



RESEARCH FOR *ACTION*

# Making the Most *of* Interim Assessment Data

*Lessons from Philadelphia*

June 2009



**Executive Summary**



## **RESEARCH FOR ACTION**

*Research for Action (RFA) is a Philadelphia-based, non-profit organization engaged in policy and evaluation research on urban education. Founded in 1992, RFA seeks to improve the education opportunities and outcomes of urban youth by strengthening public schools and enriching the civic and community dialogue about public education.*

### **About this study**

This report is part of a larger multi-year research effort led by Research for Action to analyze school reform in Philadelphia under state takeover. The project, *Learning from Philadelphia's School Reform*, is supported with lead funding from the William Penn Foundation and related grants from Carnegie Corporation of New York, the Samuel S. Fels Fund, the Edward Hazen Foundation, the Charles Stewart Mott Foundation, The Pew Charitable Trusts, The Philadelphia Foundation, the Spencer Foundation, Surdna Foundation, and others.

### **Acknowledgements**

We are deeply appreciative of the numerous communities of education researchers, practitioners, and activists of which we are a part. This research project, like so many others, has benefitted from these relationships.

The Spencer Foundation and the William Penn Foundation provided generous financial support for the research reported here.

This report would not have been possible without the cooperation of the School District of Philadelphia—especially staff in the Office of Accountability and Assessment.

Principals, teacher leaders, and teachers in the ten schools in our qualitative sample gave graciously of their time, were patient with our many requests, and responded candidly to our questions. We are also grateful for all that they do for Philadelphia young people every day.

# Competing Claims for Interim Assessments

In urban districts, the press for data-driven decision-making has intensified in the stringent accountability environment of No Child Left Behind (NCLB), where schools look for ways to increase their students' performance on state assessments. These districts increasingly are turning to the significant for-profit industry that has sprung up to sell them curricula aligned with state standards, data management systems, and interim assessments.<sup>1</sup>

In this report, Research for Action examines the use and impact of interim assessments in elementary schools in the School District of Philadelphia. Philadelphia was an early adopter of these assessments, implementing them district-wide in September 2003. Interim assessments are standardized assessments that are administered at regular intervals during the school year in order to help educators gauge student achievement before the annual state exams used to measure Adequate Yearly Progress (AYP) under NCLB. Interim assessments are a central component of Philadelphia's "Managed Instruction System" (MIS) which includes a Core Curriculum and interim assessments which are called Benchmarks in Philadelphia.

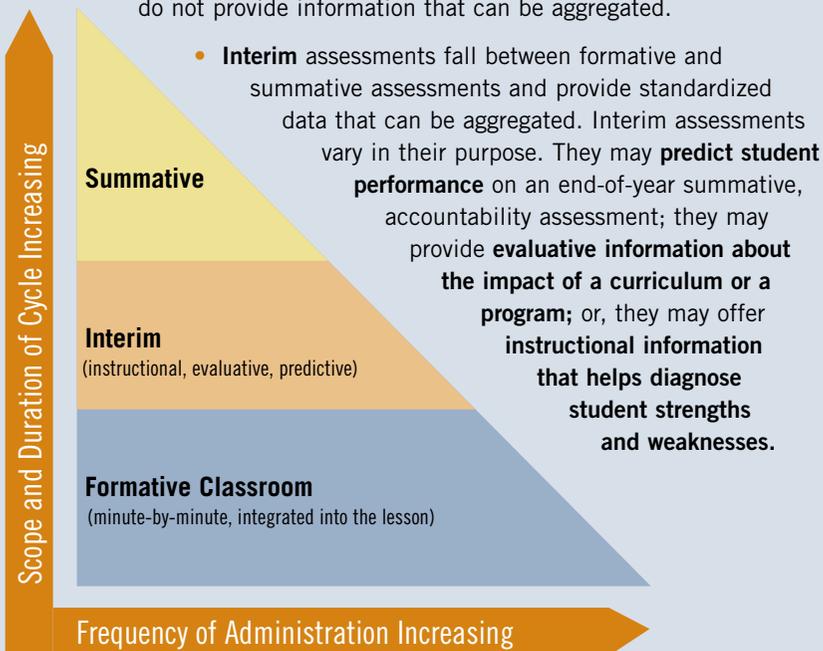
The introduction of interim assessments in urban districts across the country has not been without controversy, as researchers, district leaders, teachers, and the testing industry make conflicting claims for the efficacy of these assessments for guiding instruction and improving student achievement. Many educators and assessment experts, alarmed by the growing market in off-the-shelf commercial products labeled as "formative assessments," insist that the only true formative assessments "must blend seamlessly into classroom instruction itself."<sup>2</sup> (See page 2 for a description of three kinds of assessment – summative, formative, and interim.) There is good evidence that these instructionally embedded assessments have a positive effect on student learning.<sup>3</sup> To date, however, there is not an empirical base for the claim that interim assessments have the power of classroom-based assessments.

In this study, we draw on data from a district-wide teacher survey, student-level demographic and achievement data, and qualitative data obtained from field observations and interviews to examine the implementation and impact of Benchmark assessments. We use an organizational learning framework to offer specific recommendations for what district leaders can do to help school staff make the most of interim assessment results.

## Three Kinds of Assessments

*Perie, Marion, Gong, and Wurtzel (2007)\* have categorized the three kinds of assessments currently in use — summative, formative, and interim — by their intended purposes, audiences, and the frequency of their administration.*

- **Summative** assessments are given at the end of a semester or year to measure students' performance against district or state content standards. These standardized assessments are often part of an accountability system and are not designed to provide teachers with timely information about their current students' learning.
- **Formative** assessments occur in the natural course of teaching and learning. They are built into classroom instructional activities and provide teachers and students with ongoing, daily information about what students are learning and how teachers might improve instruction so that learning gaps and misunderstandings can be remedied. These assessments do not provide information that can be aggregated.



