2

Taking a Hard Look at Hard Data

Most of us have structured our daily lives to remind ourselves of the things we should do. We hang a motivating photograph on the refrigerator door to prompt us to make healthy food choices. We keep a packed gym bag in our car to cue us to work out at the end of the day. We place hand sanitizer dispensers in strategic locations so we remember to use it at warranted times. These subtle messages are called nudges, and they can be incorporated into the environment to nudge ourselves and others into making good choices (Thaler & Sunstein, 2008). An entire field of study called behavioral economics examines how environments can be engineered to guide people to make safe, healthy, and environmentally conscious decisions. For example, rumble strips on the road encourage you to slow down and signal when you are drifting out of your lane. Painted footsteps on the sidewalk lead you to a trashcan to throw away your litter. Lunch rooms are structured so that you must walk past the salad bar before selecting less healthy food items.

At the high school where three of us—Doug, Nancy, and Diane—work, we use hard data as a nudge. Although many schools have developed rooms for displaying testing results and attendance figures, access is usually limited to teachers and administrators. We realized that our data analysis could be used more effectively if we increased the number of people who view it. Therefore, four years of test score and high school exit exam data are displayed on the cafeteria walls above the food service areas so that students can view these results. The data are charted so that students can follow their graduating

class's progress from freshman to senior year. Each mathematics teacher conducts occasional classes in the cafeteria to use the data charts in novel ways to teach principles of algebra and statistics to her students. At the top of the main staircase, daily attendance data are posted for each grade level, as well as the monthly averages. When 11th grade's attendance dipped midyear, the teachers met with students to discuss the issue and made plans to increase support for stressed students.

In the context of education, data are any information that we can glean about teaching and learning. To improve schools, educators must consistently use data to determine where they are and what needs to be done next. For data to be effective, they must be widely available and be used to affect instructional and policy changes. As we stated in the previous chapter, we need to view data analysis at its point of use: the teaching and learning that occurs in the classroom, in front of students.

In this chapter, we will focus our attention on hard data—how the data can be quantified, disaggregated, and most important, leveraged to nudge the changes needed for continuous instructional improvement. It's all about becoming "choice architect[s]… [who have] the responsibility for organizing the context in which people make decisions" (Thaler & Sunstein, 2008, p. 3).

Hard Data Defined

Hard data are quantifiable; hard data can be described with a given degree of specificity and tangibility. In addition, hard data are relatively stable in that they aren't changed significantly by the method in which they are collected. The most common form of hard data generated for schools is derived from student assessment results. Other readily available hard data include attendance, suspension rates, and teacher credentialing. Hard data are gathered by some form of counting, and most hard data come from official or organizational sources. A difference between hard data and soft data is in collection and analysis. Hard data are usually converted into numbers and displayed as graphs. Soft data are displayed using qualitative methods such as anecdotes, testimonials, and quotes that reflect patterns and trends. Both are actionable and should be viewed as complementary. The soft data of an individual student's story breathe life into the numbers. Likewise, the hard data of attendance affirm the difference between an isolated incident and a noteworthy trend.

Hard data are reported using descriptive statistics that answer three questions: *Who? What? When?* The following statements are examples of how hard data are reported.

- Thirty-seven percent of the students in 4th grade scored proficient or advanced on the algebra and functions strand of the state content standards exam in the spring.
- Seventy-eight percent of 5th grade boys scored 340 points or higher on the math section of the second district benchmark assessment.
- Fifty-two percent of 2nd graders scored proficient in written conventions on the grade level writing rubric in May.
- Fourteen percent of 9th grade students missed five or more days of school during the first quarter of the school year.

These data can be collected using teacher-based, criterion-based, or norm-referenced assessments, and can be formative or summative in nature. However, if the reports of data are limited to snapshots of achievement, they fail to illuminate a pathway of improvement. For this reason, we advocate that comparative statistical analyses are also used to determine trends. These answer a fourth question: *Compared with what?*

- Thirty-seven percent of the students in 4th grade scored proficient or advanced on the algebra and functions strand of the state content standards exam in the spring, which is a 3 percent increase for this cohort over last year's performance in 3rd grade.
- Seventy-eight percent of 5th grade boys scored 340 points or higher on the math section of the second district benchmark assessment. On the first benchmark, 66 percent of these students scored in this range.
- Fifty-two percent of 2nd graders scored proficient in written conventions on the grade level writing rubric in May, compared with 49 percent in March.
- Fourteen percent of 9th grade students missed five or more days of school during the first quarter of the school year. Last year, 19 percent of freshmen missed five or more days in the same quarter.

