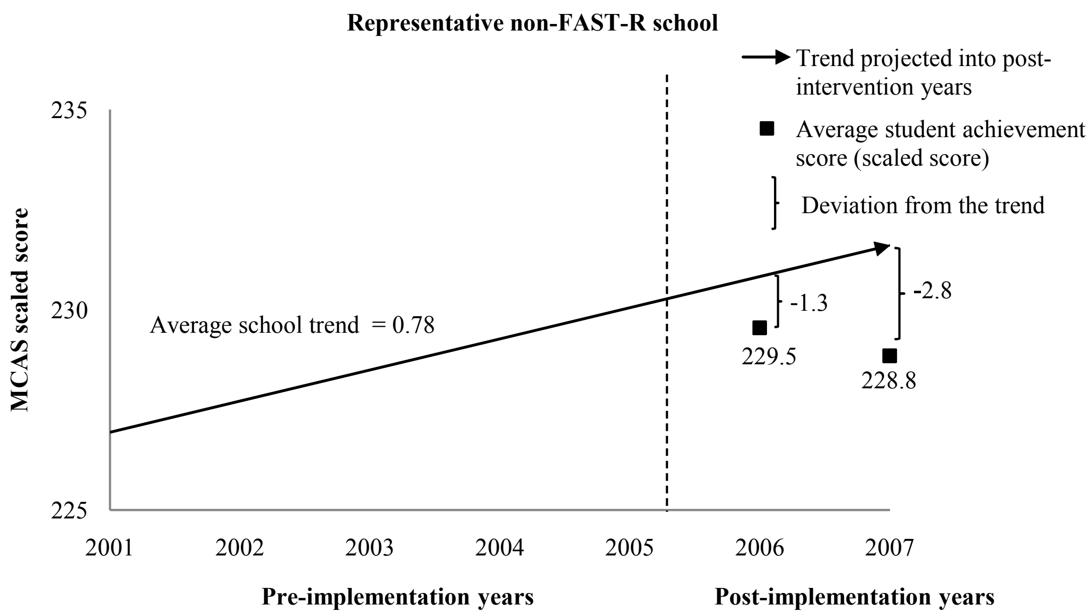
[c]The sample size includes all fourth-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 reading total is 15,816 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Figure 4.2

Trends in Fourth-Grade Students' MCAS ELA Scores





(continued)

Figure 4.2 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: To calculate the trends and intercepts for a representative FAST-R and non-FAST-R school, the student-level baseline characteristics were first centered by subtracting the school-level mean from the characteristic for each student. The model was run using these centered variables to find the trend and intercept for the average student in each school. The trend and intercept for each school was then multiplied by the percentage of the total students in the school. The results for all FAST-R schools were summed to find the trend and intercept for a "representative" FAST-R school, and the results for all non-FAST-R schools were summed to find the average trend and intercept for a "representative" non-FAST-R school.

The deviation from trend for each post-implementation year for the FAST-R schools is the regression-adjusted impact added to the non-FAST-R schools' deviation from trend. The average student achievement scores are calculated by subtracting the deviation from trend from the predicted average for a given follow-up year, calculated using the average school trend and intercept.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4.

students in both FAST-R and non-FAST-R schools, the trends are positive, suggesting a general growth in reading achievement for all students in the study. As with the third grade, the deviations from the trend for both FAST-R and non-FAST-R students are negative during both the 2005-2006 and the 2006-2007 school year. This finding suggests that students' scores were generally lower than would be expected given the trend in scores over the five years prior to implementation. Still, in both years of the intervention, the negative deviation from the trend is smaller for the students in FAST-R schools than for the students in non-FAST-R schools. The estimated impacts of FAST-R, as shown in Table 4.2, are positive in both years, but these impacts are not statistically significant. The estimated impact in the first year of implementation is 0.8 scaled score point, with an effect size of 0.06 standard deviation. The estimated impact for the second year after implementation is 1.8 scaled score points, with an effect size of 0.14 standard deviation. This effect size — 14 percent of a standard deviation between students in FAST-R and non-FAST-R schools — is notable but not statistically significant.

The estimated impacts of FAST-R on the percentage of students scoring at a proficient or advanced level on the MCAS are 4.8 percentage points and 4.7 percentage points, respectively, for the first and second years after implementation.[9] These estimated impacts are not statistically significant. The estimated impact for the SAT-9 Total Reading scaled score was 1.5 in the fall of 2006, but this impact was not statistically significant. The estimated impact for the SAT-9 Total Reading score in the fall of 2007 was –1.0 and was also not statistically signifi-

[9]The average yearly trends for the percentage of students who scored proficient or advanced on the MCAS for both the representative FAST-R and the representative non-FAST-R schools during the five baseline years are positive (0.63 percentage point and 1.52 percentage points per year, respectively), suggesting that, in general, schools were showing some progress in the years before the FAST-R program was implemented.

cant.[10] Overall, there are no statistically significant impacts of FAST-R on fourth-grade students' MCAS ELA or SAT-9 Total Reading scores.

Effects of the FAST-R Program on Students' Ability to Make Inferences and Find Evidence

Tables 4.3 and 4.4 display the estimated impacts of the FAST-R program on students' ability to answer two types of questions found on the MCAS, making inferences from and finding evidence in reading passages.[11] Table 4.3 displays the third-grade impacts. The first row shows FAST-R and non-FAST-R students' average distance from the state mean on the third-grade MCAS reading test for the set of questions related to making inferences.[12] Students in a representative FAST-R school, for instance, scored an average of 14.8 percentage points below the state average during the first school year after implementation. Students at a representative non-FAST-R school scored an average of 15.7 percentage points below the state average during the first year after the FAST-R program was implemented.[13] The estimated impact of FAST-R on students' ability to answer "making inferences" questions correctly during the first year after implementation is small and not statistically significant. The estimated impact on students' ability to answer "finding evidence" questions correctly is negative and also not statistically significant. In the second year after implementation, the estimated impact on students' ability to make inferences is larger, at 2.9 percentage points, with an effect size of 0.14 standard deviation. The estimated impact on students' ability to find evidence is 1.8 percentage points with an effect size of 0.08 standard deviation. Although the estimated impacts on making inferences are

[10]The average yearly trends for the SAT-9 Total Reading score for both the representative FAST-R and the representative non-FAST-R schools during the five baseline years are positive (0.68 and 0.92 scaled score points per year, respectively), suggesting that, in general, schools were showing some progress in the years before the FAST-R program was implemented.

[11]The comparative interrupted time series analysis of the questions on the MCAS that measure students' ability to make inferences and find evidence includes only three instead of five years of baseline scores in the trend (spring 2003 through spring 2005) because BPE has not classified the items on the spring 2001 and 2002 tests.

[12]The average distance from the state mean on questions that were designed to measure students' ability to making inferences and find evidence was calculated by subtracting the percentage of students who answered the question correctly across the state from each student's score on a particular question and then calculating each student's average score on all questions of a particular type.

[13]The average yearly trend for the questions related to making inferences and finding evidence for both the representative FAST-R schools (–1.8 points per year for making inferences and –0.34 points for finding evidence) and the representative non-FAST-R schools (–1.08 points for making inferences and –0.08 for finding evidence) during the five baseline years are negative, suggesting that all schools in the study were showing some decline on these test scores before the FAST-R program was implemented compared with the state average.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 4.3

Impacts on Third-Grade Students' Performance on MCAS "Making Inferences" and "Finding Evidence" Questions

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS "making inferences" questions	-14.8	-15.7			
Deviation from trend	2.2	1.3	1.0	0.05	0.628
MCAS "finding evidence" questions	-14.5	-14.5			
Deviation from trend	0.6	1.4	-0.8	-0.04	0.711
Second year after implementation					
MCAS "making inferences" questions	-13.1	-15.2			
Deviation from trend	5.7	2.8	2.9	0.14	0.284
MCAS "finding evidence" questions	-13.9	-16.1			
Deviation from trend	1.6	-0.2	1.8	0.08	0.527
Sample size[c]			13,715		

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using three years of data before the FAST-R program was implemented (the 2002-2003 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS = Massachusetts Comprehensive Assessment System.

The scores for the MCAS questions on making inferences and finding evidence are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two school years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

"Making inferences" and "finding evidence" questions are measured as the average percentage point deviation from the state mean.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students in the entire sample (by grade) during the 2002-2003 through 2004-2005 school years.

[c]The sample size includes all third-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the five years included in the analysis.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Table 4.4

Impacts on Fourth-Grade Students' Performance on MCAS "Making Inferences" and "Finding Evidence" Questions

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS "making inferences" questions	-9.7	-11.2			
Deviation from trend	0.9	0.0	0.8	0.04	0.708
MCAS "finding evidence" questions	-10.4	-11.0			
Deviation from trend	1.3	1.9	-0.6	-0.03	0.806
Second year after implementation					
MCAS "making inferences" questions	-10.9	-13.5			
Deviation from trend	-1.5	-3.4	2.0	0.10	0.501
MCAS "finding evidence" questions	-13.4	-17.3			
Deviation from trend	-1.5	-4.2	2.7	0.12	0.407
Sample size[c]			12,722		

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using three years of data before the FAST-R program was implemented (the 2002-2003 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS = Massachusetts Comprehensive Assessment System.

The scores for the MCAS questions on making inferences and finding evidence are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

"Making inferences" and "finding evidence" questions are measured as the average percentage point deviation from the state mean.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students in the entire sample (by grade) during the 2002-2003 through 2004-2005 school years.

[c]The sample size includes all fourth-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the five years included in the analysis.

larger than the estimated impacts on finding evidence, neither shows a statistically significant difference from zero.

Table 4.4 displays the estimated impacts of FAST-R on fourth-grade students' ability to make inferences from and find evidence in text.[14] There are no statistically significant impacts for fourth-grade students on either of these types of questions. In the second year after implementation, the magnitudes of the impacts for questions related to making inferences and finding evidence are similar, at 2.0 and 2.7, respectively, suggesting that the effects of FAST-R on fourth-grade students' ability to answer these different types of questions are similar.

Effects of the FAST-R Program on Students by Gender, Socioeconomic Status, Performance on a Pre-Program Reading Test, and Special Education Status

The general findings for the analyses of student subgroups were similar to those of the overall impact analyses.[15] There were no statistically significant impacts found for students of lower or higher socioeconomic status or students performing below or above proficiency on a pre-program reading test. The impacts on the percentage of fourth-grade boys who scored at the proficient or advanced level on the MCAS are 8.7 percentage points during the first year after implementation and 7.5 percentage points during the second year after implementation. Both these impacts are statistically significant. However, the difference in impacts for girls and boys is not statistically significant, and, therefore, the higher impact for boys should be interpreted with caution. The impact on MCAS ELA total scores for third-graders receiving special education services is statistically significant during the second year after implementation, but the difference in impacts on this measure for students receiving special education services versus students who are not receiving such services is not statistically significant and, again, should be viewed with caution. For fourth-grade students, there is a statistically significant impact on both the MCAS ELA score and the percentage of students who tested at a proficient or advanced level during the second year after implementation among students not receiving special education services. The difference between impacts on students receiving special education services and students who did not receive such services is not statistically significant for the MCAS ELA but is significant for the percentage of students performing at a proficient or advanced level.

[14]The average yearly trend for the questions related to making inferences is positive for both the representative FAST-R schools (1.09 points per year) and the representative non-FAST-R schools (1.15 points per year) during the five baseline years. The average yearly trend for questions on finding evidence is slightly negative (–0.13 point for FAST-R students and –0.28 point for non-FAST-R students).

[15]Subgroup analyses at the school level were also attempted, including comparing schools where teachers reported spending more or less time with a FAST-R data coach and comparing schools that the Boston Plan for Excellence designated as high-implementers with schools that were still initiating their implementation at the time of the study. Neither of these cases had an adequate number of schools in each subgroup to conduct the analysis.

There is also a statistically significant impact on the SAT-9 Total Reading score during the second year after implementation for students receiving special education services. The difference in impacts for this measure is also statistically significant. Still, the inconsistency of these impacts (showing up whether or not students receive special education services), and the small number of statistically significant impacts over all the analyses in this study suggest that these impacts should be interpreted cautiously.[16]

[16]See Appendix C for a more detailed description of the findings for each of the student subgroups.

Chapter 5

Reflections and Conclusions

The findings in the preceding chapters present a mixed picture of the Formative Assessments of Student Thinking in Reading (FAST-R) intervention. The program was generally implemented as intended, and third- and fourth-grade teachers at schools participating in the initiative believed that they had benefited from the program. At least two-thirds of the teachers at these schools who responded to a survey administered as part of the MDRC evaluation reported that they had used FAST-R to assess students' general reading comprehension and particular reading skills, and that it had helped them understand how to use data to inform their teaching practices.

On the other hand, the large majority of these teachers received only a few hours of FAST-R coaching over the course of the 2006-2007 school year. Moreover, their responses, as well as those of their counterparts at a set of matched comparison schools, make it clear that FAST-R was only one among many types of professional development activities that teachers in the Boston Public Schools system (BPS) received during the study period. BPS has made instructional improvement a district-wide goal and has publicly proclaimed that using data to identify student needs, improve teaching, and assess progress is one of "The Six Essentials for Whole-School Improvement." It is not surprising, therefore, that teachers at the comparison schools also reported getting a good deal of professional development assistance, as well as spending time analyzing data, including data from formative assessments (albeit of a type different from the formative assessments used in FAST-R).

The central question posed by the impact analysis is whether students at FAST-R schools registered gains in reading scores that were significantly greater than the gains experienced by similar students at the comparison schools, whose teachers were not exposed to FAST-R. The answer, in general, is that they did not. Differences in the achievement gains registered by students at the two sets of schools were generally small and were not statistically significant (that is, they could have arisen as a result of chance).

