

STUDYING MICRO-PROCESSES OF SCHOOL-BASED EDUCATORS' USE OF DATA IN CROSS-SCHOOL QUALITATIVE RESEARCH

**Suzanne Blanc, Research for Action
Katrina Bulkley, Montclair State University
Jolley Bruce Christman, Research for Action
Ruth Curran Neild, Johns Hopkins University
Eva Travers, Swarthmore College**

Paper presented at the annual meeting
of the American Education Research Association, April 2007, Chicago

INTRODUCTION

*“Benchmarks replace religion around here.”
–McKinley Math Teacher Leader, May 2006*

While there is a great deal of enthusiasm for collection and use of student data by educators, data-driven decision-making represents a new way of thinking for most teachers and principals, who traditionally have based their instructional decisions on “intuition, teaching philosophy, or personal experiences” (Cromey, 2000, p. 3). Characterizations of data-driven decision-making typically describe a logical progression of fact-finding, analysis, and decisions about action. However, theory and research on school change suggest that the translation of student data into instructional decisions that increase student achievement is likely to be complex and fraught with uncertainty, partial understandings, and subtle differences in use and meaning. Little is known about how teachers and principals actually try to make sense of data and how they apply this knowledge to making instructional decisions.

In order to gain a deeper understanding of how educators learn to use data, we began a multi-year research project involving case studies of “low-performing” elementary and middle schools that were recommended to us as sites where relatively

sophisticated use of data was occurring. Our goal was to understand the process of learning to use student data as well as the policies, interventions, and practices that support educators as they begin to think in new ways about making instructional decisions. In this paper, we describe some of the methodological and definitional challenges that arose as we undertook this research project and the decisions we made as we wrestled with these issues.

Methodological Challenges

Specifically, we focus on two challenges that we have encountered. The **first challenge** is how to understand and describe the events and subtle meanings and understandings associated with the micro activities in which school leaders and instructional communities engage as they work to make sense of data and how various aspects of school capacity influence data use. Spillane and his colleagues (2001) call for more detailed examinations of educators' work. They distinguish between "macro functions" (e.g., the use of data to drive decision-making) and "micro tasks" (e.g., displaying the data and formulating substantive questions about the data and their implications) and urge researchers to understand how educators "define, present, and carry out these micro tasks, exploring how they interact with others in the process" (Spillane et al 2001:24). A **second challenge** relates to our goal of linking educators' examinations of student performance data to what we initially called "instructional decisions" and finding examples of such decisions and connecting them to educators' analysis of data.

At its heart this paper is an attempt to make apparent to others how we have worked to triangulate both data sources and research methods to enhance the rigor,

breadth and depth of our research. In doing so, we seek to address the concerns of qualitative researchers who urge their colleagues to open up their methodological and analytic procedures to build credibility for the trustworthiness of their findings (Anfara, Brown, and Mangione, 2002). Specifically, we focus on the ways that we have been triangulating three data sources and methods:

- the collection and review of *artifacts* – in this case, tools available on a computer-based data management system and those developed by the district itself to help school staff analyze and interpret student performance data;
- *observations* of settings in which teachers and administrators talk about data;
- *in-depth, semi-structured interviews* with classroom teachers, teacher leaders, and principals about using data.

Our story is necessarily incomplete as we are in the second year of a 2 ½ year study. Nevertheless, our intention is to shed light on our process of puzzling through how to capitalize on multiple methods and data sources

Conceptual Framework

The conceptual framework for this research (shown in Figure 1) recognizes that data use takes place within multiple contexts, including: policies at the *federal and state levels* – especially the No Child Left Behind legislation, policies and programmatic interventions designed and implemented by the *school district and education management providers* operating in Philadelphia, the *overall school organization* and its processes for school improvement, and the *instructional communities within schools* established to advance effective data use practices. While we acknowledge these other levels of influence, in this paper we focus on how school-level educators use data.

Increasingly, educational researchers have framed school improvement as a problem of organizational learning (Little, 1999; Wagner, 1998; Knapp, 1997; Spillane & Thompson, 1997). The concept of a “learning organization” emerged in the 1980s to describe organizations that were able to refine their understandings of the challenges that confronted them and to respond successfully by systematically examining data in an organizational culture characterized by collaboration, openness, and inquiry (Senge, 1990; Argyris & Schon, 1978). Recent research on schools as learning organizations has begun to examine the multiple elements of school capacity that affect data use and organizational learning (Watson & Mason, 2003; Leithwood et al., 2001). School capacity is the ability of the school to advance the quality of the educational program and to help all students reach standards (Massell, 2000; Newmann, King, & Youngs, 2000; O’Day, Goertz, & Floden, 1995). The literature suggests that organizational capacity has four dimensions—human capital (the knowledge, dispositions, and skills of individual actors); social capital (social relationships characterized by trust and collective investment in improved organizational outcomes); material resources (the financial and technological assets of the organization) (Spillane & Thompson, 1997; Newmann, King, & Rigdon, 1997; Corcoran & Goertz, 1995; Louis, Kruse, & Associates, 1995; Talbert & McLaughlin, 1994); and structural capacity (an organization’s policies, procedures, and formal practices) (Century, 2000).

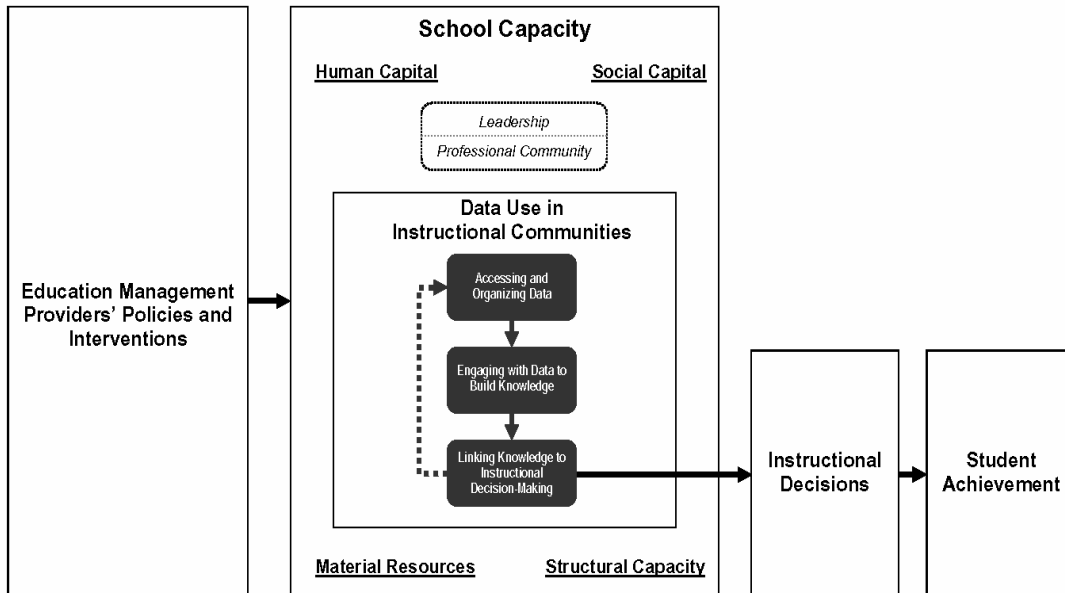
McLaughlin and Talbert’s (2002) evaluation of the efforts of the Bay Area School Reform Collaborative (BASRC) points to the relevance of school capacity for our research. Schools with weak organizational capacity had the greatest difficulty capitalizing on BASRC’s “cycle of inquiry” as a strategy for informing school

improvement. At least two studies (McLaughlin & Talbert, 2002; Mason, 2003) have also indicated that school leadership and professional community—which are dimensions of human and social capital—are important influences on school capacity.