*\*Source: Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007, November). The role of interim assessments in a comprehensive assessment system. Washington, DC: The Aspen Institute.*

# Organizational Learning: A Framework for Examining the Use of Benchmark Assessment Data

The logic behind how interim assessment data can assist teachers is straightforward: a teacher acquires data about what her students have learned; she examines the data to see where her students are strong and weak; she custom-tailors what and how she teaches so that individuals and groups of students learn more; and as teachers across the school engage in this process, the school as a whole improves.

While we recognize the importance of an individual teacher’s use of student performance data, this report views use of student data through a different lens. Specifically, we explore how an organizational learning framework can inform how to strengthen the capacity of schools to capitalize on interim assessments. Organizational learning is the ability of school practitioners to identify and problem-solve around constantly changing challenges. It is an important means to school improvement.<sup>4</sup>

The organizational learning framework that guided our research is illustrated on the next page. On the left, the figure depicts the larger policy context that is likely to influence use of Benchmark data – the school district’s Managed Instruction System (MIS) and the larger accountability environment of No Child Left Behind (NCLB).

The middle box depicts four dimensions of school capacity that are likely to influence data use.

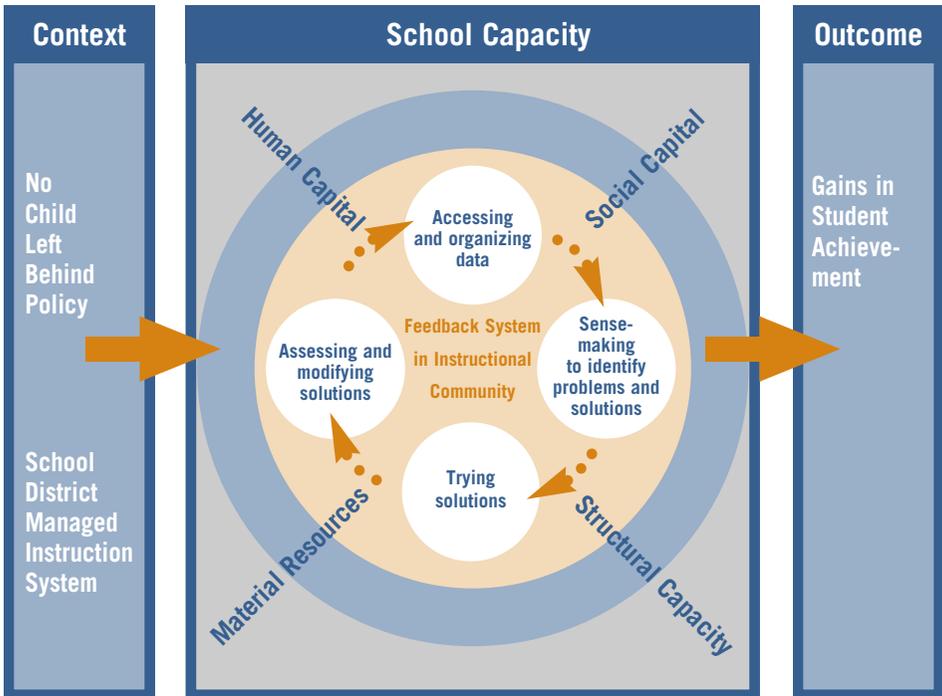
- 1 Human capital (the knowledge, dispositions, and skills of individual actors);
- 2 Social capital (social relationships characterized by trust and collective responsibility for improved organizational outcomes);
- 3 Material resources (the financial and technological assets of the organization);<sup>5</sup> and,
- 4 Structural capacity (an organization’s policies, procedures, and formal practices).<sup>6</sup>

The circle represents a four-step feedback system embedded within overall school capacity and the work of instructional communities. Feedback systems are the necessary foundation for transforming data into actionable knowledge.<sup>7</sup> The four steps in the feedback system are:

- 1 Accessing and organizing data;
- 2 Making sense of data to identify problems and solutions;
- 3 Trying the solutions; and
- 4 Assessing and modifying the solutions.

Multiple feedback systems operate simultaneously in a school; they are most likely to be iterative; and, in the ideal, knowledge generated from one feedback system informs other feedback systems. Finally, in the right hand box, we anticipate that the outcome of these processes will be reflected in gains in student achievement.

### Organizational Learning Framework



## Research Questions

- What were district leaders' expectations for how school staff would use Benchmark data and what supports did they provide to help practitioners become proficient in using data to guide instruction?
- Were teachers receptive to the Managed Instruction System, particularly the Benchmark assessments? Did they use them? Did they find them helpful?
- Did students experience greater learning gains at schools where the conditions were supportive of data use: that is, where the Managed Instruction System was more widely accepted and used and where analysis of student data was more extensive?
- What organizational practices ensure that the use of Benchmark data contributes to organizational learning and ongoing instructional improvement within and across instructional communities?

## Research Methodology: Sources of Data

- A district-wide teacher survey administered in the spring of 2006 and 2007 to all of Philadelphia's approximately 10,500 teachers.
- Student-level demographic and achievement data from standardized tests administered in spring 2005, 2006, and 2007.
- Qualitative data obtained from intensive fieldwork in ten elementary schools (selected from among the 86 schools identified as "low-performing" schools) and interviews with district staff and others who worked with the schools, as well as further in-depth case study analysis of five schools in 2006-2007.