But if the numbers are relatively straightforward, what readers should make of the numbers is not. Complicating efforts at interpretation is the fact that the information about both the amount of professional development that teachers received and other implementation matters is grounded in an imperfect source of data, the teacher survey. The response rate to the teacher survey was low — only 54 percent overall — and it is impossible to know whether the responses of teachers who completed the survey are an accurate representation of their counterparts who did not complete it.[1] To prompt further reflection on the issues, however, the following discussion assumes that the survey responses do present an accurate picture of conditions at the FAST-R and comparison schools, and considers the findings and several possible explanations in light of that assumption. These explanations must necessarily be regarded as speculative, however.

Since the FAST-R program did not, in general, show an improvement over the status quo, one possible conclusion is that FAST-R is simply not an effective tool for raising student achievement — that is, that it would not work under any circumstances. But it is also possible that the specific conditions under which it operated in Boston undercut the likelihood that a potentially powerful intervention would show significant impacts.

Thus, one plausible conclusion is that the intervention did not have positive impacts because it was not intensive enough — that is, teachers did not receive a large enough "dose" of it for it to affect teaching and learning in meaningful ways. More than two-thirds of the teachers at FAST-R schools reported receiving five or fewer hours of FAST-R training over the course of the school year, and almost 10 percent of these teachers reported receiving no FAST-R professional development at all. Teachers may not have had enough training in how to assess students' learning needs and remediate instruction to address those needs.[2]

One way of testing this hypothesis would be to see whether FAST-R produced larger impacts on student achievement at those schools where teachers received a larger dose of the treatment. Unfortunately, such an analysis is infeasible because there were only three schools where teachers reported, on average, receiving 11 or more hours of FAST-R coaching. This number is too small to permit results to be reliable and generalizable.

In any case, it should be noted that the Boston Plan for Excellence (BPE), the nonprofit organization that developed and runs FAST-R, would need to change its mode of operation in order to put in place a more intensive intervention at all schools. It would need to increase the number of instructional data coaches in order to provide more professional development to

[1]As noted in Chapter 2, the low response rate is probably attributable, at least in part, to the fact that contract negotiations had strained relationships between the teacher's union and the district, and some teachers may have been unwilling to take on any efforts unrelated to their official duties.

[2]It is also possible that the FAST-R program taught teachers how to analyze data and respond to students' individual instructional needs, but that teachers did not consistently put these lessons into practice in their classrooms. This evaluation did not gather the systematic evidence that would allow researchers to examine this potential explanation. While the study cannot shed light on the matter, the explanation itself suggests that professional development is only the first step in a series of activities leading to instructional improvement. Teachers must use what they have learned in planning group lessons and in responding to individual students' learning needs. School administrators, in their role as instructional leaders, must go beyond arranging the delivery of professional development services to their teachers to monitor the effectiveness with which that professional development is translated into practice.

more teachers. It would also need to change its way of working with schools. As FAST-R operated during the study period, each school could elect to receive as much or as little FAST-R coaching as it wanted, in line with its perceived needs and the availability of time for this particular form of professional development. Schools signing up for FAST-R were not necessarily looking to add many more hours of professional development to their teachers' schedules. If its aim were to increase the amount of professional development teachers receive, BPE would have to take a more proactive and forceful stance and leave less to the discretion of individual schools.

Another plausible explanation for the absence of impacts is that while FAST-R may have increased student achievement over time, so did the other professional development efforts in place in the Boston Public Schools, which also included training in how to examine data. The survey findings indicate that teachers at the comparison schools received at least as much professional development of several different kinds as did teachers at the FAST-R schools. Teachers at the comparison schools also reported spending more time analyzing the results of the state reading assessment, and as much time studying formative assessment data, as did teachers at the FAST-R schools. In other words, the FAST-R "treatment" was not unique; while it may have represented a contribution to teachers' knowledge and skills, this contribution was not very different from what teachers would have received had they and their schools not participated in FAST-R.

There is, however, another way of reading the data, one that suggests that FAST-R may have worked *better* than other training efforts and was, in fact, an "energy-efficient" and effective form of professional development. In general, teachers at FAST-R schools reported spending less time on professional development than their counterparts at the comparison schools — but students at the two groups of schools nonetheless experienced the same learning gains.

Thus, the data presented in Table 3.2 indicate that for every type of professional development shown, a smaller proportion of non-FAST-R teachers reported receiving no professional development of that type, and a larger proportion reported receiving 11 or more hours of such professional development. For only one type of professional development — training specifically related to the curriculum — is the difference between the two groups of teachers statistically significant, but the pattern is consistent across all types. The same pattern holds with respect to examination of the Massachusetts Comprehensive Assessment System (MCAS) data: According to Table 3.5, non-FAST-R teachers were more likely to examine these data at all, and spent more hours doing so, than teachers in the FAST-R schools.

If more professional development is associated with better instruction and (as several studies have shown) better instruction is associated with better student outcomes, then achievement gains at the non-FAST-R schools might be expected to surpass those at the FAST-R schools.[3] But they did not. It may be that the limited amount of professional development that BPE provided to the FAST-R schools was more useful to teachers than more diffuse professional development efforts that took up more time. In judging the usefulness of FAST-R vis-à-vis other forms of professional development, it would be useful to have information about the costs of various professional development efforts. Such data were not collected as part of this study.

This study has produced the best available data about FAST-R's impacts, or the absence of such impacts. But, as the preceding discussion indicates, the meaning of those data is very much open to discussion. Even if the correct interpretation were clear and incontrovertible, however, the study would not yield definitive evidence about the potential and policy relevance of the FAST-R approach; indeed, a single study in a single site is insufficient to establish whether *any* approach can be replicated with strong results and is therefore policy-relevant. It is the researchers' hope that training in the use of formative assessments to guide instruction will be tested in many variations and in many places, in order to build a cumulative body of evidence about whether and how use of these assessments promotes better teaching and learning in America's schools.

[3] Brophy (1986).

Appendix A

The Analytic Model Used in the FAST-R Impact Analysis

The analytic model used to estimate impacts was structured to calculate each school's deviation from its respective trend, constructed using the five years of baseline data. It then compares the program schools' deviations with the deviations of the comparison schools to estimate impacts for the overall program, by grade.[1] The executed model is presented below:

$$Y_{it} = \Sigma P_j * (F_t) + \Sigma S_j * (X_{ij}) + \Sigma S_j + \Sigma S_j * (T_t) + \Sigma F_t + \upsilon_t + e_i$$

Where:

Y_{it}	=	Student outcome for student, *i*, at time, *t*
P_j	=	Dummy variable indicating whether school, *j*, is a program school
F_t	=	Follow-up year at time, *t*
X_{ij}	=	Student-level baseline characteristics for student, *i,* in school, *j*
S_j	=	School, *j*, fixed effect
T_t	=	Trend at time, *t*
υ_t	=	Random error term for cohort at time, *t*
e_i	=	Random error term for student, *i*

As noted in the analytic model above, the impact estimation takes advantage of the student-level data and controls for variation in student characteristics through the use of covariates. Student-level covariates include age, gender, race/ethnicity, special education status, Limited English Proficient (LEP) or formerly LEP status, free or reduced-price lunch status, and Stanford Achievement Test, version 9 (SAT-9), Total Reading scaled score from the beginning of the third-grade school year.

The coefficient on the program variable and its t-statistic provide a test of the statistical significance of the impact of the FAST-R program at follow-up time, *t*. This is essentially the difference between the deviation from the trend for the FAST-R schools and the deviation from the trend for the non-FAST-R schools. The error terms in this model account for the hierarchical nature of the data, with students nested within cohorts.

[1]All comparison (non-FAST-R) schools matched to particular FAST-R schools were weighted equally so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed to create the weight given to each comparison school student.

Appendix B

List of FAST-R and Non-FAST-R Schools

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table B.1

FAST-R and Non-FAST-R Schools: Third-Grade Sample

	FAST-R Schools	Non-FAST-R Schools
	James J. Chittick Elementary School	Dante Alighieri Elementary School
	George H. Conley Elementary School	Phineas Bates Elementary School
	David A. Ellis Elementary School	Ludwig van Beethoven Elementary School
	Edward Everett Elementary School	William E. Channing Elementary School
	Dennis Haley Elementary School	Roger Clap Elementary School
	Harvard/Kent Elementary School	Paul A. Dever Elementary School
	Rafael Hernández K-8 Elementary School	Quincy E. Dickerman Elementary School
	Jackson/Mann K-8 School	John Eliot K-8 School
	Mary Lyon K-8 School	Ralph Waldo Emerson Elementary School
	Joseph P. Manning Elementary School	David G. Farragut Elementary School
	John Marshall Elementary School	Emily A. Fifield Elementary School
	Samuel W. Mason Elementary School	Elihu Greenwood Elementary School
	Mather Elementary School	Henry Grew Elementary School
	Mattahunt Elementary School	James W. Hennigan Elementary School
	Donald McKay K-8 School	Henry L. Higginson Elementary School
	John D. Philbrick Elementary School	John P. Holland Elementary School
	Franklin D. Roosevelt K-8 School	Joseph J. Hurley K-8 School
	William E. Russell Elementary School	John F. Kennedy Elementary School
	Warren/Prescott K-8 School	Patrick J. Kennedy Elementary School
	F. Lyman Winship Elementary School	Thomas J. Kenny Elementary School
		Joseph Lee Elementary School
		Wolfgang A. Mozart Elementary School
		Patrick O'Hearn Elementary School
		William Ohrenberger Elementary School
		Pauline A. Shaw Elementary School
		Lucy Stone Elementary School
		Charles Sumner Elementary School
		Charles H. Taylor Elementary School
		William Monroe Trotter Elementary School
		Joseph P. Tynan Elementary School
		John Winthrop Elementary School
Sample size	20	31

SOURCE: School names cited are from the Boston Public Schools online school directory.

NOTE: FAST-R schools were matched with comparison non-FAST-R schools based on similarities in student demographics and achievement scores. Some comparison schools were matched to multiple FAST-R schools.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table B.2

FAST-R and Non-FAST-R Schools: Fourth-Grade Sample

	FAST-R Schools	Non-FAST-R Schools
	James J. Chittick Elementary School	Phineas Bates Elementary School
	George H. Conley Elementary School	Ludwig van Beethoven Elementary School
	David A. Ellis Elementary School	Roger Clap Elementary School
	Edward Everett Elementary School	Paul A. Dever Elementary School
	Dennis C. Haley Elementary School	Quincy E. Dickerman Elementary School
	Harvard/Kent Elementary School	John Eliot K-8 School
	Rafael Hernández K-8 Elementary School	David G. Farragut Elementary School
	Jackson/Mann K-8 School	Emily A. Fifield Elementary School
	John Marshall Elementary School	Elihu Greenwood Elementary School
	Samuel W. Mason Elementary School	Sarah Greenwood K-8 School
	Mather Elementary School	Henry Grew Elementary School
	Mattahunt Elementary School	Nathan Hale Elementary School
	Donald McKay K-8 School	Alexander Hamilton Elementary School
	John D. Philbrick Elementary School	John P. Holland Elementary School
	Franklin D. Roosevelt K-8 School	Joseph J. Hurley K-8 School
	William E. Russell Elementary School	John F. Kennedy Elementary School
	Maurice J. Tobin K-8 School	Thomas J. Kenny Elementary School
	Warren/Prescott K-8 School	Joseph Lee Elementary School
	F. Lyman Winship Elementary School	Wolfgang A. Mozart Elementary School
		William Ohrenberger Elementary School
		Michael J. Perkins Elementary School
		Oliver Hazard Perry K-8 School
		Pauline A. Shaw Elementary School
		Lucy Stone Elementary School
		Charles Sumner Elementary School
		Charles H. Taylor Elementary School
		William Monroe Trotter Elementary School
		Joseph P. Tynan Elementary School
		John Winthrop Elementary School
Sample size	19	29

SOURCE: School names cited are from the Boston Public Schools online school directory.

NOTE: FAST-R schools were matched with comparison non-FAST-R schools based on similarities in student demographics and achievement scores. Some comparison schools were matched to multiple FAST-R schools.

Appendix C

Subgroup Analyses of the Effects of the FAST-R Program

To better understand how the FAST-R program affected different groups of students, several subgroup analyses were conducted, including examining the effects of FAST-R on boys and girls, on students of lower and higher socioeconomic status as defined by whether or not a student was receiving free or reduced-price lunch, on students who performed below or above proficiency on a pre-program reading test, and on whether or not a student was receiving special education services. This appendix describes the findings of those analyses.

Effects of the FAST-R Program on Students by Gender

Tables C.1 and C.2 display the effects of FAST-R on third-grade boys' and third-grade girls' achievement in reading. Although there are no statistically significant impacts on the Massachusetts Comprehensive Assessment System (MCAS) results for third-grade boys, the general pattern of estimated impacts for boys is positive. As shown in Table C.1, the estimated effect (deviation from trend) of FAST-R on third-grade boys' distance from proficiency on the MCAS English Language Arts (ELA) test during the second year of implementation is 1.3 raw score points, with an effect size of 0.16 standard deviation. As shown in Table C.2, the girls' estimated impact for the MCAS distance from proficiency is negative in the first year of implementation and positive but small in the second year after implementation (0.6 raw score point, with an effect size of 0.07 standard deviation). Still, a test of the difference in the impacts for boys and the impacts for girls shows that this difference is not statistically significant, and, therefore, it cannot be concluded that the impacts are different from each other.[1]

Tables C.3 and C.4 display the effects of FAST-R on fourth-grade boys' and fourth-grade girls' reading achievement. In Table C.3, there is a statistically significant impact on the percentage of fourth-grade boys who scored at the proficient or advanced level on the MCAS during both the first and second year of implementation, with an estimated impact of 8.7 percentage points during the first year after implementation and 7.5 percentage points during the second year after implementation. However, the difference in impacts for girls and boys is not statistically significant and, therefore, the higher impact on boys should be interpreted with caution. Although not statistically significant, the estimated impact on the MCAS ELA score for fourth-grade girls during the second year after implementation, shown in Table C.4, is notable at 2.4 scaled score points, with an effect size of 0.19 standard deviation. The difference in impacts between boys and girls on this measure is also not statistically significant.