Organizational learning posits that informal communities of practice are critical to an organization’s ability to innovate. In such communities of practice, workers are engaged in a *joint enterprise*; they are bound together in *mutual relationships* characterized by common norms for participation and performance; they employ a *shared repertoire of habits of mind* that involves them in thinking through problems of practice (Brown and Duguid, 2000; Wenger, 1998; Lave & Wenger, 1990). In this project, we focus particularly on collaborative data use within school-level *instructional communities*. We define an instructional community as a group of school staff members that has 1) established parameters for group membership, 2) allotted time for group meetings, and 3) shared responsibilities related to accessing and organizing data, examining and interpreting data, and/or linking data to instructional decisions. In Philadelphia, the most salient groups that meet these criteria at the elementary and middle school levels are school instructional leadership teams and grade level groups of teachers.¹

¹ It should be made clear at this point that “instructional communities” are not synonymous with “professional community,” which is also an important concept in the research literature on schools and in this proposal. The concept of professional community refers to the quality of collaboration among members of a school staff—openness, trust, shared responsibility for student progress (Kruse, Louis, and Bryk, 1995)—while instructional communities are specific structural arrangements that may or may not feature or foster the desirable qualities of professional communities.

Figure 1: Conceptual Framework



Our study focuses on student performance data provided through externally-mandated, standardized assessments and school-based assessments that are used beyond an individual classroom. For this paper, we have chosen to narrow our focus further to data from formative, interim assessments that are aligned with a district’s curriculum and that all students in grades 3-8 take every six weeks. We use “formative” to mean assessments that are used primarily to inform ongoing instructional decisions rather than assessments that are used to measure outcomes or for accountability purposes.

THE STUDY

Our study focuses on data use in the School District of Philadelphia which experienced a “state takeover” in December 2001. Shortly after the state assumed control of the district, 86 elementary and middle schools were identified as especially low-

performing and in need of special intervention. Forty-five of these schools were targeted for private management by a group of for-profit managers, non-profit managers, and university partners. A special office of the school district itself – the Office of Restructured Schools – designed and implemented an intense intervention for 21 of the 86 schools. All of these schools received additional per pupil funding ranging from \$750-\$450. Sixteen of the schools, called "the sweet sixteen" and initially slated for outside interventions argued successfully that they had improved student performance over the previous two years. They were permitted to run their schools, but they still received additional per pupil funding. All of these schools are under significant pressure to improve student achievement and to use student data to accomplish that goal.

Our original intent was to take advantage of the diverse provider model, using the natural variation that we assumed it would introduce in program interventions to learn what kinds of guidance and support were most effective in helping teachers and administrators use data well. The potential for variation in data use has been mitigated by the way in which the diverse provider model has evolved in Philadelphia. Nevertheless, Philadelphia remains an interesting site for this study because it became an early adopter of an Instructional Management System, a step that other urban districts have now taken. All of the providers have adopted, in large part, the Instructional Management System (IMS) that the district created for its schools.² The IMS includes a core curriculum in three subject areas and is aligned with state standards. Additionally, interim Benchmark Assessments of student mastery of the core curriculum are administered at the end of a

² Only Edison Schools, Inc., arrived with a fully developed instructional and assessment model. And so, even though, the providers' contracts with the district granted them full authority over curriculum and instruction and thus the potential for distinctive innovation, in fact, the threat of NCLB sanctions coupled with the lack of fully developed instructional models compelled providers to adopt the district's IMS.

five week instructional cycle. During the sixth week, teachers review the data reports generated to determine their students' weaknesses and shape instructional interventions to be used in the sixth week. The system was designed to support teachers and principals as they work to meet their Adequate Yearly Progress (AYP) targets under the No Child Left Behind legislation. Lessons from implementation of its Instructional Management System will be useful to others across the country.

Research Questions and Methodology

1. What do school leaders and formal instructional communities in schools actually do as they work with data to inform instructional improvement?
2. What kinds of instructional decisions do they make based on their interpretations of data?
3. What kinds of policies, interventions, and resources positively influence how school leaders and instructional communities use data?
4. What are the relationships among policies and interventions, school leadership and formal instructional communities' use of data, the kinds of instructional decisions that school leaders and instructional communities make, and gains in student achievement?

The conceptual framework and research questions call for a multi-method, multi-level, longitudinal research design. This study uses three kinds of data: qualitative data—observations, interviews, and documents—obtained from schools and from instructional communities within the schools and from education management providers and district staff; survey data collected from teachers; and district databases on student achievement and teacher characteristics. Our research design features:

- *School-based qualitative research and in-depth case study research* to develop a fine-grained analysis of the dynamic interactions among provider interventions, school capacity, school leadership, data use by instructional communities, and instructional decisions;
- *District and provider-level qualitative research*, including analysis of providers' policies and interventions and how they relate to provider theories of change; and
- *Survey research using hierarchical linear modeling* to 1) develop a broad descriptive picture of data use and instructional decision-making within schools and 2) determine the influence of instructional groups, professional community, and leadership on data use and instructional decision-making in schools;
- *Value-added analysis* that relates survey data on instructional decision-making, use of data, and organizational factors such as professional community and leadership to gains in student achievement at the school level.

In this paper we are discussing the school-based qualitative research and in-depth case study research.

Progress to Date in Our Data Collection and Analysis

Exploratory qualitative research for this project began in spring 2003 and continued through summer 2005. This early research helped us refine our conceptual framework, and informed the development of protocols for subsequent qualitative research and for our teacher survey. Research activities during this phase included:

- A series of 4 interviews with each of 20 principals in state identified low-performing elementary and middle schools in winter/spring 2003 just after these schools began receiving a “treatment” whether it was (a) assignment to an education management organization, (b) assignment to the district’s Office of Restructured Schools or (c) additional funding.
- In-depth interviews with staff in provider organizations and the district central office about the role of data use in their approaches to school improvement.

- Two days of observation at a district-wide professional development institute that focused on the use of SchoolNet, the review of school-level data related to *NCLB's Adequate Yearly Progress*, and leadership strategies for engaging staff in the review of data.
- In-depth interviews conducted in June 2004 with principals and teacher leaders in five elementary and middle schools that represented a range of education management organizations on their schools' use of data, and return visits to those schools during the opening days of school in September 2004 to observe faculty meetings in which staff examined student performance data and made revisions to their school improvement plans.
- A pilot teacher survey in 8 schools.

In Fall 05, with funding from the Spencer and William Penn Foundations, we began the full blown study. During the 2005-06 school year we:

- *Developed a purposive sample of 10 elementary schools* that had been identified as low-performing and targeted for intervention after the state assumed control. Originally, we had planned to include some middle schools in the sample, but the district's decision to convert almost all of its middle schools to K-8 schools meant that there were no appropriate stand-alone middle schools to examine. In constructing this sample, we aimed to include at least one school managed by each of the providers, including the district's Office of Restructured Schools. We also wanted to include schools that represented a range in terms of geographic location and, as much as possible, student demographics. We asked provider staff and district staff to nominate schools that were good examples of their approach

to curriculum, instruction and assessment, and the use of data to inform decision making in order to ensure that the schools in our sample would indeed represent the provider's approach and have identifiable and interesting strategies for data use.