# Philadelphia's Managed Instruction System

*I tell my teachers, 'The Core Curriculum is your Bible.'* —Principal

*Benchmarks replace religion around here.* —Teacher Leader

In response to accountability pressures from No Child Left Behind, School District of Philadelphia's CEO, Paul Vallas, instituted a Managed Instruction System that represented a more prescriptive approach to curriculum, instruction, and assessment than the district had taken in previous reform eras.

## Components of the Managed Instruction System

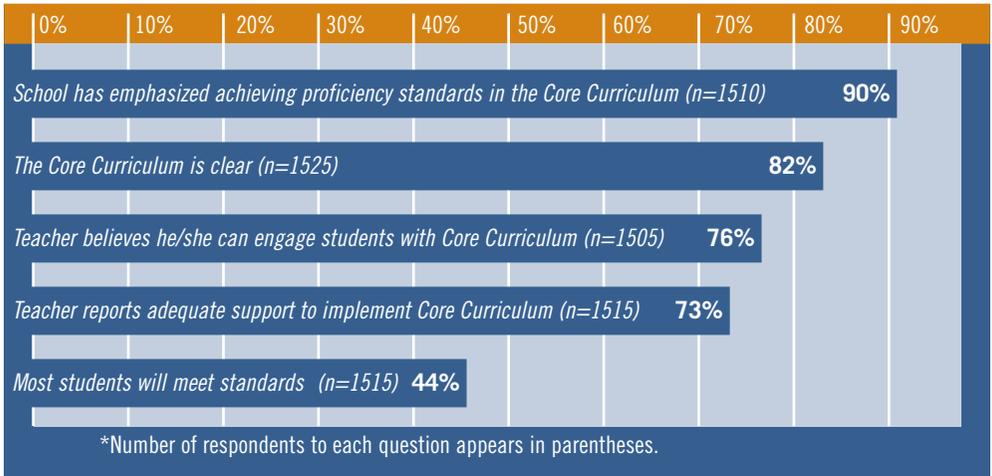
- **The Core Curriculum.** In grades K-8, the Core Curriculum includes performance goals that specify what students must know and be able to do by the end of the school year, while indicating the intermediate levels of proficiency students should attain to be on track to meet state standards. The curriculum includes a specific pacing schedule that is organized by six-week instructional cycles.<sup>8</sup>
- **SchoolNet.** SchoolNet is a district-wide database for the Benchmark assessments and other student data. It is intended to make assessment data immediately accessible to every classroom teacher and building principal, and to provide analysis and instructional tools for educators' use.<sup>9</sup>
- **Benchmark Assessments.** Benchmark assessments were developed by the School District of Philadelphia. They are aligned with the district's Core Curriculum and administered throughout the school year during the fifth week of each six-week instructional cycle. They were intended to give teachers feedback about students' mastery of topics and skills in the Core Curriculum. At the time of our study, Benchmark assessments were conducted in reading and mathematics in grades 3-8, and in science in grades 3, 7, and 8.<sup>10</sup>

## Findings

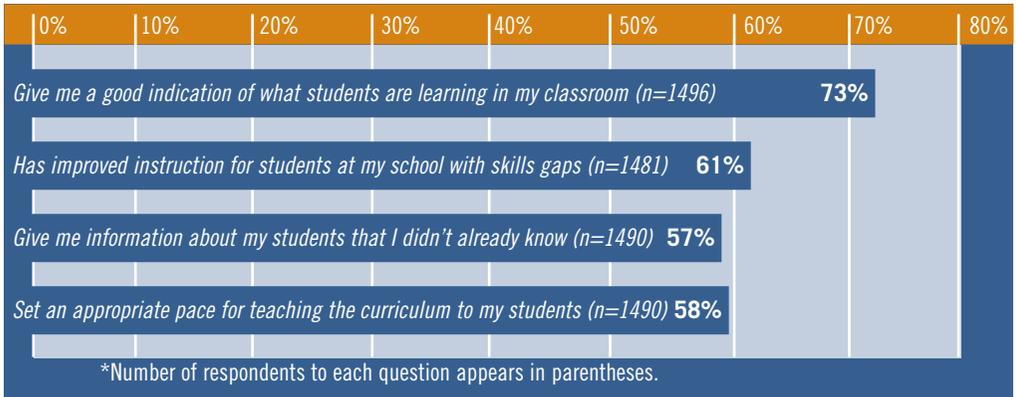
### *Broad Adoption of Core Curriculum and Benchmark Assessments*

- The Managed Instruction System was exerting considerable influence on classroom instruction.** Almost all teachers in grades 3-8 reported that they used the Core Curriculum and data from the Benchmark assessments and most found them useful (see graphs below). Our visits to ten schools between September 2005 and June 2007 corroborated findings from the teacher survey.