[1]A t-test of the difference in impacts between boys and girls was done for each outcome. None of the outcomes showed statistically significant differences between the impacts for boys and girls, suggesting that caution should be used when interpreting the difference in the magnitude of impacts between these groups.

Effects of the FAST-R Program on Students by Socioeconomic Status

Approximately 80 percent of students in both the FAST-R and comparison schools were receiving free or reduced-price lunch, an indication of lower socioeconomic status. Tables C.5 and C.6 describe the effects of FAST-R on third-grade students' achievement in reading by students' socioeconomic status, and Tables C.7 and C.8 show the outcomes for fourth-grade students. In general, the patterns of effects for third- and fourth-grade students are similar to those of all students. The estimated impacts on the MCAS and the Stanford Achievement Test, version 9 (SAT-9) Total Reading scores for third-grade students receiving free and reduced-price lunch, as shown in Table C.5, are small and not statistically significant. The estimated impact on third-grade students' distance from proficiency during the first year after implementation of FAST-R is 0.0 raw score points while in the second year after implementation the estimated impact is 0.9 raw score points with an effect size of 0.11 standard deviation, but is not statistically significant. The estimated impacts on SAT-9 scaled scores for third-grade students receiving free or reduced price lunch are negative and not statistically significant. The estimated impacts on third-grade students not receiving free or reduced-price lunch, as shown in Table C.6, are negative and not statistically significant.

The effects of FAST-R on low-income fourth-grade students' MCAS scaled scores, as shown in Table C.7, are positive but are not statistically significant. During the first year after implementation, the estimated impact on the MCAS ELA test is 0.9 scaled score point, with an effect size of 0.07 standard deviation. The magnitude of the estimated impact during the second year is 2.3 scaled score points, with an effect size of 0.18 and a p-value of 0.106. The impacts of FAST-R for fourth-grade students not receiving free or reduced-price lunch, as shown in Table C.8, are, generally, small and not statistically significant. Although the magnitude of the impact on MCAS ELA score during the second year is notable for students of a lower socioeconomic status, the difference in impacts between the two groups is not statistically significant; thus, this finding should be viewed with caution.

Effects of the FAST-R Program on Students by Baseline Reading Achievement

Since the FAST-R program trains teachers to focus on individual student ability levels in reading, it is possible that FAST-R may be particularly successful with students at certain skill levels. Tables C.9 and C.10 display the effects of FAST-R on fourth-graders' reading achievement by student performance on the SAT-9 test during the fall of their third-grade year, before the FAST-R program was implemented.[2] Table C.9 displays the effects of FAST-R on students who performed below the proficient level. A majority of the students in the study sample were scoring at this level, and the impacts are similar to those for the full sample of students. Although there are no statistically significant impacts, the magnitude of the impact on the MCAS scaled score during the second year of implementation is notable at 1.9 scaled score points, with an effect size of 0.14 standard deviation. The effects on students performing at or above proficient, as shown in Table C.10, are generally small and not statistically significant for MCAS, and negative and not statistically significant for SAT-9. The difference in impacts between these two groups of students is not statistically significant for any of the outcomes.

Effects of the FAST-R Program on Students by Special Education Status

Tables C.11 and C.12 display the effects of FAST-R on third-grade students' achievement in reading by students' special education status. The effect of FAST-R during the second year after implementation was statistically significant for students receiving special education services. A test of the difference in the impacts for students receiving special education services and those not receiving any special education services shows that this difference is not statistically significant. Therefore, it cannot be concluded that the impacts are different from each other.

Tables C.13 and C.14 display the effects of FAST-R on fourth-grade students' achievement in reading by students' special education status. During the second year after implementation, there are statistically significant impacts on students' MCAS Total Reading scores (2.8 scaled score points, with an effect size of 0.22 standard deviation) and the percentage of students scoring at the proficient or advanced level on the MCAS (9.4 percentage points) for students not receiving special education services. The difference in impacts for students receiving and not receiving special education services is not statistically significant for the MCAS reading total measure but is significant for the percent of students scoring at a proficient or advanced level. On the other hand, there is a statistically significant impact on the SAT-9 reading total score during the second year after implementation for students receiving special education services. The difference in impacts on this outcome is also statistically significant. Although these impacts are of a notable magnitude, they do not offer a coherent story. Since there are positive and conflicting impacts on students both receiving and not receiving special education ser-

[2]Only students who took the SAT-9 test in the fall of their third-grade year are included in this analysis, approximately 74 percent of the students in the sample. These students likely represent a less mobile group of students, since a principal reason students would not have taken the SAT-9 test in third grade is that they were not enrolled in the district at that time. A sensitivity test was conducted where the analysis was run on only the students with SAT-9 pretest scores available. Patterns of impacts on this group of students were similar to those for the full sample.

vices, these impacts seem to be random and may be the consequence of testing error. The small number of statistically significant impacts of any magnitude over all the analyses in this study, coupled with the fact that the impacts on these subgroups are significant only at the 10 percent level (meaning there is a 10 percent chance that these findings are random and not due to actual differences in outcomes), suggests that these impacts should be interpreted with caution.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.1

Impacts on Third-Grade Boys' Reading Test Scores

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (distance from proficiency)	-8.6	-9.4			
Deviation from trend	-2.0	-2.8	0.8	0.10	0.273
MCAS ELA proficient or above (% of students)	24.2	23.6			
Deviation from trend	-4.4	-4.2	-0.2	NA	0.950
SAT-9 Total Reading, fall fourth grade (scaled score)	600.4	600.2			
Deviation from trend	0.9	2.1	-1.1	-0.03	0.725
Second year after implementation					
MCAS ELA (distance from proficiency)	-7.4	-8.6			
Deviation from trend	-0.7	-2.1	1.3	0.16	0.125
MCAS ELA proficient or above (% of students)	25.4	23.2			
Deviation from trend	-3.4	-4.9	1.4	NA	0.727
SAT-9 Total Reading, fall fourth grade (scaled score)	600.8	600.4			
Deviation from trend	1.4	1.9	-0.5	-0.01	0.887
Sample size[c]			10,427		

(continued)

Appendix Table C.1 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in third grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fourth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in third grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fourth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or above outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all third-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 8,479 students and includes all third-grade students with SAT-9 scores reported in the fall of their fourth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.2

Impacts on Third-Grade Girls' Reading Test Scores

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (distance from proficiency)	-6.8	-6.7			
Deviation from trend	-2.3	-1.7	-0.6	-0.08	0.417
MCAS ELA proficient or above (% of students)	31.4	32.4			
Deviation from trend	-6.2	-2.6	-3.6	NA	0.397
SAT-9 Total Reading, fall fourth grade (scaled score)	609.3	611.7			
Deviation from trend	0.1	5.1	-5.1	-0.13	0.106
Second year after implementation					
MCAS ELA (distance from proficiency)	-5.9	-7.0			
Deviation from trend	-1.4	-2.0	0.6	0.07	0.536
MCAS ELA proficient or above (% of students)	32.9	29.1			
Deviation from trend	-5.5	-6.4	0.9	NA	0.848
SAT-9 Total Reading, fall fourth grade (scaled score)	609.0	607.8			
Deviation from trend	-0.9	0.9	-1.8	-0.05	0.612
Sample size[c]			9,585		

(continued)

Appendix Table C.2 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in third grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fourth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in third grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fourth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or above outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all third-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 8,055 students and includes all third-grade students with SAT-9 scores reported in the fall of their fourth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.3

Impacts on Fourth-Grade Boys' Reading Test Scores

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (scaled score)	228.7	226.9			
Deviation from trend	-0.5	-1.6	1.1	0.08	0.409
MCAS ELA proficient or advanced (% of students)	22.3	16.3			
Deviation from trend	2.1	-6.6	8.7	NA **	0.018
SAT-9 Total Reading, fall fifth grade (scaled score)	630.2	627.8			
Deviation from trend	3.7	0.9	2.8	0.08	0.371
Second year after implementation					
MCAS ELA (scaled score)	228.6	226.7			
Deviation from trend	-1.3	-2.5	1.2	0.09	0.423
MCAS ELA proficient or advanced (% of students)	22.2	18.5			
Deviation from trend	1.4	-6.0	7.5	NA *	0.080
SAT-9 Total Reading, fall fifth grade (scaled score)	627.3	627.1			
Deviation from trend	0.2	-1.1	1.3	0.03	0.724
Sample size[c]			9,512		

(continued)

Appendix Table C.3 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all fourth-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 7,957 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.4

Impacts on Fourth-Grade Girls' Reading Test Scores

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (scaled score)	232.6	232.2			
Deviation from trend	-0.6	-1.2	0.6	0.05	0.662
MCAS ELA proficient or advanced (% of students)	29.5	26.8			
Deviation from trend	-2.1	-4.7	2.6	NA	0.605
SAT-9 Total Reading, fall fifth grade (scaled score)	638.3	637.2			
Deviation from trend	3.5	3.2	0.3	0.01	0.920
Second year after implementation					
MCAS ELA (scaled score)	233.4	231.5			
Deviation from trend	-0.2	-2.7	2.4	0.19	0.140
MCAS ELA proficient or advanced (% of students)	34.2	31.5			
Deviation from trend	1.9	-1.3	3.2	NA	0.572
SAT-9 Total Reading, fall fifth grade (scaled score)	634.3	637.7			
Deviation from trend	-1.3	3.1	-4.4	-0.12	0.211
Sample size[c]			8,921		

(continued)

Appendix Table C.4 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all fourth-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 7,600 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.5

Impacts on Third-Grade Students' Reading Test Scores for Students Receiving Free or Reduced-Price Lunch

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (distance from proficiency)	-8.4	-8.6			
Deviation from trend	-2.5	-2.5	0.0	0.00	0.960
MCAS ELA proficient or above (% of students)	24.5	25.6			
Deviation from trend	-5.8	-4.2	-1.6	NA	0.642
SAT-9 Total Reading, fall fourth grade (scaled score)	601.8	602.6			
Deviation from trend	1.5	2.7	-1.1	-0.03	0.676
Second year after implementation					
MCAS ELA (distance from proficiency)	-7.3	-8.3			
Deviation from trend	-1.4	-2.3	0.9	0.11	0.281
MCAS ELA proficient or above (% of students)	26.8	24.2			
Deviation from trend	-4.0	-6.1	2.1	NA	0.594
SAT-9 Total Reading, fall fourth grade (scaled score)	601.7	601.6			
Deviation from Trend	1.2	1.3	-0.1	0.00	0.987
Sample size[c]			17,033		

(continued)

Appendix Table C.5 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in third grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fourth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in third grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fourth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students in the entire sample (by grade) during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or above outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all third-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is13,955 students and includes all third-grade students with SAT-9 scores reported in the fall of their fourth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.6

Impacts on Third-Grade Students' Reading Test Scores for Students Not Receiving Free or Reduced-Price Lunch

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (distance from proficiency)	-5.1	-5.8			
Deviation from trend	-0.8	-0.4	-0.4	-0.05	0.673
MCAS ELA proficient or above (% of students)	42.8	37.6			
Deviation from trend	-3.7	0.0	-3.7	NA	0.570
SAT-9 Total Reading, fall fourth grade (scaled score)	619.2	617.1			
Deviation from trend	-0.1	5.9	-6.0	-0.16	0.274
Second year after implementation					
MCAS ELA (distance from proficiency)	-5.5	-5.7			
Deviation from trend	-1.1	0.0	-1.1	-0.13	0.376
MCAS ELA proficient or above (% of students)	37.1	34.5			
Deviation from trend	-10.1	-3.2	-6.9	NA	0.371
SAT-9 Total Reading, fall fourth grade (scaled score)	617.7	615.7			
Deviation from trend	-1.6	4.4	-6.0	-0.16	0.341
Sample size[c]			2,979		

(continued)

Appendix Table C.6 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in third grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fourth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in third grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fourth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students in the entire sample (by grade) during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or above outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all third-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 2,579 students and includes all third-grade students with SAT-9 scores reported in the fall of their fourth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.7

Impacts on Fourth-Grade Students' Reading Test Scores for Students Receiving Free or Reduced-Price Lunch

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (scaled score)	229.9	228.9			
Deviation from trend	-0.2	-1.0	0.9	0.07	0.477
MCAS ELA proficient or advanced (% of students)	23.5	19.8			
Deviation from trend	0.3	-3.9	4.3	NA	0.251
SAT-9 Total Reading, fall fifth grade (scaled score)	631.9	630.3			
Deviation from trend	3.9	2.9	1.0	0.03	0.689
Second year after implementation					
MCAS ELA (scaled score)	230.3	228.0			
Deviation from trend	-0.3	-2.7	2.3	0.18	0.106
MCAS ELA proficient or advanced (% of students)	25.1	21.5			
Deviation from trend	1.4	-3.4	4.8	NA	0.266
SAT-9 Total Reading, fall fifth grade (scaled score)	628.3	629.4			
Deviation from trend	-0.4	1.0	-1.4	-0.04	0.629
Sample size[c]			15,872		