- *Conducted two rounds of fieldwork in each school:* Two researchers spent about three days in each school (for a total of six researcher days). In all, we conducted more than 50 interviews of school staff and 19 observations of relevant school events, including leadership team meetings, grade group meetings, faculty meetings, and professional development sessions. Data collection was based on interview and observation protocols that included questions related to the history of the school's improvement/reform efforts, leadership, formal instructional communities, recent instructional changes, and school practices related to use of student performance data.
- *Conducted district and provider level interviews:* We also interviewed five central office staff who was responsible for developing and implementing assessment initiatives and at least one staff member from each of the provider organizations, including the Office of Restructured Schools.
- *Collected documents:* We have also begun collecting examples of "tools" that teachers and administrators use to collect, organize and understand data. We are asking interviewees about the development and use of such tools. This effort is providing us additional insight into routines around data use in schools.
- *Participated in the fielding of a district-wide survey of teachers and principals*
We were able to work with the Office of Accountability and Assessment to

include on the survey a number of questions on data use, particularly on the use of the district's "benchmark" assessments that are given every six weeks and are expected to be used to guide instruction. The items distributed well, and we were able to create some distinct scales.

- *Began preliminary data analysis* Researchers have typed up all fieldnotes and entered them into Atlas.ti qualitative data management software. In January 2006, the research team read all fieldnotes and identified early themes found in the school level and central office data sets. Over the summer, we refined the themes and developed seven broad codes. These codes created the foundation for a more detailed outline that guided researchers in the write up of case studies.
 - School's experience with school reform and historical context;
 - Supports for data use;
 - Roles and responsibilities associated with data use;
 - Purposes for data use;
 - Use of technology;
 - Challenges to data use;
 - School capacity.

During the 2006-07 school year, we developed a purposive sample of five schools.

Based on the analysis we conducted in summer 2006, we developed criteria for selecting five of the original ten schools for continued and more in-depth fieldwork. The criteria included how a school used data, the principal's role in data use, the professional community of a school, and the school's AYP status. As a result of team analysis, schools were placed along continua that represented each of these criteria. Five schools

were chosen that represented the range and diversity of data use practices as well as the range of other criteria mentioned. Schools with profiles that were especially different from the average district school (e.g. a school that was particularly small) were also taken out of the sample. We did not include a school from each provider in the sample given our finding that the majority of providers were using the district's system of managed instruction. We are now in the middle of our fieldwork in these five schools and the second teacher survey will be administered this May. Data analysis will continue in summer 2007.

GETTING GOOD INFORMATION ABOUT INSTRUCTIONAL CHANGES AND THEIR RELATIONSHIP TO EDUCATORS' REVIEW OF STUDENT PERFORMANCE DATA

As discussed earlier, this study seeks connections between the ways in which school staff engages with data and school and classroom practices, with the ultimate goal being to connect these instructional practices to student achievement. However, an ongoing challenge for this project has been finding such decisions about specific instructional practices and determining the role that student performance data have played in making them.

From the beginning, we have viewed semi-structured interviews with teachers and administrators as a critically important data source. Interviews with practitioners have the potential benefit of providing a window into their thinking, to learn how they are making sense of ideas, events, interactions with colleagues and students, and interpreting data. It is in this meaning making that practitioners combine their deep, inside knowledge about their school and their students and their expert knowledge about

curriculum, instruction, and assessment with student performance data to create the actionable knowledge that can translate into concrete changes in practice. Interviews can elicit stories of changes over time, in this case, changes in how practitioners engage with data. Both observations and surveys are very limited in their ability to enhance our understanding of meaning making.

Despite the potential power of interviewing to provide rich data on the complex processes under study, we have encountered significant challenges in capitalizing on its strength. We trace this difficulty back to our conceptual framework which provided an elaborated perspective on issues of school capacity and collective meaning making around data. Considerably less elaborated was our perspective on “instructional decision-making.”

A second problem revolved around the term “data use.” Our original conceptual framework identified several components of data use – accessing and organizing data, engaging with data, and linking knowledge generated from data to instructional decision making. However, protocols used during our exploratory fieldwork asked general questions that used “data use” ambiguously: for example, “Has there been any change in how teachers are using data to drive instruction?” Not surprisingly, informants gave general responses. In other instances, they focused solely on how teachers were learning to look at data, describing how, in one case, teachers graphed test results for their small learning communities during professional development time. We learned little from these initial interviews about the links between the examination of data and instructional practices. Since then, we have noticed in the literature that “data use” is often used to mean both analysis and discussions of data, as well as decisions that result from this

examination, when in fact these are two distinct issues. In our own work, the ambiguity of “data use” resulted in considerable inconsistencies in our early data.

By the time we began this study in earnest in 2005-06, we realized that we needed to explicitly separate questions about how data were examined from questions about the impact of such analyses on instructional practices. Specifically, we began asking more concrete questions about conversations related to data, and followed those up by asking, “Did any decisions or actions come out of the conversation?” This was a difficult question for people to answer, which, with the benefit of hindsight, is not surprising. School meetings often end inconclusively, with many brainstormed ideas and suggestions, but no specific decisions. Sometimes, people were able to identify important insights that emerged from data conversations, but not instructional decisions.

In spring 2006, we made a critical shift in our questioning strategy. We decided to ask school staff about *changes* that they had made in their classrooms and *changes* that the school had made more broadly. This question was much more accessible to our informants; they were more easily able to recount *changes*. We then asked people to map back from these changes and recall whether the examination of student performance data had played a role in those specific cases. The next challenge has been linking instructional changes to collaborative engagement with data, including the roles of professional communities and school leaders. We have addressed this by including questions about with whom the informant discussed the change and in what context. In addition, recent protocols have targeted the role of teacher leaders and sought to understand the ways in which teacher leaders work with teachers (individuals and groups) around data to help them implement changes resulting from data.

This shift in our approach has yielded some of the richest interview data yet, responses that get at decisions and decision making at the more “micro” level in which we are interested. For example, we’ve learned more from a number of teachers about the connections between examination of data and changes in student grouping practices (one of the more common instructional changes that we learned about). These groupings were more flexible (in one school, they were called “transient” groups) than traditional Red Bird/Blue Bird reading groups, and they were established around students who shared a common skill area weakness. One teacher described how she would group together six or seven students that all missed the same question on the Benchmark test:

I’ll pull those six or seven kids back during the review week, and say, okay, we all got this question wrong. Let’s read through it and see where they made the mistakes. And it’s funny, because they get so used to their one type of group that when you’re calling the seven kids back, and it’s usually from all across the board, from my highest reading levels to my lower, and they’re like, ‘well, how come we’re all getting pulled together?’ (Teacher, June 06)

A teacher at another school described how “the groups constantly changed” so that she could “target specific kids and their specific needs and group kids according to where they were lacking.” When she felt it was appropriate, she would also assign different homework to different students based on their needs. In other schools, teachers described how they had begun creating groups that cut across classrooms based on shared student weaknesses. And one principal described how his leadership team had reworked the instructional schedule and deployment of teacher leaders and volunteer tutors to accommodate the frequent use of these “transient” groups. These data point to interventions at several different levels – classroom, grade, school-wide – that have focused on differentiating instruction for students who are encountering difficulty with particular test items, skills, and concepts. Along with data collected from observations

and document analysis, these data have contributed to our preliminary understanding of the kinds of instructional “decisions” teachers and principals are making and how these are related to the examination of data.

Several categories of interventions have enriched our conceptualization of instructional decisions considerably and have provided a framework for continuing our investigation of data driven decision making. They include strategic interventions for individual students, interventions in classrooms, professional development for teachers, and programmatic interventions.