### Percentage of Teachers Indicating Agreement with Statements about the Core Curriculum



### Teacher Reports on Benchmarks: Percentage of Respondents Reporting Agreement with Statements about the Core Curriculum



## *Factors Contributing to Teachers' Acceptance of the Core Curriculum and Benchmark Data*

Our qualitative research indicated that a number of factors contributed to Philadelphia's elementary teachers' acceptance of the Core Curriculum and Benchmarks:

- The **historical context** of the School District of Philadelphia and the policy environment of NCLB contributed to teachers' embrace of the MIS. Philadelphia teachers were ready for the Core Curriculum; they saw the value of strong curricular guidance in an era of high-stakes accountability.
- The **design** of Philadelphia's Benchmark assessments had two notable advantages in promoting use in the classroom: alignment with the Core Curriculum and the six-week instructional cycle. Alignment made Benchmark results very relevant to teachers' instructional planning. The sixth week for remediation and extension of topics offered the opportunity for Benchmarks to serve instructional purposes by providing teachers with formative information that could guide their follow-up with students.
- District and school **infrastructure** for supporting use of the Core Curriculum and Benchmark results contributed to teachers' acceptance of the Core Curriculum and Benchmarks. Most teachers reported that their school emphasized the proficiency standards in the Core Curriculum and that they received adequate support for using the Core Curriculum. Most reported that they received the Benchmark data in a timely way and that they had participated in professional development on how to access data. Teachers reported that they had opportunities to review data with colleagues, and had received help from math and literacy teacher leaders in using data.

## *District Intentions and School Realities*

- **District leaders intended that the use of Benchmark assessments would reinforce use of the Core Curriculum and it did.** As one district leader noted, “[It helps to] create some kind of a pacing, and sequence, program” (2005). Principals and teachers confirmed that the Benchmarks provided a curriculum roadmap with specific destinations demarcated along the way. One principal described the reaction of teachers at her school: “When teachers saw kids’ results on the Benchmarks, they really knew ‘I didn’t cover this. I should have covered this.’” At another school, a fourth grade teacher remarked,

*Tests that I give in the classroom are maybe targeting one story or one particular skill whereas [Benchmarks] give you the big picture of what you have done in the last 6 weeks and whether you achieved what you were supposed to teach them in the last 6 weeks (2007).*

- **District leaders' intentions regarding the instructional use of Benchmark data were based on some assumptions that did not necessarily match school realities.** This mismatch limited the instructional impact that leaders expected. First, district leaders expected teachers to test for mastery again at the end of the re-teaching week. Our qualitative research indicated that classroom-based assessments during the sixth week were infrequent. Second, the design of the Managed Instruction System assumed strong leadership capacity at the school level.

In the next sections, we delve into whether these expectations of school leaders were realistic.

## The Impact of Benchmarks on Student Achievement

The ultimate goal of systematically tracking student progress is to increase student learning. To address the question of whether students experienced greater learning gains at schools where the Managed Instructional System was more widely accepted and analysis of student data was more extensive, we used two types of data: teacher surveys and student scores on standardized tests for students who were in grades 4 through 8 during 2005-2006 and/or 2006-2007.

To create a measure of student academic growth, we examined changes in students' performance on standardized tests given at the end of successive school years. This strategy sometimes is known as a value-added approach because it examines the "value added" to learning by attending school in a given year. A complicating factor for this analysis was that some of the tests students took in different years were not vertically scaled. To address this incompatibility between tests, we converted the student's score on each test to a ranking within the district. Students who made learning gains relative to other students in the district in a given year received a positive value for their learning during that

year; those whose learning did not keep up with other students in the district received a negative value for the year's learning.

We examined the associations between student learning and measures of school leadership, professional climate, and the Managed Instruction System derived from a district-wide survey of teachers. (See specific measures in the box on page 11.)

## Findings

### *Schools that are Good in One Thing Tend to be Good in Another*

Our first analytic step was to examine the extent to which teachers' reports about each school condition were correlated with their reports about other school conditions – in other words, whether school characteristics tend to occur as “packages.”

- **All of the school characteristics—instructional leadership, professional climate, and use of, and satisfaction with, the Managed Instruction System — were positively correlated.** These correlations suggest that schools that have one type of good quality—such as strong instructional leadership—tend to have other good qualities, such as a strong professional climate. These also tend to be the schools that provide more support for data use and where teachers are more satisfied with, and make greater use of, the Managed Instruction System. Instructional leadership had the highest correlation with other characteristics.

### *Teacher Satisfaction with Benchmarks Not Predictive of Student Growth*

Our second step was a regression analysis to examine associations between student achievement and each school condition separately, controlling for individual student characteristics and the percentage of low-income students at the school.

This analysis showed that almost every variable is a statistically significant predictor of learning growth. It indicated that, generally speaking, the instructional leadership and professional climate variables had slightly larger impacts on achievement than the Managed Instruction variables. Importantly for this study:

## Key Survey Variables

### *School Leadership*

- **Instructional Leadership.** The quality of school leadership in the use of student data, monitoring of instructional quality, and setting clear goals and high expectations for teachers. (The emphasis on leadership as related to use of student data is an important distinction for this study.)

### *Professional Climate*

- **Commitment to the School.** The extent to which teachers would prefer to work at their school rather than at any other school and would recommend their school to parents.
- **Instructional Innovation and Improvement.** Whether teachers' colleagues try to improve their teaching and are willing to try new strategies.
- **Teacher Collective Responsibility.** Teachers' sense of responsibility for their students' academic progress and for the overall climate of the school.

### *Satisfaction with, and Use of, the Managed Instruction System*

- **Use of the Core Curriculum.** How much the Core Curriculum guides teachers' topic coverage, instructional activities, and assessment strategies.
- **Satisfaction with Benchmarks.** Teachers' beliefs and attitudes about whether the Benchmark assessments provide useful information about student progress in a timely and clear manner.
- **Collegial Instructional Responses to Student Data.** How often teachers met with colleagues at their school to discuss re-teaching a subject or re-grouping students, based on examination of Benchmark scores.
- **Technology Access and Support.** Classroom Internet access, working computers, and technology support for teachers.
- **Professional Development on Data Use.** Whether the school offered professional development on how to access and interpret student performance data.