(continued)

Appendix Table C.7 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all fourth-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 13,395 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.8

Impacts on Fourth-Grade Students' Reading Test Scores for Students Not Receiving Free or Reduced-Price Lunch

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (scaled score)	234.3	233.1			
Deviation from trend	-1.7	-2.8	1.1	0.08	0.618
MCAS ELA proficient or advanced (% of students)	37.5	31.3			
Deviation from trend	-1.0	-12.1	11.2	NA	0.159
SAT-9 Total Reading, fall fifth grade (scaled score)	646.6	642.5			
Deviation from trend	2.1	-0.7	2.8	0.07	0.653
Second year after implementation					
MCAS ELA (scaled score)	234.4	233.9			
Deviation from trend	-2.6	-3.2	0.5	0.04	0.829
MCAS ELA proficient or advanced (% of students)	42.0	42.9			
Deviation from trend	1.9	-4.0	5.8	NA	0.524
SAT-9 Total Reading, fall fifth grade (scaled score)	644.8	644.3			
Deviation from trend	-1.2	0.5	-1.7	-0.05	0.811
Sample size[c]			2,561		

(continued)

Appendix Table C.8 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all fourth-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 2,162 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.9

Impacts on Fourth-Grade Students' Reading Test Scores for Students Performing Below Proficient on SAT-9 Pretest

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year of implementation					
MCAS ELA (scaled score)	228.1	226.9			
Deviation from trend	-1.1	-0.8	-0.3	-0.02	0.842
MCAS ELA proficient or advanced (% of students)	16.2	14.3			
Deviation from trend	-3.5	-2.9	-0.6	NA	0.876
SAT-9 Total Reading, fall fifth grade (scaled score)	627.7	622.2			
Deviation from trend	1.9	1.7	0.2	0.00	0.953
Second year of implementation					
MCAS ELA (scaled score)	228.8	225.2			
Deviation from trend	-1.2	-3.0	1.9	0.14	0.214
MCAS ELA proficient or advanced (% of students)	20.3	13.9			
Deviation from trend	-0.6	-4.4	3.9	NA	0.417
SAT-9 Total Reading, fall fifth grade (scaled score)	624.9	621.8			
Deviation from trend	-1.9	1.0	-2.9	-0.08	0.362
Sample size[c]			10,392		

(continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all fourth-grade students in the FAST-R and non-FAST-R schools during the seven years included in the analysis who performed below the proficiency threshold on the SAT-9 during the fall of their third-grade year and had reported MCAS scores in fourth grade. The sample size for the SAT-9 Total Reading is 8,785 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.10

Impacts on Fourth-Grade Students' Reading Test Scores for Students Performing At or Above Proficient on SAT-9 Pretest

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year of implementation					
MCAS ELA (scaled score)	244.2	242.0			
Deviation from trend	-0.3	-0.6	0.3	0.02	0.889
MCAS ELA proficient or advanced (% of students)	68.4	53.9			
Deviation from trend	2.9	-8.6	11.5	NA	0.162
SAT-9 Total Reading, fall fifth grade (scaled score)	670.6	673.3			
Deviation from trend	1.6	7.2	-5.6	-0.18	0.237
Second year of implementation					
MCAS ELA (scaled score)	244.4	242.2			
Deviation from trend	-0.4	-0.7	0.2	0.02	0.925
MCAS ELA proficient or advanced (% of students)	67.2	60.8			
Deviation from trend	2.0	-2.4	4.4	NA	0.648
SAT-9 Total Reading, fall fifth grade (scaled score)	667.6	669.0			
Deviation from trend	-0.8	3.9	-4.7	-0.15	0.393
Sample size[c]			3,295		

(continued)

Appendix Table C.10 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all fourth-grade students in the FAST-R and non-FAST-R schools during the seven years included in the analysis who performed at or above the proficiency threshold on the SAT-9 during the fall of their third-grade year and had reported MCAS scores in fourth grade. The sample size for the SAT-9 Total Reading is 2,950 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.11

Impacts on Third-Grade Students' Reading Test Scores for Students Receiving Special Education Services

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (distance from proficiency)	-14.4	-14.7			
Deviation from trend	-3.0	-3.3	0.3	0.04	0.781
MCAS ELA proficient or above (% of students)	9.2	11.5			
Deviation from trend	-4.4	-2.9	-1.6	NA	0.737
SAT-9 Total Reading, fall fourth grade (scaled score)	577.6	573.4			
Deviation from trend	5.0	1.5	3.6	0.09	0.485
Second year after implementation					
MCAS ELA (distance from proficiency)	-12.5	-14.5			
Deviation from trend	-1.0	-3.4	2.4	0.29 *	0.089
MCAS ELA proficient or above (% of students)	11.1	12.2			
Deviation from trend	-3.1	-3.3	0.2	NA	0.975
SAT-9 Total Reading, fall fourth grade (scaled score)	576.1	573.2			
Deviation from trend	4.8	1.0	3.8	0.10	0.524
Sample size[c]			4,038		

(continued)

Appendix Table C.11 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in third grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fourth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in third grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fourth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students in the entire sample (by grade) during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or above outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all third-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 2,987 students and includes all third-grade students with SAT-9 scores reported in the fall of their fourth-grade year.

Appendix Table C.12

Impacts on Third-Grade Students' Reading Test Scores for Students Not Receiving Special Education Services

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (distance from proficiency)	-5.8	-6.4			
Deviation from trend	-1.8	-1.9	0.2	0.02	0.797
MCAS ELA proficient or above (% of students)	33.1	32.1			
Deviation from trend	-5.2	-3.6	-1.6	NA	0.662
SAT-9 Total Reading, fall fourth grade (scaled score)	612.7	613.7			
Deviation from trend	0.6	3.5	-2.9	-0.08	0.305
Second year after implementation					
MCAS ELA (distance from proficiency)	-5.1	-6.2			
Deviation from trend	-1.1	-1.7	0.7	0.08	0.398
MCAS ELA proficient or above (% of students)	34.3	30.0			
Deviation from trend	-4.4	-5.9	1.5	NA	0.727
SAT-9 Total Reading, fall fourth grade (scaled score)	612.1	611.7			
Deviation from trend	-0.6	1.1	-1.7	-0.04	0.606
Sample size[c]			15,974		

(continued)

Appendix Table C.12 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Reading Comprehension, Grade 3; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in third grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fourth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in third grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fourth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students in the entire sample (by grade) during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or above outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all third-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 13,547 students and includes all third-grade students with SAT-9 scores reported in the fall of their fourth-grade year.

Formative Assessments of Student Thinking in Reading (FAST-R) Evaluation

Appendix Table C.13

Impacts on Fourth-Grade Students' Reading Test Scores for Students Receiving Special Education Services

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (scaled score)	220.6	219.9			
Deviation from trend	-1.3	-0.8	-0.4	-0.03	0.769
MCAS ELA proficient or advanced (%)	3.5	3.4			
Deviation from trend	-6.1	-2.0	-4.1	NA	0.243
SAT-9 Total Reading, fall fifth grade (scaled score)	603.3	603.9			
Deviation from trend	2.8	3.1	-0.2	-0.01	0.958
Second year after implementation					
MCAS ELA (scaled score)	220.8	219.3			
Deviation from trend	-1.6	-2.2	0.6	0.04	0.745
MCAS ELA proficient or advanced (% of students)	4.8	4.6			
Deviation from trend	-6.1	-1.2	-4.9	NA	0.227
SAT-9 Total Reading, fall fifth grade (scaled score)	604.6	596.2			
Deviation from trend	4.2	-6.0	10.2	0.27 *	0.058
Sample size[c]			3,953		

(continued)

Appendix Table C.13 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all fourth-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Reading Total is 3,206 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Appendix Table C.14

Impacts on Fourth-Grade Students' Reading Test Scores for Students Not Receiving Special Education Services

Outcome	FAST-R	Non-FAST-R[a]	Estimated Impact	Estimated Impact Effect Size[b]	P-Value for Estimated Impact
First year after implementation					
MCAS ELA (scaled score)	233.2	232.2			
Deviation from trend	-0.2	-1.2	1.0	0.08	0.442
MCAS ELA proficient or advanced (%)	31.3	27.0			
Deviation from trend	0.9	-5.9	6.8	NA	0.121
SAT-9 Total Reading, fall fifth grade (scaled score)	642.5	640.5			
Deviation from trend	4.3	1.9	2.3	0.06	0.379
Second year after implementation					
MCAS ELA (scaled score)	234.0	231.4			
Deviation from trend	0.0	-2.8	2.8	0.22 *	0.074
MCAS ELA proficient or advanced (% of students)	35.2	29.6			
Deviation from trend	4.4	-5.0	9.4	NA *	0.065
SAT-9 Total Reading, fall fifth grade (scaled score)	637.8	641.8			
Deviation from trend	-1.3	2.3	-3.5	-0.09	0.256
Sample size[c]			14,480		

(continued)

Appendix Table C.14 (continued)

SOURCES: MDRC calculations from individual student records from Boston Public Schools between the 2000-2001 and 2006-2007 academic years.

NOTES: The trend is constructed using five years of data before the FAST-R program was implemented (the 2000-2001 through 2004-2005 academic years). The first year after implementation of FAST-R is the 2005-2006 academic year, and the second year after implementation is the 2006-2007 academic year.

MCAS ELA = Massachusetts Comprehensive Assessment System, English Language Arts, Grade 4; SAT-9 = Stanford Achievement Test, version 9; NA = not applicable.

In the top panel, "First year after implementation," all outcomes are for the group of students in fourth grade during the 2005-2006 school year. These students took the SAT-9 in the fall of 2006 at the beginning of their fifth-grade year. In the bottom panel, "Second year after implementation," all outcomes are for the group of students in fourth grade during the 2006-2007 school year. These students took the SAT-9 in the fall of 2007 at the beginning of their fifth-grade year.

The MCAS and SAT-9 scores are the adjusted means during the years after implementation for the average, or "representative," school. To calculate these means, the trends and intercepts for each school were summed for all FAST-R and non-FAST-R schools, respectively. The trend and intercept were then used to calculate the predicted outcomes during each of the two years after implementation. The deviation from the trend was added to these predicted outcomes to find the average student achievement scores. Estimated impacts are regression-adjusted for students' background characteristics and prior achievement. The deviation from trend for each follow-up year for the non-FAST-R schools is regression-adjusted. The deviation from trend for each follow-up year for the FAST-R schools is the impact added to the non-FAST-R schools' deviation from trend.

A two-tailed t-test was applied to the estimated impact of the FAST-R program. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]FAST-R schools were matched with multiple comparison schools. All comparison schools matched to a particular FAST-R school are weighted so that the total weight of all the matching non-FAST-R schools equals one. Since each comparison school could also be matched to several FAST-R schools, all the weights of a particular non-FAST-R school are summed.

[b]The estimated impact effect size was calculated by dividing the estimated impact by the standard deviation of outcomes for all students during the 2000-2001 through 2004-2005 school years. Since the MCAS proficient or advanced outcome is in percentage points, effect size cannot be calculated.

[c]The sample size includes all fourth-grade students with reported MCAS scores in the FAST-R and non-FAST-R schools during the seven years included in the analysis. The sample size for the SAT-9 Total Reading is 12,351 students and includes all fourth-grade students with SAT-9 scores reported in the fall of their fifth-grade year.

Appendix D

Sample of FAST-R Assessment Student and Teacher Materials

Appendix D features sample FAST-R assessment materials created by the Boston Plan for Excellence (BPE).

The sample FAST-R assessment shown on the following pages was given to third- and fourth-grade students in the Boston Public Schools. The assessment passage is entitled "Apartment Building," by Eileen Spinelli. After reading the story, students answer 11 questions based on the "drawing inferences" and "finding evidence" reading skills.

The assessment materials also include a FAST-R teacher's guide explaining the scoring process BPE used to grade students' assessments. Teachers can see the reading skills and vocabulary highlighted in the text. The guide also provides teachers with ideas for follow-up lessons connected to the exercise.

Teachers can tell whether a question is assessing the "making inferences" or "finding evidence" reading skills in the FAST-R assessment. In addition, all wrong answers to a question are coded as OOP1, indicating answers that are true but irrelevant based on the text; OOP2, indicating "misread" answers based on a misunderstanding of the reading text; or OOB, which represents "out of bounds" answers that are not based at all on the text.

Following the teacher's guide in this appendix is a sample of a FAST-R assessment data report that teachers receive outlining their students' performance on the assessment. Teachers can see how students performed overall as a class, individually, and by subgroup (such as race, ethnicity, gender, limited English proficient, and special education).

Finally, a FAST-R "Student Performance Summary" explains the various types of questions that measure students' ability to make inferences and find evidence in reading passages. Teachers can pinpoint the strengths and weaknesses of their students with this enhanced level of question-item analysis.

FAST-R

Formative Assessments of Student Thinking in Reading

Name | "Apartment Building" • Fiction

Date | Teacher/Class

Apartment Building

by Eileen Spinelli

In this story, a boy named Sam solves a problem in a very interesting way. Read this story to find out what his problem is, and how he solves it. Answer the questions that follow.

1 Sam wanted a pet. A little one. A big one. One that flew or walked. Any kind of pet at all.

2 But Mrs. Cabot, the landlady, said: "NO PETS!" And that was that.

3 Sam's mom and dad bought him fuzzy stuffed bears and a kite in the shape of a kitten and a plastic parrot on a stick. But he wanted a live pet. And that was that.

4 Sam tried to change Mrs. Cabot's mind. He told her he would get a quiet pet. He told her he would keep it clean. He told her he would not let the pet scratch the woodwork or jump through the hedges.