Strategic interventions for individual students. One common use of data from Benchmarks is to identify individual students for interventions including pull-out, extended day and tutoring programs. In many cases, the students identified are “bubble kids,” or those students who are on the cusp of scoring “proficient.” (The district has provided guidelines for what constitutes advanced, proficient, basic and below basic performance on the Benchmarks, the same categories of performance as on the Pennsylvania State System Assessment (PSSA).) One principal commented that:

When we go to identify the students, especially on my testing grades, my third and fifth, usually always, they'll have that (benchmark item analysis report) beside them. And our Math Leader prints that out for them so they'll that information when we try to identify the students. Because I'll ask them what are their scores on their benchmarks, and we use that quite a bit. (Principal, June 06)

Interventions in classrooms Among the most frequent uses of Benchmark data was re-teaching information or skills at the end of six-week cycles that were shown as weaknesses through item analysis at the classroom level (see above). In some cases, this was done through whole-class instruction. Different versions of re-teaching have also emerged. In some instances, teachers seem simply to place more time on certain content

or skills, while in other instances, teachers explore new instructional strategies for re-teaching an area of weakness. Regrouping as discussed above, was also a common classroom intervention.

Professional development for teachers Benchmarks, along with other data, can be a tool for identifying teacher needs (both in terms of pedagogy and content knowledge). Formal professional development, including training to strengthen teachers' skills in analyzing data as well as to improve the pedagogical skills and/or knowledge of teachers was a common response to teacher needs. Alongside these formal professional development sessions were various forms of less formal professional development and learning opportunities for teachers. For example, at several schools, curriculum coordinators or coaches provided "on-the-spot" professional development in areas identified as weaknesses through analysis of Benchmark and other data. One teacher described how a weakness identified through Benchmarks was addressed in her school:

We actually had a professional development about it, where [the principal] did a lesson to show us, and then we went to two other teachers' rooms and saw them do a lesson. And then pretty much that whole week that followed, [the principal] came around to see how we were using it, if we needed any help, what other support we needed to get this going and into play. (Teacher, June 06)

Programmatic interventions A few schools identified programmatic interventions, where changes beyond the individual classroom were the focus. In each of these schools, the interventions were directly tied to tests and performance on tests. For example, at two schools, teacher assignments were changed in order to move teachers who were perceived as stronger to tested grades. (At another school, there is a daily seminar period where every grade is divided into two sections, based on students' performance, to prepare for the PSSA's. While these "seminars" can nominally be

considered test prep, in fact the test prep is embedded into authentic instruction that involves teaching and re-teaching math and reading content. In another school, the school leader ordered new materials for social studies in order to help support efforts to improve language arts instruction and learning.

Asking specific questions about instructional *changes* enabled us to identify the four types of interventions that schools used to improve student learning, and to help us move beyond general and vague descriptions of data use to a more concrete understanding of the specific ways in which data played a role in shaping school and classroom practices. Based on data from a variety of sources, all schools used all four of these interventions to a greater or lesser degree to support and enhance student achievement.

ANALYZING ARTIFACTS TO REVEAL RECURRENT AND TAKEN FOR GRANTED DATA USE PRACTICES

Organizational learning theory suggests the importance of making the micro-practices of educators visible. It posits that informal communities of practice are critical to an organization's ability to innovate. In such communities of practice, workers are engaged in a *joint enterprise*; they are bound together in *mutual relationships* characterized by common norms for participation and performance; they employ a *shared repertoire of habits of mind* that involves them in thinking through problems of practice (Brown and Duguid, 2000; Wenger, 1998; Lave & Wenger, 1990).

Organizational learning theory also emphasizes the actual *practices* of communities, for example, pointing to the importance of *tools* used by teachers as they work together (Cobb, McClain, Lamberg, & Cean, 2003; Supovitz & Christman, 2003;

Wenger, 1998; Lave & Wenger, 1990). As Cobb and his colleagues argue, “the use of tools and artifacts is a relatively inconspicuous, recurrent, and taken-for-granted aspect of school life that is underdeveloped in the research literature both on teacher professional development (Marx, Blumenfeld, Karjick, and Soloway, 1998; Putnam and Borko, 2000) and on policy and educational leadership (Spillane, Halverson and Diamond, 1999)” (Cobb, McClain, Lamberg, and Dean, 2003). Examination of tools is important because these artifacts can tell us things that educators themselves may not be able to articulate. Analysis of tools also provides insights into the cultural features of the settings in which tools are used and how they structure different ways of thinking about and discussing data.

Before moving to a description of some of the artifacts we have collected, it is important to discuss SchoolNet, a central feature of the district’s Instructional Management System. SchoolNet provides users with online access to a variety of information including Benchmark data and other student performance data, student demographic data, student attendance data, and human resource data. SchoolNet has the ability to present the data in different formats, as well as to sort, organize, and integrate the data in different ways. Over the course of three years, schools were hardwired with equipment to gain access to SchoolNet and principals and teacher leaders were trained in its use. Teachers and administrators can access Benchmark data in a variety of formats and the district and the developers of School Net tout this feature. In many schools, however, teachers do not have the ability or time to download data themselves and our interviews and observations suggest that it is a more common practice for the principal and/or a teacher leader to download data reports and distribute them to classroom

teachers. Over time the school district and schools have developed a number of additional tools to support school based educators' use of Benchmark data.

SchoolNet reports and other tools designed by the district to analyze data are significant artifacts for our research. However, our focus on tools has introduced its own set of challenges for our team. Remembering to collect tools that are mentioned in an interview, labeling tools and referring to their labels consistently in field notes, cross-filing tools and fieldnotes, and figuring out how to handle tools in our qualitative data analysis software have been problematic.

We first encountered tools in our early interviews with principals and teachers. Our informants spontaneously brought them up as they described how they were using student performance data. Often the tools were not easily available to look at as we continued the interview and, as a result, our informants' descriptions of them were partial and often hard to follow. Therefore, in our May and June 2006 round of field research we decided to ask informants directly about tools and their use. We asked the following questions of the principal, school leaders and teachers: *What tool(s) do you use to guide your understanding of how students are learning? (Give an examples.) Where did the tool come from? Did you look at this by yourself? With someone else? With a group of people?* Information in response to these questions was more complete. However, we continued to have difficulty with simultaneously leading the interview and looking carefully at the specific aspects of a tool being discussed. We knew that our data about tools were incomplete and we were uncertain about how they fit into our larger process of analysis. However, our team discussions continued to highlight the potential value of the tools and accompanying interview questions, and so we began an analysis of the actual

artifacts related to Benchmark assessments with the goal of establishing a richer description of these tools' properties. This analysis enabled us to triangulate information about specific tools with information from interview and field observations. Here are some examples of the tools, and some of what we have been able to learn from them about micro practices around Benchmark data.

Classroom Benchmark Item Analysis Report

The Classroom Benchmark Item Analysis Report (Document 1) offers information on individual students in a classroom, providing a visual overview of class performance. It indicates the numbers of questions (out of twenty) that each student answered correctly and the numbers of students who responded correctly to each of the twenty questions. A check mark in a box indicates that the student answered the question correctly and a letter in the box indicates the student's incorrect response.

We have observed or heard about this tool being used in several ways. Most frequently, teachers identified individual students who needed additional support, especially "below basic" students and "basic" students who were on the cusp of becoming "proficient." In some cases, a teacher leader or the principal had already color-coded the data in a way that indicates where students fall – below basic, basic and proficient/advanced.

Well we have a chart, and what we did is we tracked... you know, the teachers put stars next to those kids that they're going to target, and we made sure that those kids had interventions, from Saturday school to extended day, to Read 180 which is another intervention. And then we followed their benchmark data. But those were the kids that the teachers were really gonna focus on, making sure that those kids become proficient, or move that ten percent out of the lower level so that we can make Safe Harbor next year.
(Teacher, June 06)

The Classroom Item Analysis Report is also used as a basis for grouping and re-grouping students to receive differentiated instruction and for identifying interventions and strategies for improving performance on specific standards or more typically and narrowly on types of test questions. Because these data are reported by classroom, it offers the opportunity for teachers to compare how their students are doing relative to students in other classrooms. However, this is only possible if teachers look at the data together as in the case below.