The comparative use of hard data lets us know what interventions are working and to what extent. It can also alert us to possible problems that are slowly taking root, or serve as an early warning that a previously effective intervention may need to be reevaluated. However, numbers alone don't tell the full story, and they are easily misused if assessment literacy is lacking.

Being Assessment Literate in Hard Data

We often discuss issues of literacy for our students, but we rarely apply the same conditions to our own learning. We don't mean the educational definition of literacy, which relates to reading, writing, speaking, listening, and viewing, but rather its root: *literate*. Consider the synonyms of the term: *educated, scholarly, knowledgeable*. Similarly, assessment literacy is being educated about the instruments, scholarly in analysis, and knowledgeable in interpretation. Assessment literacy "comprises two skills: first is the ability to gather dependable and quality information about student achievement; second is the ability to use that information effectively to maximize student achievement" (Stiggins, 2001, p. 20). An essential part of being assessment literate is in understanding how, and under what circumstances, data are collected. Given that the most common type of hard data used for school improvement comes from state and local assessments, let's take a closer look at norm-referenced and criterion-referenced tests.

Norm- and Criterion-Referenced Tests

Most commercially prepared tests fall into one of two categories: norm-referenced and criterion-referenced. Normed tests are used to compare the performance of an individual student with others across a larger population, usually a national one. Tests of intelligence such as the Weschler Intelligence Test for Children yield an intelligence quotient for test takers. Some achievement tests, including the Iowa Test of Basic Skills or the Stanford Achievement Test, provide similar rank ordering, and generate the familiar bell curve. Results are typically reported as percentiles, with further analyses such as stanines, which scales the test results into nine segments. Because the purpose of an IQ test is to compare participants using a bell curve, the instrument is designed so that the majority of students score near the middle. Although norm-referenced testing has some applications to education (like intelligence testing), it is of limited value when measuring the learning of students. For this reason, most standards-based tests are criterion referenced.

Criterion-referenced tests are assessments created to determine student mastery of content. In other words, they tell us to what degree a student has learned something. A teacher-created assessment on stoichiometry for her chemistry students is a criterion-referenced test, because it gives the teacher and her students information about how much they learned during this unit of instruction. Likewise, the district-created science benchmark and the state standards assessment for chemistry are criterion based. In each case, a student's level of proficiency is reported along cut points, or predetermined scores

that convert quantitative measures into performance bands. The cut points for a unit test on stoichiometry are reported as grades, whereas the science benchmark is likely to have cut points tied to identified bands of mastery or lack thereof such as *intensive*, *strategic*, *basic*, *proficient*, or *advanced*. Students in the intensive and strategic bands are said to be in the most need of intervention. Those in the proficient and advanced bands are deemed to have mastered the tested skills. The state chemistry standards test provides students, as well as their families and schools, with information about how the individual performed and often gives similar cut point information.

Using Criterion-Referenced Tests for School Improvement

At the *federal level*, hard data are collected annually about elementary and middle school student performance in the areas of reading, writing, and math; and about high school student performance in the areas of reading, math, science, and world history via the National Assessment of Educational Progress (NAEP). These assessments are administered to 4th, 8th, and 12th grade students in schools across the country and are proctored by NAEP personnel. Student results from these criterion-referenced tests are disaggregated to compare academic progress from state to state. Often called "the nation's report card," NAEP scores are not reported at the individual student level. Instead, these results are used to formulate policy and determine funding.

At the *state level*, hard data are collected annually from every school in every district beginning at 2nd or 3rd grade. Schools administer state standards-based assessments within a timeframe tied to the number of days students have been in school; that is, students in every school take the test after being in school for the same number of days. Students are assessed in the areas of language arts, math, science, and social science. Districts and schools across the state compare student results broadly by subject area and also drill down to examine the results by content cluster and strand. Unlike NAEP, the results are tied to individual students, making it possible for schools to identify those in need of intervention the following school year. A compelling purpose for developing the Common Core State Standards was so that results could be compared across states as well as within, because portions of the testing instruments will be identical.

At the *district level*, hard data are collected several times annually through criterion-referenced benchmark assessments. These tests are usually administered to every student in specific grades at the end of each quarter or trimester. Information from the district tests are used at the district level to gauge progress, compare schools, and identify schools

that might need assistance. Data that are generated include the number and percent of students scoring at or above a given cut score in each content area. The data are further analyzed to determine students' progress toward outcome standards. These district tests are administered and analyzed to provide timely interventions to students to bolster learning.

At the *school level*, similar common formative assessments are collected on a more frequent basis to improve the precision of teaching and intervention. These criterion-referenced assessments provide information about progress and are further used strategically to ensure that student performance information informs future decisions.