5 But Mrs. Cabot looked Sam square in the eye and said: "NO PETS!"

6 One day Sam heard Mrs. Cabot screaming in the hallway. He rushed over. "What's wrong, Mrs. Cabot?"

7 "A mouse!" she shrieked. "I saw a mouse!"

"I thought you had a rule, Mrs. Cabot. No pets allowed." 8

"It wasn't a pet mouse," the landlady squawked. "It was a plain old terrorize-the-building type of mouse." 9

Sam grinned. "What you need, Mrs. Cabot, is a cat." 10

While it was true that Mrs. Cabot hated cats, there was something she hated even more – mice! And so that very day, Mrs. Cabot went to the animal shelter and found herself a cat. 11

The next time Sam saw Mrs. Cabot, he said, "I see you have a cat." 12

"Yes indeed," she replied. 13

"Does that mean I can get a pet?" Sam asked. 14

"No, it does not!" snapped Mrs. Cabot. "If I let you get a pet, I'd have to let everyone get a pet. Then I'd be running a zoo and not an apartment building!" 15

Another day Sam was coming home from school. Police cars surrounded the apartment building. "What happened, Mrs. Cabot?" he asked. 16

Spotlight On: Eileen Spinelli

Eileen Spinelli has written over 20 books, including picture books and collections of poems. She is the author of the Lizzy Logan series. She has six children and thirteen grandchildren. Her husband, Jerry Spinelli, is also a writer. His work includes Maniac Magee.

17 “I was robbed!” she cried. “They took my radio and my penny bank and my entire collection of salt and pepper shakers!”

18 “That’s too bad,” said Sam, shaking his head. “What you need is a good watchdog.”

19 While it was true that Mrs. Cabot hated dogs, there was something she hated even more – robbers! And so that very day, Mrs. Cabot went to the kennel and bought herself a dog.

20 A week later Sam found Mrs. Cabot dusting the stairs. There were tears in her eyes.

21 “You look sad,” said Sam.

22 “I am sad,” replied Mrs. Cabot. “My very best friend in all the world is moving away.”

23 “I’m sorry to hear that,” said Sam.

24 “She and I would talk for hours,” sniffled Mrs. Cabot.

25 “Can’t you talk to Mr. Cabot?” asked Sam.

26 Mrs. Cabot shook her head. “Mr. Cabot likes to read the newspaper. He likes to watch TV. He likes to build shelves for his wrench collection. But he doesn’t like to talk.”

27 “What you need, Mrs. Cabot, is a parrot.”

28 Mrs. Cabot dabbed at her eyes with the hem of her apron. “A parrot?”

29 Sam nodded. “Parrots love to talk.”

30 While it was true that Mrs. Cabot hated birds, there was something she hated even more – not having anyone to talk with. And so that very day, Mrs. Cabot went to The Exotic Bird Shop and bought herself a parrot that talked all the way home.

31 In spring the grass grew green and thick and tall. Mrs. Cabot tried to mow the lawn, but every time she pushed the mower, she sneezed.

32 “Bless you,” said Sam.

33 “Thank you,” said Mrs. Cabot, between sneezes.

34 “Do you have a cold?” Sam asked.

35 Mrs. Cabot blew her nose. Sneezed. Pushed the mower. Then sneezed again. “No. I have allergies.”

“You shouldn’t be mowing grass, then,” said 36
Sam.

“I know that,” replied Mrs. Cabot, sneezing. 37
“But Mr. Cabot hurt his back. So he can’t mow the grass.”

“Maybe I could,” offered Sam. 38

“That’s kind of you,” said Mrs. Cabot. “But 39
you’re too small.”

Sam smiled. “I know just what you need.” 40

“What’s that?” asked Mrs. Cabot. 41

“A goat!” Sam laughed. “A goat will eat every 42
bit of grass. You’ll never have to mow again.”

While it was true that Mrs. Cabot hated goats, 43
there was something she hated even more – sneezing! And so that very day, she drove to a farm and came back with a goat.

A month later Sam found Mrs. Cabot hammer- 44
ing a For Sale sign into the front lawn.

“Are you selling the building?” asked Sam. 45

“I don’t want to,” sighed Mrs. Cabot. “But I’m 46
so busy taking care of the cat and the dog and the parrot and the goat that I don’t have time for anything else. The laundry room is full of cobwebs. The stairs are full of dust. And as for the hedges – well, see for yourself.”

Sam patted Mrs. Cabot on the shoulder. “What 47
you need is a pet-sitter.”

Mrs. Cabot stopped hammering. “But who on 48
earth would take care of all those pets?”

Sam’s grin was as wide as a wheelbarrow. “I 49
know just the person,” he said. “And he even lives in the building!”

FAST-R

Formative Assessments of Student Thinking in Reading

Name "Apartment Building" • Fiction

Date Teacher/Class

Fill in the circle for the best answer to each question.

Apartment Building

1. The introduction says Sam has a problem to solve. What was that problem?

- (A) He wanted a pet but wasn't allowed to have one.
- (B) He had allergies whenever someone cut the grass.
- (C) Mrs. Cabot was selling the apartment building and his family had to move.
- (D) He was very busy taking care of all Mrs. Cabot's pets.

2. According to the passage, the MOST LIKELY reason that Mrs. Cabot did not allow pets was that

- (A) she didn't like animals.
- (B) she was allergic to animals.
- (C) she wanted to run a zoo.
- (D) Mr. Cabot was afraid of animals.

3. According to paragraph 4, how did Sam try to change Mrs. Cabot's mind about letting him have a pet?

- (A) He said he would buy her a kite or a stuffed bear.
- (B) He offered to take out the garbage for a month.
- (C) He told her that pets could help solve her problems.
- (D) He promised his pet would be quiet and clean.

4. Why did Sam grin when he suggested Mrs. Cabot buy a cat?

- (A) He thought it was funny that Mrs. Cabot hated mice.
- (B) He hoped Mrs. Cabot would get a cat even though she had a rule against pets.
- (C) He liked mice.
- (D) He knew that Mrs. Cabot liked cats.

5. According to the passage, why did Sam think that Mrs. Cabot should get a goat?

- (A) to chase away robbers
- (B) so she could start a petting zoo next to the apartment building
- (C) so she wouldn't have to mow the lawn
- (D) because she was allergic to cats

Continue on the next page

Name School

Date Teacher/Class

6. What did all of Mrs. Cabot's pets have in common?

- Ⓐ She bought all of them from pet stores.
- Ⓑ She chose all of them to help her in some way.
- Ⓒ They were all pets Sam did not want.
- Ⓓ They were all good at doing tricks.

7. What was the MAIN reason Sam kept suggesting animals for Mrs. Cabot to get?

- Ⓐ He wanted Mrs. Cabot to have a pet instead of him.
- Ⓑ He liked helping Mrs. Cabot.
- Ⓒ He hoped Mrs. Cabot would let him get a pet too.
- Ⓓ He wished that Mrs. Cabot would move away.

8. Why did Mrs. Cabot want to sell the apartment building?

- Ⓐ She did not want to take care of all her pets anymore.
- Ⓑ She was afraid of robbers.
- Ⓒ She thought she could make money from selling the building.
- Ⓓ She could not take care of both the animals and the apartment building.

9. Who did Sam think would be a good pet-sitter for Mrs. Cabot?

- Ⓐ himself
- Ⓑ his parents
- Ⓒ Mr. Cabot
- Ⓓ the man who walks his dog past the building

10. What is the MAIN IDEA of this story?

- Ⓐ Sometimes you can get what you want by helping others.
- Ⓑ Sometimes you can find pets at an animal shelter.
- Ⓒ Sometimes there's not enough work to do around an apartment building.
- Ⓓ Sometimes you can meet new people near your home.

Open Response Question:

Explain Sam's problem and how he solves it. Use important information from the selection in your answer.

Teachers: Please duplicate and use this answer sheet only for students for whom you did not receive a pre-printed answer sheet!

FAST-R Answer Sheet

Name ______________________ School ______________________

Date ______________________ Grade __________ Class __________

Passage Title ______________________ Teacher Name ______________________

Completely fill the circle for the correct answer.

1. Ⓐ Ⓑ Ⓒ Ⓓ
2. Ⓐ Ⓑ Ⓒ Ⓓ
3. Ⓐ Ⓑ Ⓒ Ⓓ
4. Ⓐ Ⓑ Ⓒ Ⓓ
5. Ⓐ Ⓑ Ⓒ Ⓓ
6. Ⓐ Ⓑ Ⓒ Ⓓ
7. Ⓐ Ⓑ Ⓒ Ⓓ
8. Ⓐ Ⓑ Ⓒ Ⓓ
9. Ⓐ Ⓑ Ⓒ Ⓓ
10. Ⓐ Ⓑ Ⓒ Ⓓ

Write your answer to the open response prompt in the lined space below if your teacher directs you to do so.

OFFICE USE ONLY				
RESEARCH:	Ⓨ	Ⓝ		
OPEN RESPONSE:	①	②	③	④

Teacher Guide for FAST-R Passage: Apartment Building Realistic Fiction

FAST-R: Formative Assessments of Student Thinking in Reading

At a Glance

Approximate Grade Range: 2 3 4 5 6 7 8 9 10 11

Genre: Realistic Fiction

Topic: By convincing his landlady that various animals can help solve her problems, Sam winds up getting to "pet-sit" despite the landlady's ban on pets.

Author: Eileen Spinelli

Source: *Highlights for Children*

Difficulty Index: Considerate . . . . Challenging

Structure:
Purpose:
Richness:
Relationships:
Vocabulary:
Style:

Lexile Measure: 610L

In this story, a boy named Sam solves a problem in a very interesting way. Read this story to find out what his problem is, and how he solves it. Answer the questions that follow.

Apartment Building

by Eileen Spinelli

1 Sam wanted a pet. A little one. A big one. One that flew or walked. Any kind of pet at all.

2 But Mrs. Cabot, the landlady, said: "NO PETS!" And that was that.

3 Sam's mom and dad bought him fuzzy stuffed bears and a kite in the shape of a kitten and a plastic parrot on a stick. But he wanted a live pet. And that was that.

4 Sam tried to change Mrs. Cabot's mind. He told her he would get a quiet pet. He told her he would keep it clean. He told her he would not let the pet scratch the woodwork or jump through the hedges.

5 But Mrs. Cabot looked Sam square in the eye and said: "NO PETS!"

6 One day Sam heard Mrs. Cabot screaming in the hallway. He rushed over. "What's wrong, Mrs. Cabot?"

7 "A mouse!" she shrieked. "I saw a mouse!"

8 "I thought you had a rule, Mrs. Cabot. No pets allowed."

9 "It wasn't a pet mouse," the landlady squawked. "It was a plain old terrorize-the-building type of mouse."

10 Sam grinned. "What you need, Mrs. Cabot, is a cat."

11 While it was true that Mrs. Cabot hated cats, there was something she hated even more – mice! And so that very day, Mrs. Cabot went to the animal shelter and found herself a cat.

12 The next time Sam saw Mrs. Cabot, he said, "I see you have a cat."

13 "Yes indeed," she replied.

14 "Does that mean I can get a pet?" Sam asked.

15 "No, it does not!" snapped Mrs. Cabot. "If I let you get a pet, I'd have to let everyone get a pet. Then I'd be running a zoo and not an apartment building!"

16 Another day Sam was coming home from school. Police cars surrounded the apartment building. "What happened, Mrs. Cabot?" he asked.

17 "I was robbed!" she cried. "They took my radio and my penny bank and my entire collection of salt and pepper shakers!"

Structure: Sam's "problem," which is alluded to in the introductory blurb, is that he wants a pet but his landlady prohibits them. After Sam tries to persuade her to relent (¶4), Mrs. Cabot encounters a series of problems that Sam suggests could be solved by an animal.

Repetitive text elements add predictability and make the story particularly considerate for struggling readers. The story ends with a final twist that solves Sam's problem by solving Mrs. Cabot's.

See especially: Questions 1, 4, 6, 10

Relationships: Sam identifies Mrs. Cabot's needs and suggests animals as a solution, thereby hoping to meet his own "need" as well.

See especially: Questions 4, 6, 7, 10

Continued on next page

18 “That’s too bad,” said Sam, shaking his head. “What you need is a good watchdog.”

19 While it was true that Mrs. Cabot hated dogs, there was something she hated even more – robbers! And so that very day, Mrs. Cabot went to the kennel and bought herself a dog.

Structure: Repetition leads to expected outcome. The repetition of “While it was true that Mrs. Cabot hated [each animal]…” adds a predicable element that lends itself to reading aloud and predicting how Sam will suggest solving each new problem.

See especially: Question 5

20 A week later Sam found Mrs. Cabot dusting the stairs. There were tears in her eyes.

21 “You look sad,” said Sam.

22 “I am sad,” replied Mrs. Cabot. “My very best friend in all the world is moving away.”

23 “I’m sorry to hear that,” said Sam.

24 “She and I would talk for hours,” sniffled Mrs. Cabot.

25 “Can’t you talk to Mr. Cabot?” asked Sam.

26 Mrs. Cabot shook her head. “Mr. Cabot likes to read the newspaper. He likes to watch TV. He likes to build shelves for his <u>wrench</u> collection. But he doesn’t like to talk.”

Vocabulary: Dabbed, mowing, wrench, exotic

27 “What you need, Mrs. Cabot, is a parrot.”

28 Mrs. Cabot <u>dabbed</u> at her eyes with the hem of her apron. “A parrot?”

29 Sam nodded. “Parrots love to talk.”

30 While it was true that Mrs. Cabot hated birds, there was something she hated even more – not having anyone to talk with. And so that very day, Mrs. Cabot went to The <u>Exotic</u> Bird Shop and bought herself a parrot that talked all the way home.