I can see how my whole class is doing. And they (members of grade group) can say, this one question, only four of your twenty kids got it right. So, I know that if only four kids got it right, that's something I need to go back and re-teach or, you know, get a fresh idea about how to give them that information. *(Teacher, May 06)*

District Benchmark Analysis Protocol

The District Benchmark Analysis Protocol (Document 2) provides a set of questions for teachers to answer after they have analyzed the Classroom Benchmark Item Analysis Report. Some of the questions ask teachers to summarize information in the report; others ask teachers to identify the implications of the report for grouping students, for instructional strategies, for instructional materials, and for ongoing assessment. District staff explained to us that teachers are required to complete the form, hand it into their principal, who then summarizes all of the forms and integrates the data into ongoing revisions of the School Improvement Plan. In addition, in a number of the schools we visited, teachers discussed the District Benchmark Analysis Protocol in their grade groups as described in the case below:

Although it [Benchmark Analysis Protocol) was really annoying to fill out, I think that the dialogue about the benchmark assessment really helped, like it was very time consuming the first couple of times I did it but then it actually

does have good data... And then we have to turn in intervention lesson plans using that. I think they should streamline that but I think that actually it's a helpful piece of data. (Teacher, June 06)

Benchmark Aggregate Item Analysis Report

The Benchmark Aggregate Item Analysis Report (Document 3) shows the percentage of students in a grade who responded to four possible answers on twenty questions on benchmark tests in Reading and Math for each cycle. The print out includes the percentage of students in a *grade* who chose the correct answer – which is identified – and the percentage of students who chose each of the three remaining answers. At the bottom of the form, a key is provided that links each item to the state standard it is assessing. District informants explained that this tool is intended to show patterns of strengths and weaknesses of student performance on the benchmark tests at the grade level and to pinpoint the *content* of particular standards that needs to be re-taught in the 6th week of the benchmark cycle. In our observations, we have frequently noticed that this report often spurs discussion in grade groups about why students picked certain incorrect responses. For example, fieldnotes from an observation of a fifth grade group meeting in February 2006 document the following.

Teacher #1 notes that most students had trouble telling what the ‘main point’ of the story was. The Principal suggests that maybe this is a language/vocabulary issue and they [students] aren’t understanding the question. The Principal asks, “What is the story mainly about?” Teacher #2 notes that some students had trouble with the word "pioneer." She says this is a difficult word because students are more used to it from social studies, and it wouldn’t be familiar in this context.

The most common reasons discussed by school personnel as to why students have difficulty with particular items have included: the ways in which questions are worded; vocabulary that students don’t know; and the testing of a concept in a way that was not

parallel to how the concept was taught in class. Discussions of students' conceptual understanding of content and why their reasoning went awry are considerably less frequent.

Benchmark Test Performance Report

The Benchmark Test Performance Report (Document 4) provides information that can be used to compare Benchmark results for grades 3-8 across Benchmark testing cycles. Its format does not make it easy to do this, however. The example here includes data from the October, November, January and March cycles (2005-06 school year) and indicates the percentages of students in each grade scoring 80-100% , 60-79%, 40-59% and below 39% on each Reading and Math test.

We observed this report being used in leadership team meeting, meetings of the whole faculty, and meetings of a school's staff with its School Assistance Team (SAT). Schools in Corrective Action under NCLB are assigned School Assistance Teams that are composed of members of the school leadership team, plus one or more individuals appointed by the district to oversee the SAT process. At a School Assistance Team meeting in March 2006, the SAT team case manager was reviewing the Benchmark Test Performance Report with the school leadership team at School X and noted, "Overall the 5th grade has gone from 54% of students scoring at proficient in Reading in January to 44% in March. What happened?" He then said that the school had reduced the percentage at the below basic level, but decreased the percentage in the proficient range. "So that's why the figures are the way they are. What happened? Did the 5th grade teachers say anything?" The Literacy Leader explained that she understood that it was not valid to draw comparisons across testing cycles because different content is tested in

different cycles of the benchmarks, and so they can't be equated. Additionally, the Benchmark tests frequently included material that had not been taught in the six-week cycle and that this misalignment was worse for some cycles than others.

The SAT Leader was not deterred. He continued: "In January there was a lot of concern that 5th grade scores in math were not moving in the right direction at the school, but between January and March the number of students scoring proficient and above has increased about 8%." For comparison, he said to look at changes in Math scores across that region of the city. He said that math scores in the Region had "stagnated. But something must have happened here. You achieved greater growth." He then thanks the leadership team for this accomplishment.

While the comparative analysis of Benchmark results from cycle to cycle is stressed by many in the district, in fact using the data in this way is inappropriate, because, as the School X Literacy Leader explains, the questions on each test are different and test items do not always reflect the content of the core curriculum during a cycle. It appears to us that the main use of this tool is to monitor grade level teachers since the tool is not providing information about student performance that could be used to guide instruction.

We found that a close look at these artifacts outside the context of our observational and interview data gave us a more concrete understanding of what the tools were telling teachers and asking of them. In fact, examining the tools *outside* of the school setting enabled us to go back to our field notes and interviews and more

confidently analyze and interpret the data than had been possible when we had looked at information about tool use from the interviews and/or fieldnotes by themselves.³

Our review of these tools and some preliminary analysis clearly show that there is not a one to one correspondence between what appears to be the intended purpose of the tool and how the tools are actually used in schools. In the following section, we look more closely at instructional communities in four schools to see how they make sense of information from one of the tools discussed above – the Classroom Item Analysis Report.

ANALYZING OBSERVATIONS OF INSTRUCTIONAL COMMUNITIES: THE VALUE OF HOLDING TOOLS AND SETTINGS CONSTANT

We had a clear goal of observing grade groups, leadership teams, and other settings where we might be able to document and analyze practices related to data use in instructional communities. And we have met this goal: each site visit to the schools has included interviews with school leaders or classroom teachers and observations of instructional communities.

However, once we began looking systematically across the fieldnotes from our observations, we faced an unanticipated analytic challenge: a great deal of variation existed in what we had collected in our research. Variation in itself was not a surprise;

³ In addition to School Net reports and district designed tools, we found several examples of tools that were designed by individual schools to further explore Benchmark data. These included:

- A Benchmark Analysis template that requires teachers to complete a form that asks them to discuss specific goals and interventions for each individual student in their class.
- A template that asks *students* to consider their Benchmark performance on a number of dimensions, and provide their teacher with feedback on how they learn best and how she or he could support their learning. (This tool shifts the way Benchmark data is discussed in schools to include student input.)
- A template that asks teachers to consider the implications of Benchmark data in combination with data from other sources for making instructional decisions. (In many schools, Benchmark data was used in combination with other data to plan interventions for individual students or groups of students.)

for example, we had initially expected to find variation across different intervention models. But a descriptive index of our fieldnotes showed us variations in features that might or might not be relevant to understanding how communities of practice develop within schools. For example, the organizational context of observations varied (e.g. grade groups met during the regular school day and also as part of half day or whole day professional development sessions). The grade groups observed might have been looking at different Benchmark reports (as discussed in the previous section). Other features such as the timing of a meeting within a school's trajectory of experience with data use, the timing of the meeting within the PSSA testing cycle, or the grades represented in our observations also varied.⁴ In order to minimize some of the variation, we decided to begin our analysis of fieldnotes from observations of instructional community meetings by using a smaller number of observations in which the organizational context of the observations and the templates for looking at data were constant. Examining the discussion of Benchmark data in grade groups was our first attempt at analyzing data in this way. Grade groups are intended to provide opportunities for classroom teachers and principals or other school-based leaders to discuss grade-appropriate curriculum, instructional strategies, and assessment for students within one grade or a span of 2-3 grades. Because we have held the organizational context constant and held the data and data formats constant, it becomes possible to see some of the emerging contours of the micro-processes – particularly the actions of school leaders – that shape differences in how and whether instructional communities learn to learn with data.