## Key School Variables Predicting Growth in Student Learning

	Reading 2005-06		Math 2005-06		Reading 2006-07		Math 2006-07	
	estimate	p*	estimate	p	estimate	p	estimate	p
<b>Instructional Leadership</b>	<b>.080**</b>	.010	<b>.100</b>	.002	<b>.150</b>	.000	<b>.150</b>	.000
Use of the Core Curriculum	<b>.150</b>	.002	<b>.100</b>	.030	.040	.300	.001	.976
<i>R-squared at Level 2 (school level)</i>	.080		.060		.120		.090	
<b>Collective Responsibility</b>	<b>.170</b>	.000	<b>.170</b>	.000	<b>.130</b>	.000	<b>.140</b>	.000
Use of the Core Curriculum	<b>.120</b>	.004	.080	.060	.080	.053	.030	.476
<i>R-squared at Level 2 (school level)</i>	.130		.100		.090		.070	

\*The p-value is the probability that the estimate is simply the result of chance.

\*\* Statistical significance is indicated in bold type.

- A measure of satisfaction with Benchmarks was not significantly associated with either reading or math achievement growth** for either 2005-2006 or 2006-2007 (although it approached statistical significance at  $\alpha = .05$  in 2006-2007). Likewise, a measure of collegial instructional responses to student data was not a significant predictor in 2006-2007. The direction of the coefficients was positive in all cases.

The organizational learning framework that informs this study provides some insight into the weak relationship between satisfaction with Benchmarks and achievement. The measure of satisfaction with Benchmarks tells us about only a small piece of the feedback system: whether the teachers felt that Benchmarks provided useful, clear, and timely information about student progress. It does not tell us whether teachers had good ideas about how to respond to the data. As the qualitative data of the following chapters show, the ability of teachers to make sense of the data and plan appropriate instructional responses is heavily contingent on school resources, especially the quality of leadership and support provided by the principal and content area teacher leaders.

## *Instructional Leadership, Collective Responsibility, and Use of the Core Curriculum Were the Strongest Predictors of Student Growth*

In our final step, we used multivariate regression to identify school characteristics that had an especially strong relationship with achievement. Our purpose in so doing was to assess whether there were particular organizational characteristics on which education leaders could focus in order to help teachers make the most of student data. This analysis showed that:

- **Instructional leadership and teacher collective responsibility were the school leadership and professional climate variables that had the strongest and most consistent relationships with student achievement across years and subjects.**
- **Use of the Core Curriculum was also a statistically significant predictor of student achievement growth** when each of the five Managed Instruction variables was added to a regression with either the instructional leadership or collective responsibility measures.

## **Implications**

Importantly, we found that certain levers for improvement in student learning were stronger than others. Use of Philadelphia’s Core Curriculum proved to be a strong predictor of student growth, likely because of its contribution to instructional coherence in a school. Instructional leadership and collective responsibility were also strong predictors of learning growth.

The instructional leadership and collective responsibility measures imply that translating student data into student achievement requires a strong learning community at a school. In such a community, school leaders and faculty feel accountable to one another, they are diligent in monitoring student progress, and they are willing to use data as a starting point for inquiry. While Benchmarks may be helpful, they are not in themselves sufficient to bring about increases in achievement without a community of school leaders and faculty who are willing and able to be both teachers and learners.

# Making the Most of Benchmark Data

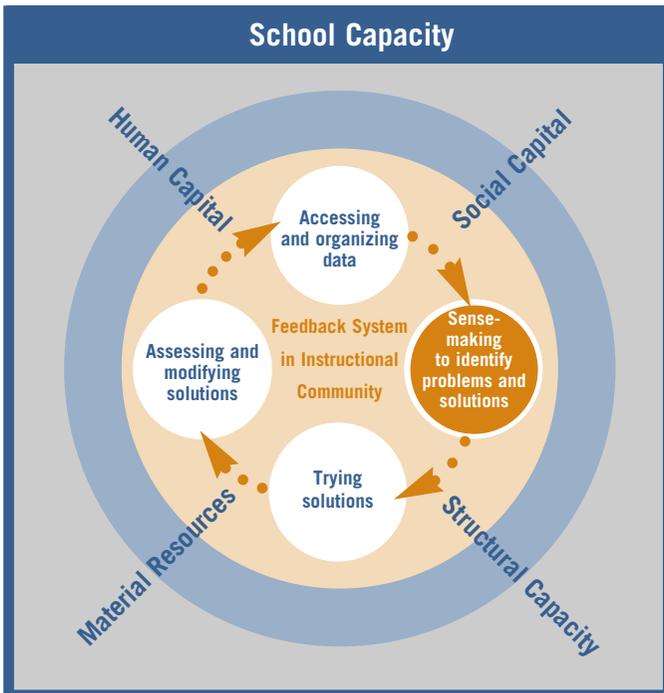
In this section, we draw on our observations of formal instructional communities, in this case, grade group meetings, to shed light on the role of leadership and professional community in the interpretation of Benchmark data. How did practitioners make sense of the data and what action did they take as a result? In the four-step feedback system, sense-making is represented in the second step, “sense-making to identify problems and solutions.”

## Findings

### *Three Kinds of Sense-Making: Strategic, Affective, and Reflective*

Our observations of grade group meetings suggest that practitioners engaged in three major types of sense-making.