31 In spring the grass grew green and thick and tall. Mrs. Cabot tried to <u>mow</u> the lawn, but every time she pushed the mower, she sneezed.

32 “Bless you,” said Sam.

33 “Thank you,” said Mrs. Cabot, between sneezes.

34 “Do you have a cold?” Sam asked.

35 Mrs. Cabot blew her nose. Sneezed. Pushed the mower. Then sneezed again. “No. I have allergies.”

36 “You shouldn’t be mowing grass, then,” said Sam.

37 “I know that,” replied Mrs. Cabot, sneezing. “But Mr. Cabot hurt his back. So he can’t mow the grass.”

38 “Maybe I could,” offered Sam.

39 “That’s kind of you,” said Mrs. Cabot. “But you’re too small.”

40 Sam smiled. “I know just what you need.”

41 “What’s that?” asked Mrs. Cabot.

42 “A goat!” Sam laughed. “A goat will eat every bit of grass. You’ll never have to mow again.”

43 While it was true that Mrs. Cabot hated goats, there was something she hated even more – sneezing! And so that very day, she drove to a farm and came back with a goat.

Spotlight On: Eileen Spinelli

Eileen Spinelli has written over 20 books, including picture books and collections of poems. She is the author of the Lizzy Logan series. She has six children and thirteen grandchildren. Her husband, Jerry Spinelli, is also a writer. His work includes <u>Maniac Magee.</u>

Open Response Prompt:

Explain Sam’s problem and how he solves it. Use important information from the selection in your answer.

44 A month later Sam found Mrs. Cabot hammering a For Sale sign into the front lawn.

45 “Are you selling the building?” asked Sam.

46 “I don’t want to,” sighed Mrs. Cabot. “But I’m so busy taking care of the cat and the dog and the parrot and the goat that I don’t have time for anything else. The laundry room is full of cobwebs. The stairs are full of dust. And as for the hedges – well, see for yourself.”

47 Sam patted Mrs. Cabot on the shoulder. “What you need is a pet-sitter.”

Structure: The final twist solves Sam’s “problem” while once again solving one of Mrs. Cabot’s problems. The text leaves implicit just who Sam intends to suggest as a pet-sitter, but his overall approach to the problem-solution pattern, as well as his “grin as wide as a wheelbarrow” provide strong clues.

See especially: Question 9

48 Mrs. Cabot stopped hammering. “But who on earth would take care of all those pets?”

49 Sam’s grin was as wide as a wheelbarrow. “I know just the person,” he said. “And he even lives in the building!”

Teacher Guide for FAST-R Passage: Apartment Building

FAST-R: Formative Assessments of Student Thinking in Reading

Realistic Fiction

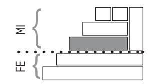
The annotated answer key below highlights common reasons students might choose each answer, and the sidebar gives more insight into the question types, to help you understand patterns of student responses. Always make time to follow up with students in conferences or small groups to probe their thinking, teach in response to patterns, and help them apply effective reading and thinking strategies to their everyday reading.

Note: You may find it helpful to refer to the **"Types and Levels of Thinking Assessed on FAST-R"** sheet from your teacher resource folder as you examine your students' responses. The icon in the right-hand column, below, corresponds to that sheet's more detailed explanations of the kinds of thinking each type of question asks of readers.

1. The introduction says Sam has a problem to solve. What was that problem?

✓ A. He wanted a pet but wasn't allowed to have one. (¶2)
B. He had allergies whenever someone cut the grass. (OOP2, ¶35)
C. Mrs. Cabot was selling the apartment building and his family had to move. (OOP2, ¶44-46)
D. He was very busy taking care of all Mrs. Cabot's pets. (OOP2, ¶48)

MI: *Determine implicit meaning from ideas in context*

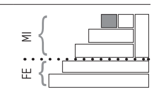


---

2. According to the passage, the MOST LIKELY reason that Mrs. Cabot did not allow pets was that

✓ **A. she didn't like animals.** — *can be inferred from the fact that each time Sam suggests an animal to solve her problem, it says Mrs. Cabot "hated" the animal*
B. she didn't want too many animals in the building. — (OOP1) *at the start it says he didn't want ANY animals in her building*
C. she wanted to run a zoo. — (OOP2) *Mrs. Cabot snapped, "Then I'd be running a zoo and not an apartment building," implying disfavor at the idea*
D. Mr. Cabot was afraid of animals. — (OOB) *no evidence suggests this*

MI: *Determine implicit meaning by understanding the organization of information in the text*

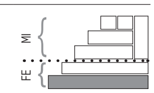


---

3. According to paragraph 4, how did Sam try to change Mrs. Cabot's mind about letting him have a pet?

A. He said he would buy her a kite or a stuffed bear. — (OOP2, ¶3) *Sam's parents bought him a kite and a fuzzy stuffed bear*
B. He offered to take out the garbage for a month. — (OOB) *a persuasive tactic students may have tried on adults in their lives, but not in the text*
C. He told her that pets could help solve her problems. — (OOP1, *true, but that did not persuade her to let him get a pet of his own* [*see* ¶*15-16 "Does that mean I can get a pet?" "No, it does not!"*])
✓ **D. He promised his pet would be quiet and clean.** — (¶4)

FE: *Identify evidence explicitly stated in the text*

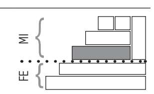


---

4. Why did Sam grin when he suggested Mrs. Cabot buy a cat?

A. He thought it was funny that Mrs. Cabot hated mice. — (OOB) *a grin may signal humor, but Sam "rushed over" out of concern when she screamed*
✓ **B. He hoped Mrs. Cabot would get a cat in spite of her rule against pets.** — *can be inferred from his questions to Mrs. Cabot about getting a pet*
C. He liked mice.— (OOP1) *Probably true, since he wanted "any kind of pet at all," but not the reason he grinned*
D. He knew that Mrs. Cabot liked cats. — (OOP2) *Sam himself liked cats, but Mrs. Cabot "hated cats"*

MI: *Determine implicit meaning from words in context*

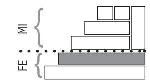


---

5. According to the passage, why did Sam think Mrs. Cabot should get a goat?

A. to chase away robbers — (OOP2) *he suggested a* dog *to chase robbers*
B. so she could start a petting zoo next to the apartment building — (OOB) *goats are common in petting zoos, but no evidence supports this choice*
✓ **C. so she wouldn't have to mow the lawn** — *(* ¶ 42)
D. to cure her allergies — (OOP1) *the goat won't "cure" her grass allergies, but would help alleviate them by reducing her need to mow the lawn*

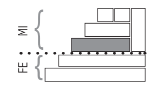
FE: *Recognize the explicit meaning from varied wording in the text*



6. What did all of Mrs. Cabot's pets have in common?

A. She bought all of them from pet stores. — (OOP1) *only the bird came from a pet store; the others came from an animal shelter, kennel, or farm*
✓ **B. She chose all of them to help her in some way.** — *can be inferred from evidence throughout the passage that each pet was bought for a specific purpose*
C. They were all pets Sam did not want. — (OOP2) *clearly not true, since he eagerly offered to care for them at the end*
D. They were all good at doing tricks. — (OOB) *though they were intended to help her, it was not by doing tricks*

MI: *Determine implicit meaning from words in context*



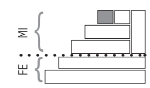
7. What was the MAIN reason Sam kept suggesting animals for Mrs. Cabot to get?

A. He wanted Mrs. Cabot to have a pet instead of him. — (OOP2) *Sam hoped he **and** Mrs. Cabot could have pets*

B. He liked helping Mrs. Cabot. — (OOP1) *Sam offered to mow the lawn and pet-sit, but helpfulness is not the **main** reason he wants her to get pets*
✓ **C. He hoped Mrs. Cabot would let him get a pet too.** - (¶ 14)

D. He wished that Mrs. Cabot would move away. — (OOB)

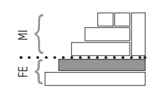
MI: *Determine implicit meaning by understanding the organization of information in the text*



8. Why did Mrs. Cabot want to sell the apartment building?

A. She did not want to take care of all her pets anymore. — (OOP2) *she's too busy to care for animals and the building, but is keeping the animals*
B. She was afraid of robbers. — (OOP1) *she was afraid of robbers, but that's why she got a dog, not why she wants to sell the building*
C. She thought she could make money from selling the building. — (OOB) *logical (especially in Boston's real estate market!) but not implied*
✓ **D. She could not take care of both the animals and the apartment building.** — (¶ 46)

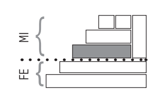
FE: *Recognize the explicit meaning from varied wording in the text*



9. Who did Sam think would be a good pet-sitter for Mrs. Cabot?

✓ A. himself
B. his parents (OOP2)
C. Mr. Cabot (OOP2)
D. the man who walks his dog past the building (OOB)

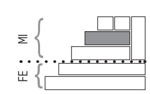
MI: *Determine implicit meaning from ideas in context*



10. What is the MAIN IDEA of this story?

✓ **A. Sometimes you can get what you want by helping others.** — *by helping Mrs. Cabot solve her various problems, Sam gets what he wants*
B. Sometimes you can find pets at an animal shelter. — (OOP1) *the cat was found at an animal shelter, but this is not the main idea of the story*
C. Sometimes there's not enough work to do around an apartment building. — (OOP2) *there was too much work to be done*
D. Sometimes you can meet new people near your home. — (OOB) *this may sound reasonable to students, but is not related to the story*

MI: *Determine a singular meaning from the total text*

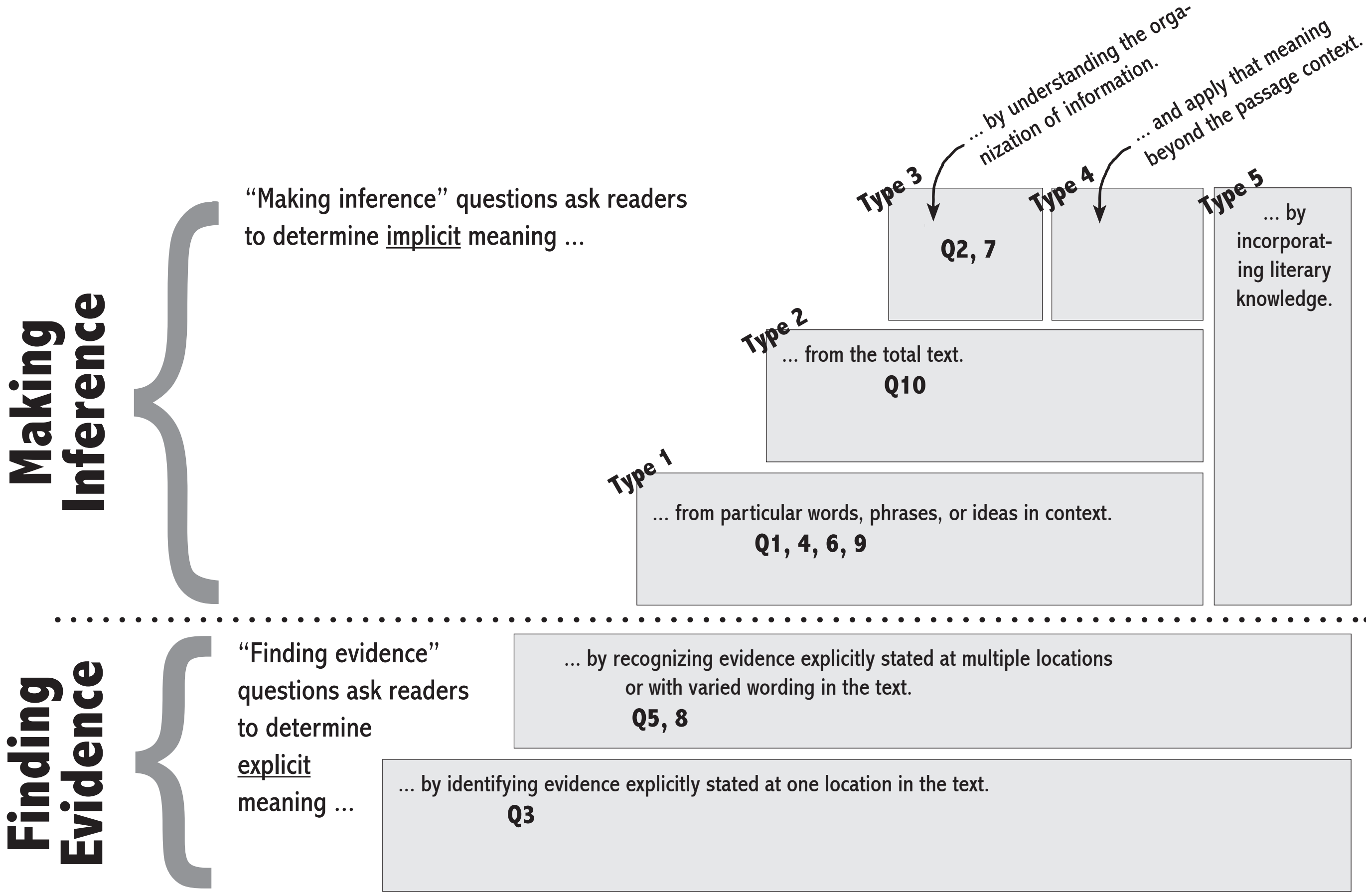


Teacher Guide for FAST-R Passage: Apartment Building

FAST-R: Formative Assessments of Student Thinking in Reading

Realistic Fiction

What levels and types of thinking are assessed by questions on this FAST-R question set?