⁴ To a large extent, the level of variation in our observations was a result of our “opportunistic” sampling strategy. We had the goal of observing professional development, leadership meetings, and grade groups at each school, but the actual observations were primarily determined by what meetings were already scheduled during each round of our site visits, what meetings we had access to, and what tasks needed to be accomplished by the schools.

School leaders play a critical role, as they are in a position to encourage and support instructional communities to engage with data to transform practice. Recently, the idea of principal as “learning leader” has emphasized the role of principals in creating a climate in which adult learning is central to school improvement (DuFour, 2002; Spiri, 2001; Scribner, Madrone, & Hager, 2000; Elmore, 2000). In looking at data use, however, we take a broader view of leadership. Spillane, Halverson, and Diamond (2000) have argued for a conception of leadership that is *distributed* across the roles of principal, teachers, and parents/community members. In addition to principals, teachers can assume important leadership functions relative to data use.

Overview of Grade Group Observations Used in This Analysis

Our grade group observations took place at four different schools (Schools A, B, C, and D) during the 2004-2005 school year or the 2005-2006 school year. Each of these observations took place during a regularly scheduled grade group meeting and was part of the regular school day. At each of the meetings, participants were working with the district’s Classroom Item Analysis report. Attendance at each meeting consisted of the school’s principal, at least one other school leader, and between two to four classroom teachers. It is likely that our request to observe the use of data shaped who attended these meetings and the conversations that occurred at these particular grade groups. However, participants’ levels of knowledge and familiarity with the materials and processes suggest that the activities occurred with some regularity in the schools.

Tools and Processes: In three of the four observations, principals handed teachers print-outs containing their class’s Benchmark scores in literacy and mathematics. In each case, the tool used was the student level item-analysis form

discussed above. In one grade group, the classroom teachers had already reviewed their students' Benchmarks. In general, the principals asked teachers the same questions: "How many students are proficient or advanced? How many are close to "proficient" or "advanced"? What are the questions that gave the students the most problems?" In some grade groups principals played particularly prominent roles, but in every grade group, classroom teachers and teacher leaders also made recommendations about strategies for improvement. Most of the staff observed appeared familiar with the Benchmark print-outs and with AYP goals.

Framing Benchmarks: Different Responses to the Same Set of Tools

These observations indicate that Benchmark data are used to identify short term strategies related to improving test scores in all the schools. In some schools, Benchmark data can also be a jumping off point for discussions of deeper changes in instructional practice. However, use of Benchmarks for instructional improvement is limited when schools are facing pressure from high stakes testing and when school leaders do not have adequate understanding of the intersections between Benchmarks and the curricula used in their schools.

All observations involved conversations about specific approaches to improving test outcomes. For example, teachers were advised to give students practice in test-taking strategies such as using a process of elimination in multiple choice problems, reading problems to students in mathematics tests, and using calculators, where allowed.

However, there is also a pattern of differences in these observations suggesting that teachers in two schools (A and B) participated in a grade group meeting that structured conversations about Benchmarks as a conversation about curriculum and

instruction, while teachers in the two other schools (C and D) participated in a grade group meeting that structured conversations about Benchmarks as a conversation about test-taking. One way of understanding these different patterns is to look closely at how the leaders at different grade group meetings approached similar problems identified in the Benchmarks:

Trick Questions: “Evil Test-makers” or an Opportunity for Inquiry and Independence?

A common concern among educators is that standardized tests, including the Benchmarks, do not capture students’ true skill level or content knowledge. We heard educators express concerns about the assessment including: the inclusion of multi-step problems of a type unfamiliar to students; questions in formats that did not parallel the form in which they were taught in class; and questions that were culturally biased. An issue that came up repeatedly in two of the meetings (at Schools C and D) was that the Benchmark questions were misleading and therefore not fair. The principal or teacher leader in these schools avoided the issues related to fairness and, instead, responded to teachers’ concerns with suggestions about test preparation strategies that might motivate students to persevere in the face of challenging questions.

For example, the 4th grade teacher at School C explained that most of her students missed a question about the length of a paper clip because students assumed that they had to start measuring at 0, even though the paper clip was placed at the 2 cm on a ruler. What they needed to do to obtain the correct answer was to subtract 2 cms. The math teacher leader responded that it was the “evil test makers at work. Nobody ever starts measuring something from 2 cm.”

The principal responded that “The re-teaching opportunity can be powerful, especially if it’s done right after students take the test and it is fresh in their mind. Sometimes it’s two or three steps that you need to get to in order to get the right answer.”

At School D, the 5th grade teachers said that their students were having a lot of problems with fractions on the Benchmark tests.

One teacher explained that one of the hard things for her students was reducing improper fractions. She described how she reviewed such a problem with her students and helped them understand what they needed to do to apply skill that they had already learned to a new situation. The teacher explained, “A lot of light bulbs went off [when students saw how to draw on what they already knew]. No one in the meeting addressed conceptual issues related to improper fractions, but the principal who was captivated by the image of light bulbs going off, suggested that teachers make posters of light bulbs for their classrooms to motivate students during the Benchmarks and other tests. “Hang up a light bulb, put on your thinking caps and say “I can do it.”

In both of the above examples, instructional issues are being skirted. The principals offer helpful tips to teachers, but the substantive instructional questions remain unaddressed.

In contrast, the principals and teacher leaders at two schools (A and B) encouraged their teachers to take different approaches to the gap between student knowledge and student performance. The principal at School B took every opportunity she could to help teachers understand the connections between the Benchmark items and the standards they were designed to assess. Instead of test preparation strategies to help students deal with tricky questions, this principal directed her 4th grade teachers to think about the relationship between the Benchmark assessments and the standards. “Look at questions that test the same standard. Are they written the same way or a different way? Is one harder than the other?” Later in the meeting, the math teacher leader gave an

example of how to do this by showing how expanded notation was used in two different questions on the most recent Benchmarks.

At School A, the math teacher leader also had ideas about how teachers could change their instruction to help students deal with tests. At the grade group we observed he told the 6th, 7th, and 8th grade teachers, “One of the reasons that people say the kids know the material, but don’t test well is that the conditions are so different...During instructional periods, teachers need to let the kids do more on their own, so it’s is more like a testing situation where they have to interpret the instructions.” He suggested that during a math period, the teachers should start with the objective, give out the supplies needed (ruler, scissor, etc.) and then have the students follow the directions for an activity without telling them what to do. He concluded, “Our students need to learn to be more independent. Then you can review and reflect with the small groups.”

In order to understand how educators use data, it’s essential to identify and analyze the tools and organizational structures that create the contexts for data use within schools. In this phase of our analysis, we held tools and structures constant and were able to start seeing patterns in how instructional communities approach Benchmark test results. The issues raised were similar across schools - especially the gap between teachers’ perceptions of student knowledge and their performance on particular test items. However, there were differences in how instructional leaders helped classroom teachers think about the gaps between student knowledge and student performance on the Benchmarks. In two settings, the emphasis was primarily on boosting students’ confidence and persistence in answering hard problems on standardized tests. In the

other settings, the analysis of Benchmarks also became a platform for talking about curriculum and instruction.

