December 2008

Using Student Data to Improve Teaching and Learning

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

Janet C. Quint, Susan Sepanik, and Janell K. Smith

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

---

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

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 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.