## Organizational Learning Framework



- **Strategic sense-making**, which dominated talk about Benchmark data, focused on the identification of short-term tactics that help a school reach its Adequate Yearly Progress (AYP) targets. Examples included conversations about “bubble students” who have the highest likelihood of moving to the next level of performance; conversations on improving test-taking conditions and test-taking strategies; and identifying strengths and weaknesses that cut across grades and classrooms so that leaders could allocate resources (staff, materials, and time) in ways that increased the odds that the school will meet its AYP goal.
- **Affective sense-making**, the second most prevalent kind of sense-making, addressed teachers’ and leaders’ professional agency, their beliefs about their students, their moral purpose, and their collective responsibility for students’ learning. These were often supportive and motivational conversations about their own role as teachers in eliciting students’ best efforts.
- **Reflective sense-making** was the least observed, but offered the most promise for building increased school and classroom instructional capacity because it focused on teachers’ learning from Benchmark data. Reflective sense-making occurred when teachers and leaders questioned and evaluated the instructional practices that they employed in their classrooms and their school. They connected what they were learning about what their students knew and didn’t know to key concepts in the Core Curriculum and identified resources that would help them strengthen instruction of those concepts.

### *From Sense-Making to Action*

The quality of practitioners’ sense-making influenced the quality of the actions that they took based on the data. As a consequence of reviewing Benchmark data, practitioners planned actions that included:

- **Identifying students who were likely to move** from “Basic” to “Proficient” or from “Below Basic” to “Basic” and targeting them for special interventions in order to increase the likelihood that the school will make AYP.
- **Identifying skills and concepts to be re-taught** in the sixth week of the instructional cycle or in subsequent units.
- **Re-thinking classroom routines** in order to emphasize greater student independence, motivation, and responsibility for their own learning.

- **Identifying content and pedagogical needs of teachers** to inform opportunities for continued professional learning and other supports that addressed those needs.

Each of these planned actions emerged from paying attention to data. However, the depth and the quality of the actions varied considerably. When practitioners focused on superficial problems—described as “the low-hanging fruit” by principals in our study—their intervention strategies were likely to be mundane. The fact that these schools had been identified as “low-performing” influenced practitioners’ perceptions of why examining data was important; it would help them to address the primary problem that they felt compelled to solve: how to make AYP. They brought the imperative to “do something”—some might say “do anything”—to their discussion and interpretation of Benchmark data.

### *Opportunities*

There are opportunities for district and school leaders to strengthen the impact of Benchmark data on teacher and student learning.

- **Balance strategic, affective, and reflective sense-making.**
- **Identify students’ conceptual misunderstandings.**
- **Capitalize on Benchmark discussions to build teachers’ “pedagogical content knowledge.”**<sup>11</sup> Teachers with strong pedagogical content knowledge understand what teaching approaches fit the content being taught; their deep understanding of content makes it possible for them to explain disciplinary concepts to students and to craft learning tasks that build students’ conceptual understanding; their broad repertoire of instructional strategies provide them with options to help students with different learning needs. The alignment of Benchmark assessments with the Core Curriculum offer the opportunity for teachers to look at results with an eye towards strengthening their pedagogical content knowledge.
- **Focus on curricular standards and instructional approaches that would address these standards, rather than on single test items.**
- **Follow up in classrooms to see if and how instructional actions are being implemented and to offer support.**

## Findings: The Case of Mahoney Elementary School

Our quantitative analyses indicated how difficult it is to unravel the multiple factors that influence the impact of Benchmark data on student achievement because a school that is good in one area is likely to be good in another. Below, we focus on the Mahoney Elementary School (a pseudonym), a school in which students made strong achievement gains during the time of our research, to illuminate how these factors interact in a school. We identify exactly what the principal and content area teacher leaders did to establish strong feedback systems and robust sense-making processes. The analysis focuses especially on building strong linkages between the steps in the feedback system depicted by the arrows in the figure on page 14.

### *Structured Grade Group Meetings Focused on Teacher Learning*

- **Grade group meetings were a key opportunity for teachers to learn from Benchmark data.** Examination of Benchmark data in grade groups at Mahoney Elementary School was a powerful vehicle for reinforcing the use of the Core Curriculum, for focusing teachers' attention on the proficiency standards of the curriculum, and for organizing conversations about student achievement in which teachers were expected to talk about ways to improve their teaching. According to teachers and school leaders, grade group meetings focused on analysis of data and/or reflection on instruction. In fact, grade group meetings were described by school leaders as the most important site in the school for teacher learning and were re-named Professional Learning Communities during the course of our study.

Meetings were held weekly and the principal and teacher leaders were consistently in attendance. (In addition, because Mahoney's leaders attended all of the grade group meetings, as well as meetings of other groups, they were able to spread insights and knowledge from one setting across the school.) The principal set high expectations for teachers' preparation for, and participation in, grade group meetings.

The meetings were tightly structured and organized around a guiding question. They included strategic, affective, and reflective sense-making. Reflective sense-making was frequent as teachers consistently questioned their own past practices, and both sought and shared new ways to approach content and new ways to help their students learn. Meetings ended with a school leader summarizing next steps.