Leadership and Instructional Community

One way of articulating this contrasting approach to the use and value of Benchmark is to consider “Deepening Instruction” as one of several dimensions that needs to be looked at to identify how different instructional communities work with data.⁵ We are exploring the hypothesis that instructional communities that are high on this dimension are characterized by principals and teacher leaders who foster a number of practices which might include: referencing curriculum content, chunking, problematizing the Benchmark system, and asking reflective questions.

Referencing curriculum content: By this we mean school leaders demonstrate an interest and an understanding of the specific content standards and curricular issues relevant to student assessment data being examined. Mintz and colleagues (2005), suggest that data should be used as “a springboard for focused conversations about academic content that the faculty believes is important for students to know.” (p. 94). For example, if teachers comment that students cannot answer a question because it wasn’t covered in the curriculum, it is a “good opportunity to investigate how content areas are addressed in the curriculum” (Mintz and colleagues et al 2005:94)

Chunking: If Benchmark items (or items in other assessments) are looked at one-by-one, it is hard to identify meaningful connections to content. Conversations about single benchmark items are often limited to comments like, “my students don’t remember how to _____.” On the other hand, it is possible to look across questions to address

⁵ We would like to acknowledge that Young (2006) presents a useful analysis of dimensions that are necessary for successful use of data within instructional communities. We are interested in understanding the intersections between our emerging analysis and the framework that she uses.

meaningful content issues. A teacher in one of the grade groups discussed above did exactly this when she commented that her students had difficulty turning a fraction into a decimal and that they didn't understand the relationship between the two. In looking across test items, this teacher flagged one of the "big ideas" in upper elementary mathematics which had the potential to open up a discussion of what is, or isn't in the curriculum, for addressing this important idea. Likewise, chunking very specific instructional changes into a larger approach offers a better chance of coherent improvement.

Problematizing the Benchmark System: Teachers have well-founded concerns about the quality of data, especially from standardized tests, and they question its usefulness to them (Cromey, 2000). Teachers repeatedly discussed problems related to the alignment between the Benchmarks, the curriculum, and the PSSA. It is obviously desirable to have the Benchmarks, the curriculum, and the PSSA aligned as closely as possible, and the district has put substantial effort into making corrections. However, as pointed out by Wenger (1998) in his study of communities of practice, a written policy or system for organizing information is a "reification" of ideas that will necessarily have to be adapted and modified in practice. It is important that school leaders have sufficient knowledge about the Benchmarks, the curriculum, and the PSSA so that they can help teachers stay focused on what useful information they can garner from the Benchmarks instead of getting sidetracked and frustrated with the alignment problems that inevitably show up.

Asking Reflective Questions: Faculty may view data use strictly as a means of compliance with disagreeable external mandates, rather than as a source for problem-

solving. Nevertheless, principals and other members of leadership teams are called upon to engage their school communities in thoughtful analyses of student assessment data (Hoachlander, Alt, & Beltranena, 2001; National Association of Elementary School Principals, 2001; Council of Chief State School Officers, 1996). In these kinds of inquiries, teachers are asked to engage with questions about the meaning and value of data that do not have simple or predetermined answers.

In our initial analysis of instructional communities that are high on the dimension of using data to deepen instruction, principals and teacher leaders encouraged their peers to understand and think about curriculum and content in relation to the Benchmarks. In contrast to leaders in instructional communities that focus on short-term strategies, leaders in communities that deepen instruction are less likely to pay concerted attention to particular Benchmark questions and more likely to direct attention to standards or larger content issues. In addition, they have enough knowledge of the Benchmark system, the curriculum, and the standardized tests that they can tell teachers what Benchmarks to ignore and which ones to pay more attention to as they proactively point out inconsistencies and misalignments. In contrast, in instructional communities where the focus is on short term strategies, classroom teachers are more likely to raise concerns and criticisms about the lack of alignment between the Benchmarks and the curriculum. Finally, in instructional communities that use discussions of Benchmarks as an opportunity to deepen instruction, the principals and teacher leaders are more likely to model a reflective process with their teachers.

Next Steps for Analysis

This phase of analysis, which is based on a very small data set, raises several questions for the team's ongoing work including, what kinds of micro-practices would become visible if we focused on observations of instructional communities using different types of data and formats for organizing data (e.g. rubrics for assessing student work).

CONCLUSION

While data-driven decision-making has become an increasing focus at the levels of state and district policies, there is still much to be learned about how those in schools actually engage with data and use it as the basis for making decisions about what happens in schools and classrooms. In this paper, we have discussed methodological strategies that we developed for shedding light on the micro-processes in schools that often remain invisible, even to those who are looking for them. Our research goals have been to learn about: (a) the kinds of instructional changes that teachers and principals make based on student performance data; (b) how they use district-developed tools to help them examine and make sense of the data; and, (c) the kinds of conversations that they are having about the data in instructional communities. Our account begins to shed light on the complexity and variability in how those in schools both engage with data and use it as the basis for altering their practices. While perhaps creating a “messier” picture of the role of data, using multiple methodological strategies has significant promise for developing much richer understandings not only of the role of data itself, but also of the role that instructional communities and leaders play in helping their schools become learning organizations.

In a few weeks we will be meeting with district central office administrators responsible for professional development, curriculum and assessment to share our preliminary findings with them. A central theme that we intend to develop is the importance of helping school leaders (teachers and principals) learn the skills of linking the discussions of Benchmarks to curriculum content, offering opportunities for chunking, reflective questioning, and problematizing Benchmarks. Our next step is to link what we have been learning about instructional decisions to what we have been learning about grade group meetings and to decide what other kinds of data we may need to develop a rich understanding of the role of these instructional communities in data use and instructional changes.

REFERENCES

Anfara, V., Brown, K., and Mangione, T. (2002) Qualitative analysis on stage: Making the research process more public. *Educational Researcher* October, 2002.

Argyris, C. & Schon, D. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley.

Brown, J. S. & Duguid, P. (2000). Organizational learning and communities of practice: Toward a unified vision of working, learning and innovation. In E. L. Lesser, M. Fontaine, & J. A. Slusher (Eds.) *Knowledge and Communities*: 99-121. Boston: Butterworth Heinemann.

Bryk, A.S., Easton, J. Q., Kerbow, D., Rollow, S. G., & Sebring, P. B. (1998). *A view from the elementary schools: The state of the reform in Chicago*. Chicago: Consortium on Chicago School Research.

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.

Cobb, P., McClain, K., Lamberg, T. S., & Cean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher*, 32(6), 13-24.

Corcoran, T. & Goertz, M. (1995). Instructional capacity and high performance. *Educational Researcher*, 24(9), 27-31.

Cromey, A. (2000, November). Using student assessment data: What can we learn from schools? NCREL Policy Issues, Issue 6. North Central Regional Educational Lab. Retrieved 5/24/04 from the world wide web: <http://www.ncrel.org/policy/pubs/pdfs/pivol6.pdf>.

Knapp, M. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation and professional learning. *Review of Educational Research*, 67, 227-266.

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Leithwood, K., Aitken, R., & Jantzi, D. (2001). *Making schools smarter: A system for monitoring school and district progress*. Thousand Oaks, CA: Corwin Press.

Little, J. W. (1999). Teachers' professional development in the context of high school reform: Findings from a three-year study of restructuring schools. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec.

Louis, K. S., Kruse, S., & Associates. (1995). *Professionalism and community: Perspectives on reforming urban schools*. Thousand Oaks, CA: Corwin Press.