FAST-R Student Performance Summary

These graphs show how many students selected each answer choice, while the annotated answer key highlights types of wrong answers, where answers were found in the text, and what thinking may have led to that answer choice.

The data can help you hypothesize about your students' reading and thinking, but talking with your students will confirm or contradict your hypotheses and help you plan mini-lessons, conferring, and guided reading instruction in response to your students' strengths and needs.

Key

- Correct answer
- OOP1 "near miss" answer; true based on text, but irrelevant
- OOP2 "mis-read" answer; based on misunderstanding of the text
- OOB answer; not based on text; plausible from prior knowledge
- No answer

This analysis represents the performance of:

School **Sample ES** Grade: **04** Class/Sec: **08** Question Set: **G4F1**

---

1. How did Sam try to change Mrs. Cabot's mind about letting him have a pet?

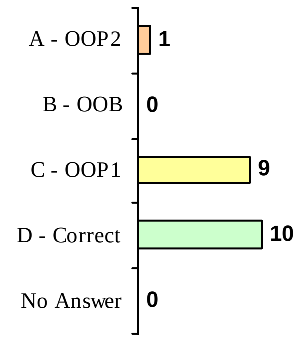
A. He said he would buy her a kite or a stuff ed bear. — (OOP2, ¶3) Sam's parents bought him a kite and a fuzzy stuffed bear

B. He offered to take out the garbage for a month. — (OOB) a persuasive tactic students may have tried on adults in their lives, but not in the text

C. He told her that pets could help solve her problems. — (OOP1) true, but did not persuade her (see ¶15-16 "Does that mean I can get a pet?" "No, it does not!"

✓ D. He promised his pet would be quiet and clean. — (¶4)

1. Finding Evidence



2. Why did Mrs. Cabot get a parrot?

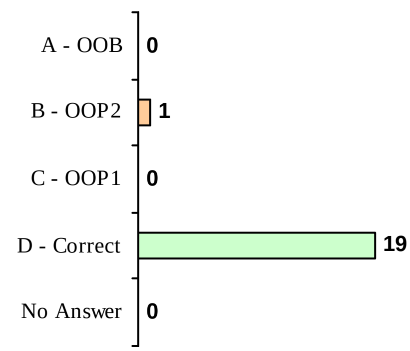
A. She liked birds. — (OOB) sounds reasonable, not supported by the text

B. She got a plastic parrot for decoration. — (OOP2) in ¶3, Sam's parents bought him a plastic parrot; not the reason she got a real parrot

C. So that Sam could help take care of it. — (OOP1) at the end, Sam offers to help care for the pets, but that's not why Mrs. Cabot gets a parrot

✓ D. She didn't have anyone to talk with. — found about two-thirds of the way through the passage

2. Finding Evidence



3. All of Mrs. Cabot's pets were the SAME in what way?

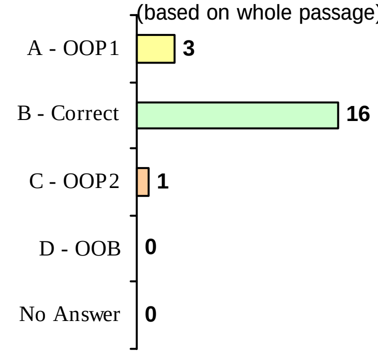
A. She bought all of them from pet stores. — (OOP1) only the bird came from a pet store; the others came from a shelter, kennel, or farm

✓ B. She chose all of them to help her in some way. — can be inferred from evidence throughout the passage that each pet was bought for a specific purpose

C. They were all pets Sam did not want. — (OOP2) clearly not true; he eagerly offered to care for them at the end

D. They were all good at doing tricks. — (OOB) though they were intended to help her, it was not by doing tricks

3. Making Inferences
(based on whole passage)



---

4. What is the MAIN IDEA of this story?

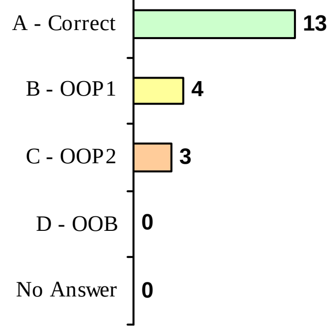
✓ A. Sometimes you can get what you want by helping others. — Sam eventually gets to take care of the pets that he wants by helping Mrs. Cabot solve her various problems

B. Sometimes you can find pets at an animal shelter. — (OOP1) the cat was, but this is not the main idea of the story

C. Sometimes there's not enough work to do around an apartment building. — (OOP2) there was too much work to be done, and it's not the main idea

D. Sometimes you can meet new people near your home. — (OOB) this may sound reasonable, but is not related to the main idea of the story

4. Making Inferences
(based on whole passage)



5. According to the passage, why did Sam think Mrs. Cabot should get a goat?

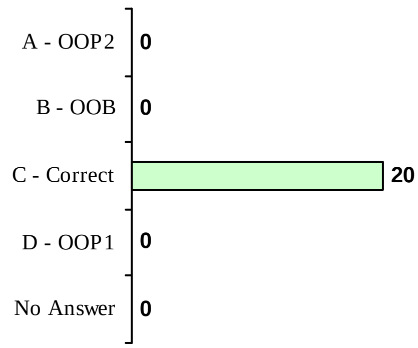
A. to chase away robbers — (OOP2) he suggested a dog to chase robbers

B. so she could start a petting zoo next to the apartment building — (OOB) goats are common in petting zoos, but this has little basis in the text

✓ C. so she wouldn't have to mow the lawn — found seven paragraphs from the end of the passage

D. to cure her allergies — (OOP1) the goat won't "cure" her grass allergies, but would help alleviate them by reducing her need to mow the lawn

5. Finding Evidence



6. What was the MAIN reason Sam kept suggesting animals for Mrs. Cabot to get?

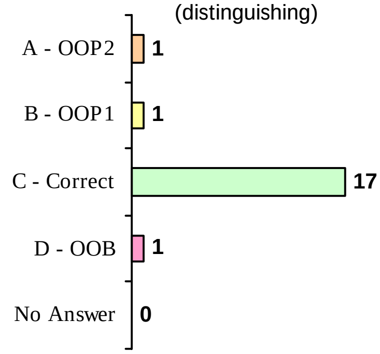
A. He wanted Mrs. Cabot to have a pet instead of him. — (OOP2) Sam hoped he AND Mrs. Cabot could have pets.

B. He liked helping Mrs. Cabot. — (OOP1) Sam offered to mow the lawn and pet-sit, but helpfulness is not the main reason he wants her to get pets

✓ C. He hoped Mrs. Cabot would let him get a pet too.

D. He wished that Mrs. Cabot would move away. — (OOB)

6. Making Inferences
(distinguishing)



7. Who did Sam think would be a good pet-sitter for Mrs. Cabot?

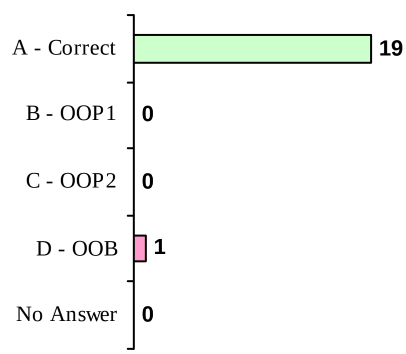
✓ A. himself — can be inferred from Sam's final statement in the passage

B. his parents — (OOP1) Sam's parents "even live in the building" and seemed to support his desire for a pet, but there's no evidence the want to care for pets

C. Mr. Cabot — (OOP2) Mr. Cabot "even lives in the building" but the paragraph describing what he likes to do does not include caring for animals

D. the man who walks his dog past the building — (OOB)

7. Making Inferences



The passage and question 4 come from the Spring 2001 Grade 3 MCAS test; questions 1, 3, and 7 are adapted from the same test.

8. Why did Sam grin when he suggested Mrs. Cabot buy a cat?

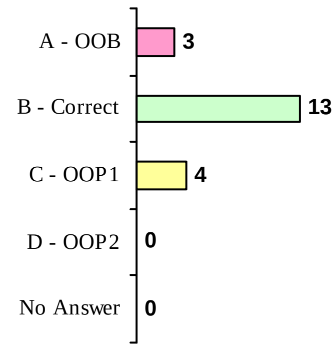
A. He thought it was funny that Mrs. Cabot hated mice. — (OOB) a grin may signal humor, but Sam "rushed over" out of concern when she screamed

✓ B. He hoped Mrs. Cabot would get a cat in spite of her rule against pets. — can be inferred from his questions to Mrs. Cabot about getting a pet

C. He liked mice.— (OOP1) Probably true, since he wanted "any kind of pet at all," but not the reason he grinned

D. He knew that Mrs. Cabot liked cats. — (OOP2) Sam himself liked cats, but Mrs. Cabot "hated cats"

8. Making Inferences



9. The MOST LIKELY reason that Mrs. Cabot did not allow pets was that

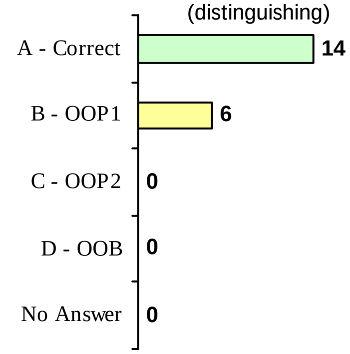
✓ A. she didn't like animals. — can be inferred from the fact that each time Sam suggests an animal to solve her problem, it says Mrs. Cabot "hated" the animal

B. she didn't want too many animals in the building. — (OOP1) at the start sit says he didn't want ANY animals in her building

C. she wanted to run a zoo. — (OOP2) Mrs. Cabot snapped, "Then I'd be running a zoo and not an apartment building," implying disfavor at the idea

D. Mr. Cabot was afraid of animals. — (OOB) no evidence suggests this

9. Making Inferences
(distinguishing)



10. Why did Mrs. Cabot want to sell the apartment building?

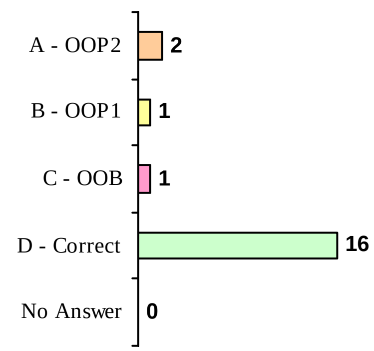
A. She did not want to take care of all her pets anymore. — (OOP2) she's too busy to care for animals and the building, but seems to be keeping the animals

B. She was afraid of robbers. — (OOP1) she was afraid of robbers, but that's why she got a dog, not why she wants to sell the building

C. She thought she could make money from selling the building. — (OOB) logical (especially in Boston's real estate market!) but not implied in the passage

✓ D. She could not take care of both the animals and the apartment building.— found four paragraphs from the end of the passage

10. Finding Evidence



The passage and question 4 come from the Spring 2001 Grade 3 MCAS test; questions 1, 3, and 7 are adapted from the same test.

Question Set: G4F1

Student-Level Data Summary

Key to colors:

green	Correct	
yellow	OOP1	"near miss"
orange	OOP2	"mis-read"
pink	OOB	"out of bounds"

School: Sample ES

Total Students: 20

Skill Summaries:

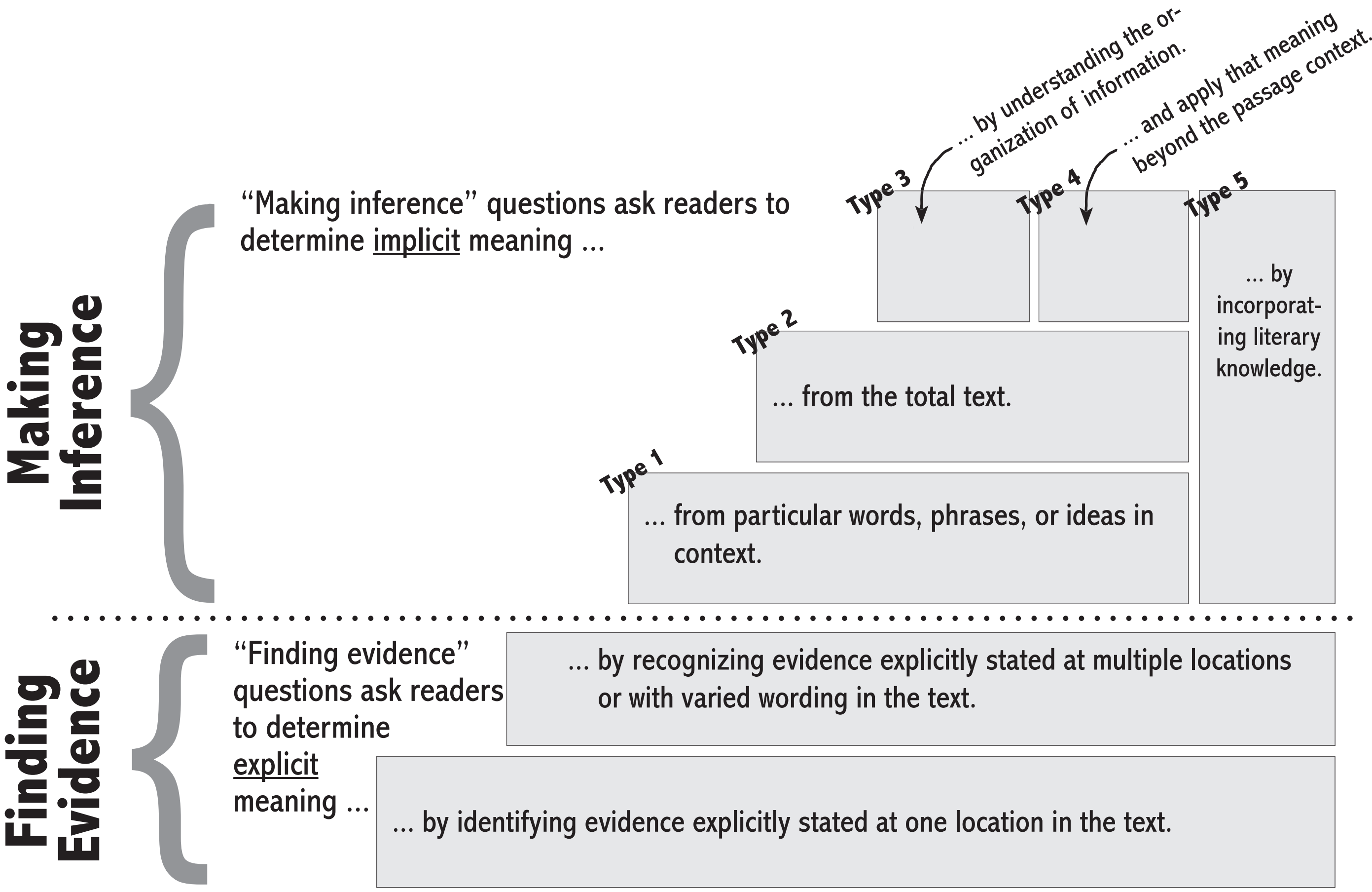
Stud ID #	First Name	Last Name	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	FE # ✓	MI # ✓	FE % ✓	MI % ✓	Total ✓	Gr.	Race	Gender	Class	ELL	SPED	OR	Res.	Adhoc
259518	Student	U	D	D	B	A	C	C	A	B	B	D	4	5	100 %	83 %	**9**	04	-	-	08	-	-	X	N	0
222744	Student	A	D	D	A	A	C	C	A	B	A	D	4	5	100 %	83 %	**9**	04	Asian	Male	08	-	R2	X	Y	0
997497	Student	N	D	D	B	A	C	C	A	A	A	D	4	5	100 %	83 %	**9**	04	Black	Female	08	-	-	X	N	0
220728	Student	J	C	D	B	A	C	C	A	B	A	D	3	6	75 %	100 %	**9**	04	White	Male	08	-	-	X	N	0
206765	Student	C	C	D	B	A	C	C	A	B	A	D	3	6	75 %	100 %	**9**	04	Hispanic	Female	08	-	-	X	Y	0
224929	Student	S	C	D	B	A	C	C	A	B	A	D	3	6	75 %	100 %	**9**	04	White	Female	08	-	-	X	Y	0
223633	Student	R	C	D	B	A	C	C	A	B	A	D	3	6	75 %	100 %	**9**	04	Black	Female	08	-	-	X	Y	0
209985	Student	G	C	D	B	A	C	C	A	B	A	D	3	6	75 %	100 %	**9**	04	Black	Male	08	-	R2	X	N	0
222746	Student	O	C	D	B	A	C	C	A	B	A	D	3	6	75 %	100 %	**9**	04	Black	Female	08	-	-	X	Y	0
210002	Student	H	D	D	B	B	C	C	A	A	A	D	4	4	100 %	66 %	**8**	04	-	-	08	-	-	X	N	0
220416	Student	I	D	D	B	B	C	C	A	B	B	D	4	4	100 %	66 %	**8**	04	Asian	Female	08	-	-	X	Y	0
221045	Student	K	D	D	B	B	C	C	A	B	B	D	4	4	100 %	66 %	**8**	04	Black	Female	08	-	-	X	N	0
222789	Student	P	D	D	A	C	C	C	A	B	A	D	4	4	100 %	66 %	**8**	04	Black	Male	08	-	-	X	N	0
225531	Student	T	C	D	B	A	C	C	A	B	B	D	3	5	75 %	83 %	**8**	04	White	Female	08	-	-	X	Y	0
202186	Student	L	C	D	B	A	C	A	A	C	A	D	3	4	75 %	66 %	**7**	04	Hispanic	Female	08	-	-	X	N	0
208357	Student	E	D	D	B	C	C	C	A	C	A	A	3	4	75 %	66 %	**7**	04	Black	Female	08	-	-	X	N	0
222451	Student	M	D	D	A	A	C	C	A	A	A	B	3	4	75 %	66 %	**7**	04	Asian	Male	08	-	-	X	N	0
208363	Student	F	C	D	B	C	C	C	A	B	B	A	2	4	50 %	66 %	**6**	04	Black	Female	08	-	-	X	Y	0
223477	Student	Q	A	D	B	B	C	B	D	C	A	D	3	2	75 %	33 %	**5**	04	Asian	Male	08	-	-	X	N	0
206594	Student	B	D	B	C	A	C	D	A	C	B	C	2	2	50 %	33 %	**4**	04	White	Male	08	-	-	X	N	0

Types and Levels of Thinking Assessed on FAST-R

Formative Assessments of Student Thinking in Reading

FAST-R assesses two primary ways that readers construct understanding as they read: **finding evidence** to determine explicit meaning in the text, and **making inferences** to determine implicit meaning in the text.

The graphic below represents how these skills build on one another as readers answer questions of increasing sophistication. FAST-R Teacher Guides annotate each question with a miniature version of this graphic to highlight the level of thinking that question assesses.



"Finding Evidence" Questions

Readers find and use evidence to determine the author's explicit meaning from details that are stated in the text. There are two main types of "finding evidence" questions on FAST-R.

Identify evidence explicitly stated at one location in the text

This type of referential question asks readers to identify and locate an answer explicitly stated somewhere in the text itself (including the introductory blurb, author spotlight, and footnotes). To answer this type of question, students must look back at the passage and "identify" the answer as it is stated in the text.

Recognize evidence explicitly stated at multiple locations or with varied wording in the text

Although answers for the second type of "finding evidence" question are also explicitly stated in the text, readers may need to gather the relevant details from several places in the text *or* recognize words and phrases that express the same meaning in slightly different ways. Students are not asked to make any new meanings or draw any new conclusions (as they do to answer inferential questions).

"Making Inference" Questions

Readers make inferences when they determine implicit meaning that is implied but not directly stated. FAST-R "making inference" questions ask readers to fill gaps based on knowledge gleaned from a close reading of the text. This close reading can help students avoid basing inferences on incorrect or irrelevant prior knowledge.

Type 1: Determine implicit meaning from words, phrases, or ideas in context

In this type of inferential question, the word, phrase, or idea to be interpreted is identified in the question stem itself. However, the question does *not* tell students what other clues to use. In other words, students may be directed to a starting place, but they may need to employ other close reading strategies and base their interpretation on additional textual clues.

For example: Vocabulary questions, cause-and-effect questions, and some basic "author's craft" questions.

- "Based on the clues in paragraph 6, what is the best definition of the word *troupe*?" (*word* in context)
- "In paragraph 13, the author describes a rainbow as 'an enormous arch of colored stripes,' suggesting ..." (*phrase* in context)
- "In paragraph 9, why did the Admiral 'make observations from different points'?" (*idea* in context)

To answer these questions correctly, students must draw conclusions and derive implied meanings from several places in the text.

Type 2: Determine a single implicit meaning from the total text

This type of inferential question asks students to draw a single, overarching new meaning from the *entire* passage (or from entire paragraphs or sections). Inferring in this way involves multiple thought processes: Students must comprehend most or all of the details in the text, then synthesize the information in order to understand the larger meaning.

For example: "Main idea" and "main purpose" questions asking about some chunk of the text.

- "In paragraph six, the author's main point is ..."
- "Which sentence best expresses the main idea of this excerpt?"

Type 3: Determine implicit meaning by understanding the organization of information in a text

This type of inferential question requires readers not only to derive new meanings, but *also* to analyze, order, or prioritize those ideas in light of the way the text is organized. This process involves "analyzing" and "evaluating" (as described in Bloom's Taxonomy), and is used as readers "determine importance" and distinguish supporting details from larger ideas.

For example: These questions often use words like "best" or "most important."

- "What would be the best heading for the information in paragraph 9?"
- "Which detail best supports the idea that the narrator has adjusted to life in America?"

Type 4: Determine implicit meaning and apply it beyond the passage context

This type of inferential question requires readers to determine a new meaning from the text *and* apply that meaning in a situation that extends beyond the text itself.

For example:

- "Based on the details in the passage, if Martin were choosing another book to read, he would most likely choose ..."
- "Which topic will the author most likely write about in the chapter immediately following this prologue?"

This thinking process often involves identifying relevant details in the text, accurately interpreting their implied meaning, and then matching or classifying that idea to make sense of a new situation.

Type 5: Incorporate literary knowledge to determine implicit meaning

These inferential questions require students to incorporate their knowledge of literary terms or concepts in order to understand the text. Depending on what kind of interpretation the question asks of readers, the question may also exhibit characteristics of Type 1-4 "making inference" questions.

For example:

- "Why did the author end paragraph 5 with '– and disaster struck'?"
- "Which of the following details helps create a tone of anticipation in the passage?"

References

Black, Paul, and Dylan Wiliam. 1998. "Inside the Black Box: Raising Standards Through Classroom Assessment." *Phi Delta Kappan* 80, 2: 139-148.

Bloom, Howard S. 2003. "Using 'Short' Interrupted Time-Series Analysis to Measure the Impacts of Whole School Reforms: With Applications to a Study of Accelerated Schools." *Evaluation Review* 27, 1: 3-49.

Boston Public Schools. 2004. "Boston Public Schools Whole-School Improvement — The Six Essentials." Web site: http://boston.k12.ma.us/bps/offices.pdf.

Boston Public Schools. 2008a. "A Comprehensive Reform Plan for the Boston Public Schools, Focus on Children II." Web site: http://boston.k12.ma.us/teach/foc.asp.

Boston Public Schools. 2008b. "Professional Development."
Web site: http://boston.k12.ma.us/teach/profdev.asp.

Brophy, Jere. 1986. "Teacher Influences on Student Achievement." *American Psychologist* 41, 10: 1069-1077.

Chrismer, Sara Schwartz. 2005. Formative Assessments of Student Thinking in Reading [FAST-R], Year II Evaluation. Cambridge, MA: Education Matters, Inc.

Harcourt Brace & Company. 1997. *Stanford Achievement Test Series, Ninth Edition, Fall Norms Book.* San Antonio: Harcourt Brace Educational Measurement.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2007. *Empirical Benchmarks for Interpreting Effect Sizes in Research.* New York: MDRC.

Massachusetts Department of Education. 2007. *Massachusetts Comprehensive Assessment System: 2007 MCAS Technical Report.* Malden, MA: Massachusetts Department of Education.

Meisels, Samuel J., and Ruth Piker. 2000. *An Analysis of Early Literacy Assessments Used for Instruction.* Tech. Rep. No 3-002. Ann Arbor: University of Michigan, Center for the Improvement of Early Reading Achievement.

Murnane, Richard J., and Nancy Sharkey. 2004. "Learning from Student Assessment Results: Lessons for New York State." Presented at the Symposium on Education Finance and Organizational Structure in New York State, Albany, March 5, 2004.

National Institute of Child Health and Human Development. 2000. Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction.
Web site: www.nichd.nih.gov/publications/nrp/smallbook.htm.

Neufeld, Barbara., and Sara Schwartz. 2004. *Formative Assessment Pilot Implementation: Final Report.* Cambridge, MA: Education Matters, Inc.

Snipes, Jason. 2003. "Using Interrupted Time Series with Comparison Groups to Estimate the Effects of Educational Interventions on Student Achievement." Presented at the annual conference of the Association for Public Policy Analysis and Management, Washington, DC, November 2003.

Snow, Catherine. E., M. Susan Burns, and Peg Griffin. 1998. *Preventing Reading Difficulties in Young Children.* Washington, DC: National Academy Press.

Spinelli, Eileen. 2007. "Apartment Building." In Boston Plan for Excellence, *FAST-R: Formative Assessments of Student Thinking in Reading.* Sample materials. Boston: Boston Plan for Excellence.

About MDRC

MDRC is a nonprofit, nonpartisan social and education policy research organization dedicated to learning what works to improve the well-being of low-income people. Through its research and the active communication of its findings, MDRC seeks to enhance the effectiveness of social and education policies and programs.

Founded in 1974 and located in New York City and Oakland, California, MDRC is best known for mounting rigorous, large-scale, real-world tests of new and existing policies and programs. Its projects are a mix of demonstrations (field tests of promising new program approaches) and evaluations of ongoing government and community initiatives. MDRC's staff bring an unusual combination of research and organizational experience to their work, providing expertise on the latest in qualitative and quantitative methods and on program design, development, implementation, and management. MDRC seeks to learn not just whether a program is effective but also how and why the program's effects occur. In addition, it tries to place each project's findings in the broader context of related research — in order to build knowledge about what works across the social and education policy fields. MDRC's findings, lessons, and best practices are proactively shared with a broad audience in the policy and practitioner community as well as with the general public and the media.

Over the years, MDRC has brought its unique approach to an ever-growing range of policy areas and target populations. Once known primarily for evaluations of state welfare-to-work programs, today MDRC is also studying public school reforms, employment programs for ex-offenders and people with disabilities, and programs to help low-income students succeed in college. MDRC's projects are organized into five areas:

- Promoting Family Well-Being and Child Development
- Improving Public Education
- Raising Academic Achievement and Persistence in College
- Supporting Low-Wage Workers and Communities
- Overcoming Barriers to Employment

Working in almost every state, all of the nation's largest cities, and Canada and the United Kingdom, MDRC conducts its projects in partnership with national, state, and local governments, public school systems, community organizations, and numerous private philanthropies.