### *School Leaders Established Consistent Feedback Systems*

- **The principal established a culture of data-driven decision-making and worked with teacher leaders to integrate review of Benchmark data into strong feedback systems.** After new practices had been identified in grade group meetings, school leaders followed up in classrooms to make certain that teachers were using them. They provided resources and coaching. One Mahoney teacher described how helpful teacher leaders were, “Knowing that my literacy leader is there [is important], and if I say to her, ‘You know, I’m not really sure how I’m going to do this lesson,’ she’s always there and very helpful.” Further, school leaders discussed with teachers how the new instructional strategies were working and how to modify them when appropriate. These steps became routine at Mahoney, thus making feedback systems strong and coherent and ensuring organizational learning and instructional improvement. (See figure page 20)

### *School Leaders Laid the Groundwork for Instructional Coherence*

- **Organizational routines at Mahoney promoted instructional coherence.** The principal and teachers leaders at Mahoney had a clear understanding of the powerful connection between the Benchmarks and the Core Curriculum. The principal allocated resources for knowledgeable teacher leaders who were well-versed in the content and assessment issues in their own curricular areas. Together, the principal and teacher leaders established a set of structures and practices that ensured that Benchmark data were used as part of a process for ensuring high quality instruction. They made grade group meetings sites for ongoing teacher learning and established feedback systems that included assessment of the interventions that were used to boost student learning.

All of these steps laid the groundwork for instructional coherence, a key feature of schools shown to make student learning gains in Chicago and elsewhere.<sup>12</sup> Instructional coherence requires a common instructional framework that “guides curriculum, teaching, assessment, and learning climate” and includes expectations for student learning, teaching materials, and evaluation of instructional strategies.<sup>13</sup>

# Making the Most of Benchmark Data: Implications for Philadelphia and Beyond

Federal, state, and district policies that use standardized tests as the central metric for accountability have fueled the fervor for student achievement data, especially in districts with large numbers of academically failing students. Controversy notwithstanding, the use of interim assessments by large urban school districts is on the rise. The findings from our research on the use and impact of these assessments in Philadelphia's K-8 schools will not end the debate. They do, however, offer formative lessons to Philadelphia and beyond about the design, implementation, and impact of interim assessments.

Teachers' high degree of satisfaction with the usefulness of Benchmarks was not a statistically significant predictor of student achievement gains; instructional leadership, collective responsibility, and use of the Core Curriculum were. Our analyses coupled with an organizational learning framework offer insights into why these findings were the case. If practitioners' sense-making does not lead them to seek or develop new and robust instructional interventions, if these interventions are not actually implemented or not implemented well, or if their effectiveness is not assessed, then teaching and learning is not likely to improve. Data can make problems more visible, but only people can solve them.

Below we offer recommendations for how districts can make the most of interim assessment data.

## Investing in School Leaders

- **As “learning leaders,” principals and teacher leaders need to know how to facilitate “learning” discussions about data.** School leaders can make a real difference in helping staff move beyond data use as a narrow exercise in preparing to “teach to the test.” But to do so, they must know how to frame conversations about assessment data so that teachers understand the connections to larger school improvement priorities and to the curriculum. They need to know how to pose questions in ways that invite teachers to talk openly about: curriculum concepts, how their students learn best, what instructional practices have worked and those that have not, what additional curricular resources they need, what they need to learn about content, and where they might seek evidence-based instructional strategies that would address the learning weaknesses of their students. They also need to be able to steer

## Making the Most of Benchmark Data at Mahoney Elementary School

### Engaged Principal:

- Built strong leadership team by allocating full-time teacher leaders in math and reading
- Worked with teacher leaders to develop long-term instructional improvement strategies and shorter-term priorities for their work with classroom teachers
- Emphasized data-driven decision-making
- Actively attended grade group meetings
- Established meeting routines that were used across the school
- Set high expectations for teachers' preparation for and participation in grade group meetings
- Used discussions of Benchmark data in grade groups to reinforce importance of proficiency standards of Core Curriculum
- Encouraged strategic, affective, and reflective sense-making, with the strongest emphasis on reflective sense-making
- Worked with teacher leaders to spread insights and knowledge about instruction across the school

### Full-time Math and Reading Teacher Leaders:

- Well-versed in the Core Curriculum, the Benchmark assessments, and the Pennsylvania System of School Assessment (PSSA) exams and understood the connections and disconnections among the three
- Continuously enhanced their knowledge of research-based instructional strategies that supported effective use of the Core Curriculum
- Helped teachers interpret Benchmark data
- Recommended specific instructional strategies based on the Benchmark results
- Moved in and out of classrooms to see if teachers were implementing curriculum well and provided coaching and demonstration where needed
- Gathered resources to supplement the curriculum
- Collaborated with principal on long- and shorter-term instructional strategies to meet school's goals

### Effective Grade Group Meetings:

- Held weekly
- Principal, teacher leaders, and classroom teachers came prepared to participate
- Discussions included strategic, affective, and reflective sense-making
- Highly structured meeting routines, focused on instructional issues and ongoing professional learning of staff
- Began with an agenda and guiding question
- Ended with school leader summarizing next steps
- Follow-up notes distributed across the school

teachers away from inappropriate use of interim assessment data. School leaders need opportunities to practice these skills and receive feedback. Understanding the value and purposes of the different types of sense-making identified in our research—*affective, strategic, and reflective*—and how to use them, offer a framework for such training.

- **As “learning leaders,” principals and teacher leaders need to know how to allocate resources and establish school organizational structures and routines that support the work of instructional communities and assure that the use of Benchmark data is embedded in the feedback systems necessary for organizational learning.**

- School budgets need to accommodate full-time content area teacher leaders.
- School schedules need to accommodate regular meetings of grade groups.
- Principals and teacher leaders need to be at these meetings and, with teachers, establish meeting routines that include agendas, discussion protocols with guiding questions, and documentation of proceedings.
- Follow up to the meetings is crucial. School leaders need to visit classrooms to see if and how teachers are using instructional strategies and to offer resources and coaching so that teachers can deepen their understanding of curriculum content and pedagogy.
- Assessing the impact of interventions is also crucial. Important steps include helping teachers to design classroom-based assessments for use during the sixth week of instruction and examining the quality of common interventions such as tutoring and after-school remediation programs.