Marx, R., Blumenfeld, P., Krajcik, J., & Soloway, E. (1998). New technologies for teacher professional development. *Teaching and Teacher Education, 14*, 33-52.

Massell, D. (2000). The district role in building capacity: Four strategies. CPRE Research Brief Series RB-32. Philadelphia: Consortium for Policy Research in Education.

Mason, S. A. (2003). Learning from data: The role of professional communities. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

McLaughlin, M.W. & Talbert, J.E. (2002). Bay Area School Reform Collaborative: Summary Report (Phase 1, 1996-2001). Stanford University: Center for Research on the Context of Teaching. Available online at:
<http://www.basrc.org/Pubs&Docs/2002CRCSUMMARY.PDF>

Mintz, E., Fiarman, S.E., and Buffet, T. (2006). Digging into data. In K.P. Boudett, E.A. City, and R.J. Murnane (Eds.), *Data wise: A step-by-step guide to using assessment to results to improve teaching and learning* (pp.81-96). Cambridge, MA: Harvard Education Press.

Newmann, F. M., King, B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review, 67*, 41-74.

Newmann, F. M., King, M. B., & Youngs, P. (2000). Professional development that addresses school capacity: Lessons from an urban elementary school. *American Journal of Education, 108*(4), 259-299.

O'Day, J., Goertz, M. E., & Floden, R. E. (1995). Building capacity for education reform. New Brunswick, NJ: Consortium for Policy Research in Education. ERIC Document Reproduction Service: ED 390 118.

Putnam, R., and Borko, H. What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4-15.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.

Spillane, J. P., Halverson, R., & Diamond, J. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, 30(3), 23-28.

Spillane, J. & Thompson, C. (1997). Reconstruction conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, 19(1).

Supovitz, J. & Christman, J. B. (2003) Developing communities of instructional practice: Lessons from Cincinnati and Philadelphia. CPRE Policy Briefs, Series RB-39. Philadelphia: Consortium for Policy Research in Education.

Wagner, T. (1998). Change as collaborative inquiry: A 'constructivist' methodology for reinventing schools. *Phi Delta Kappan*, 80(7), 378-83.

Watson, J. & Mason, S. (2003). Understanding schools' capacity to use data. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, England: Cambridge University Press.

Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education*, 112, 521-548.

APPENDIX

Document 1: Benchmark Item Analysis by Student

Questions	1	2	3	4	5	6	7	8	9
Standard	2.2.8.B	2.2.8.B	2.2.8.B	2.2.8.B	2.2.8.B	2.9.8.B	2.9.8.F	2.9.8.E	2.9.8.G
Possible Points	1	1	1	1	1	1	1	1	1
Correct Response	C	C	C	B	B	A	D	A	B
BOUNIGHT, KANSHA	✓	D	✓	D	✓	C	✓	✓	
BURGERON, ALVA	✓	✓	✓	✓	✓	C	B	✓	✓
CHRISTLEY, ANTWAN	✓	A	✓	✓	✓	B	B	✓	C
DAENAIL, OLLON									
DELANGEL, SCOTT	✓	✓	✓	A	✓	✓	✓	✓	✓
DOEDKE, LELAH	✓	✓	✓	✓	✓	✓	B	✓	✓
EASTMAN, ISIAH	✓	✓	✓	✓	✓	✓	✓	✓	✓
ESSEX, WILLIS	✓	✓	✓	✓	✓	✓	B	✓	C
FALLEN, ZITA	✓	✓	✓	✓	✓	C	✓	✓	✓
GOLDSON, BACANN	✓	✓	✓	✓	✓	D	✓	✓	D
HECRICK, JAMEL	✓	✓	✓	✓	✓	D	✓	✓	✓

Document 2: Benchmark Data Analysis Protocol for Teachers of Grades 3-8

**School District of Philadelphia
Benchmark Data Analysis Protocol
for Teachers of Grades 3-8**



Benchmark Administration Date: _____

School _____

Teacher's Name _____

Grade/Section _____

Benchmark Analysis Using the IMS (*SchoolNet*): (Refer to guide to access reports)

1. Please examine the reading and/or math **Item Analysis data** for your class or section (**Align Report**).
2. **Look closely** for patterns, outliers, weaknesses, and strengths.
3. Attach the reading and mathematics **Item Analysis** for each class to this plan.

READING	MATHEMATICS
Using the item analysis report, identify the weakest skills/concepts for your class for this benchmark period.	Using the item analysis report, identify the weakest skills/concepts for your class for this benchmark period.
How will you group or regroup students based on the information in the necessary item analysis and optional standards mastery reports? (Think about the strongest data and how those concepts were taught.)	How will you group or regroup students based on the information in the necessary item analysis and optional standards mastery reports? (Think about the strongest data and how those concepts were taught.)
What changes in teaching strategies (and resources) are indicated by your analysis of benchmark results?	What changes in teaching strategies (and resources) are indicated by your analysis of benchmark results?
How will you test for mastery?	How will you test for mastery?

School District of Philadelphia Benchmark Data Analysis Protocol for Teachers of Grades 3-8



Teacher's Reflection

1. I have discussed my classes' results and my lesson planning for week 6 with...
2. In order to effectively differentiate instruction (remediate and enrich) during week 6, I need to . . .
3. Based on patterns in my classes' results, I might need some professional development or support in . . .
4. As I think about giving students a better understanding and more ownership for their learning based on benchmark assessments, I ...
5. I receive support from....

Document 3: Benchmark Aggregate Item Analysis

Test Details						
Test Name:	NOV2003 CCAT Grade 8 Math					
Status:	Active					
Description:						
Test ID:	25					
Subject:	Mathematics					
Grade Level:	8					
Number of Questions:	14					
Test Booklet:						
Grading Rubric:						
Administration Start Date:	Wednesday, 11/19/2003					
Administration End Date:	Wednesday, 11/19/2003					
Score Due Date:	Wednesday, 11/19/2003					
Number of Students who took test:	222					

Question	Answers				No Answer Selected	Percent Correct	# Correct / # Attempted	P Value **
1	A 5%	B 11%	C 37%	D 46%	0%	37%	83 / 222	0.374
2	A 2%	B 97%	C 1%	D 0%	0%	97%	216 / 222	0.973
3	A 2%	B 7%	C 3%	D 89%	0%	89%	197 / 222	0.887
4	A 2%	B 2%	C 0%	D 95%	0%	95%	212 / 222	0.955
5	A 13%	B 29%	C 26%	D 32%	0%	29%	64 / 222	0.288
6	A 91%	B 5%	C 2%	D 2%	0%	91%	202 / 222	0.910
7	A 14%	B 0%	C 1%	D 84%	0%	84%	187 / 222	0.842
8	A 9%	B 22%	C 68%	D 1%	0%	68%	151 / 222	0.680
9	A 6%	B 39%	C 43%	D 12%	0%	43%	96 / 222	0.432
10	A 0%	B 1%	C 0%	D 98%	0%	98%	217 / 222	0.977
11	A 4%	B 4%	C 5%	D 87%	0%	87%	193 / 222	0.869
12	A 76%	B 5%	C 4%	D 16%	0%	76%	169 / 222	0.761
13	A 1%	B 89%	C 8%	D 1%	1%	89%	198 / 222	0.892
14	A 41%	B 23%	C 21%	D 15%	1%	41%	91 / 222	0.410
Summary						73%	162.57 / 222.00	0.73

Key: Correct Answer Open-ended question

**P value is an index of an item's difficulty, evaluated by dividing the total points achieved by the total points possible for an individual item.