## Designing Interim Assessments and Supports for Their Use

This research suggests additional lessons about designing interim assessments and the resources that will encourage and support the use of data from those assessments.

- **As districts and schools develop organizational structures, processes and tools to support the use of interim assessment data, they need to ask themselves two questions:**
  - Do the structures, processes and tools support the review of data as a collective learning activity of instructional communities? Do they support the

review of data as an activity which helps teachers deepen their pedagogical content knowledge and understand what their students know and how they learn?

→ Do they support the multiple steps of feedback loops? Do they encourage leaders' follow-up work with teachers in the classroom? Do they promote the assessment of interventions and modifications where necessary?

- **In order to capitalize on Benchmarks to fulfill instructional purposes, district leaders should develop interim assessments that:**

→ Test for a range of thinking skills—knowledge, comprehension, application, synthesis and evaluation;

→ Offer distractor answers that provide insight into what students do not understand;

→ Include open-ended items.

The most important message from this research is that the success of even a well-designed system of interim assessments is dependent on the knowledge and skills of the school leaders and teachers who are responsible for bringing the system to life in schools. Continued investment in interim assessments is not likely to yield the benefits intended without concomitant investments in the learning of school leaders and teachers.

# Endnotes

- 1 Burch, P. (2005, December 15). The new education privatization: Educational contracting and high stakes accountability (ID Number: 12259). [On-line]. Retrieved on January 20, 2006, from <http://www.tcrecord.org>.
- 2 Cech, S. J. (2008, September 17). Test industry split over 'formative' assessments. *Education Week*, 28(4), 1, 15, p. 1.
- 3 Black, P. & Wiliam, D. (1998, October). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* 80(2).
- 4 Little, J. W. (1999). Teachers' professional development in the context of high school reform: Findings from a three-year study of restructuring high schools. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec; Wagner, T. (1998). Change as collaborative inquiry: A 'constructivist' methodology for reinventing schools. *Phi Delta Kappan*, 80(7), 378-383; Knapp, M. S. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research*, 67(2), 227-266; Spillane, J. P. & Thompson, C. L. (1997, June). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Education Evaluation and Policy Analysis*, 19(2), 185-203.
- 5 Spillane, J. P. & Thompson, C. L., 1997.
- 6 Century, J. R. (2000). Capacity. In N. L. Webb, J. R. Century, N. Davila, D. Heck, & E. Osthoff (Eds.), *Evaluation of systemic change in mathematics and science education*. Unpublished manuscript, University of Wisconsin-Madison, Wisconsin Center for Education Research.
- 7 Halverson, R. R., Prichett, R. B., & Watson, J. G. (2007). *Formative Feedback Systems and the New Instructional Leadership* (WCER Working Paper No. 2007-3). [On-line]. Retrieved on July 16, 2007, from <http://www.wcer.wisc.edu/publications/workingPapers/index.php>.
- 8 Travers, E. (2003, September). Philadelphia school reform: Historical roots and reflections on the 2002-2003 school year under state takeover. *Penn GSE Perspectives on Urban Education*, 2(2).
- 9 Students' families also have limited access to SchoolNet data through the system's FamilyNet tool to obtain up-to-date information on their children's test scores (including Benchmark assessments), report card grades, and attendance.
- 10 Journalistic accounts of the use of interim assessments (largely in *Education Week*) led us to the conclusion that in most school districts that use interim assessments the tests are given between monthly and three times a year. Aside from Philadelphia, we did not identify any other districts where time was set aside explicitly for addressing weaknesses identified from analysis of interim assessment data.
- 11 Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
- 12 Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001, January). *Improving Chicago's schools: School instructional program coherence benefits and challenges*. Chicago: Consortium on Chicago School Research; Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, 23(4), 297-321.
- 13 Newmann et al., 2001, January.

## About the Authors

**Jolley Bruce Christman, Ph.D.** is a Principal of Research for Action and the Principal Investigator on this project. Her research focuses on instructional communities, school leadership, organizational learning, privatization in public education, and the use of research to inform policy and practice.

**Ruth Curran Neild, Ph.D.** served as a Co-Principal Investigator on this project. She is a Research Scientist at the Johns Hopkins University. Her scholarly interests, broadly speaking, focus on improving educational outcomes for urban youth through transforming their school experiences.

**Katrina Bulkley, Ph.D.** served as Co-Principal Investigator on this project. She is an Associate Professor of Educational Leadership at Montclair State University. Her work explores the role of governance changes in educational reform.

**Suzanne (Sukey) Blanc, Ph.D.** is an educational anthropologist and a former middle school math teacher. She is a senior research consultant to Research for Action and is the founder of Creative Research and Evaluation Services.

**Roseann Liu** is a Ph.D. student at the University of Pennsylvania's Graduate School of Education pursuing a dual degree in anthropology and education. She is interested in the cultural productions of youth in transnational and diasporic communities.

**Cecily Mitchell** is a Senior Research Assistant at Research for Action. She is interested in school-based interventions to improve the educational experience and outcomes for students who have been marginalized within the educational system.

**Eva Travers, Ph.D.** is Professor Emeritus at Swarthmore College where she taught urban education and education policy.





**RESEARCH  
for *ACTION***

3701 Chestnut Street  
Philadelphia, PA 19104  
Tel (215) 823-2500  
Fax (215) 823-2510  
[www.researchforaction.org](http://www.researchforaction.org)