Looking Deeper

Of course, schools are not one-dimensional, and achievement data do not fully portray the nuances of school life. Other readily available hard data are used to describe the students in a school. Demographic data are reported across many attributes, including race, ethnicity, gender, disability, and socioeconomic status. In addition, home language surveys are collected and analyzed to identify English learners. How these demographic data are reported and disaggregated is regulated according to federal and state guidelines. The purpose of these demographic data should go beyond mere description and be further analyzed to allow schools to look deeper for students who might otherwise be lost in the data deluge.

With the enactment of No Child Left Behind (NCLB), schools were required to pay attention to the achievement of all students. Schools and districts now report the performance of students by subgroups: black or African American, American Indian or Alaska Native, Asian, Filipino, Hispanic or Latino, Native Hawaiian or Pacific Islander, white, two or more races, socioeconomically disadvantaged, English learners, and students with disabilities. By disaggregating state results by subgroups, the federal government forced states to examine the achievement of all groups of students.

But you may recall an earlier time when students with identified disabilities were tucked into the corners of high-performing schools. Their numbers alone were usually not enough to affect the overall achievement of an entire school, as they typically composed only 8 to 15 percent of the student body. It was simply a byproduct of the mathematics of achievement. A school with a high overall level of achievement wasn't going to experience sufficiently depressed scores to warrant attention. Students with disabilities were always there, hidden in plain sight. But the accountability measures required by

NCLB changed that. Although the NCLB accountability measures remain controversial, most educators agree that the measures illuminated an overlooked portion of our school population. More than a decade later, it is hard to imagine achievement discussions that don't include students with special needs.

Attendance data provide another lens for examining the pulse of the school. Most educators are familiar with applying attendance data at the individual level to alert us to those students whose chronic absenteeism places them at risk. But absence rates can also be used to locate patterns among students. For example, by examining school health records, demographic information, and absence rates, a school may find a significant relationship—perhaps that habitually absent black students have poorly controlled asthma. Asthma is the most common chronic illness of childhood, and poor, urban black boys have the condition at a rate 45 percent higher than the rate for peers (Basch, 2011). What initially may have been perceived as a motivation issue may be a health one, and the solution is to coordinate family, school, and community supports. In this case, the health issue warrants having asthma action plans for students, reducing environmental triggers at school that can set off an attack, and creating a case management system for students whose asthma is severe and poorly controlled (Basch, 2011). Identifying effective health interventions for these students, rather than behavioral interventions, would not be detected as a possible solution without the know-how to look beyond standard data reports to mine information *across* reports.

A third commonly reported statistic relates to suspension and expulsion. Since the 1970s, many studies have clearly demonstrated that American Indian, Latino, and African American students are disproportionally suspended or expelled from school (Gregory, Skiba, & Noguera, 2010). Despite efforts to reduce this disparity, it persists. Causes have been proposed, including poverty, environmental and cultural attributes that differ from school, and higher rates of selection for formal discipline in school and the justice system. But one factor that is often overlooked is achievement. Low literacy achievement is often viewed as a product of high rates of disciplinary action, but more rarely recognized as an opportunity for intervention. For instance, Miles and Stipek (2006) reported that low literacy levels in the primary grades predicted aggression in 3rd grade. And though it is common for elementary schools to design and implement early literacy interventions, it is less common to look at discipline rates among participants in the literacy programs. Yet both the achievement and the behavioral data are already there, just waiting for someone to look at them together. That's just what a team of researchers

did with kindergarten students at risk for literacy failure. The researchers coupled a literacy intervention with positive behavior supports and found improvements on both measures (Volpe, Young, Piana, & Zaslofsky, 2012). By viewing both achievement and nonacademic hard data together, surprising relationships can come to light. And with that insight can come some very good ideas about what to do about it.

Demographic information about students, as well as nonacademic data such as attendance and discipline rates, can help us look deeper into the corners to address the needs of learners that we might not recognize as harbingers of patterns or trends. In addition, the ways we use hard data can herald a growing problem that we have overlooked. The ability to spot trends requires using the data both formatively and summatively.

Using Hard Data Formatively and Summatively

Hard data can be analyzed to inform or to summarize. Hard data are considered summative when used to describe the results of a process and formative when used to inform and adjust processes as they unfold. Summative assessments tell us how well students, schools, or districts did at the end of a given period of time. Teachers use summative data in their classrooms to determine students' unit, semester, and end-of-course grades. Similarly, schools and districts use hard data summatively when reporting annual growth or proficiency status. These summative analyses of hard data are used to rank states, districts, schools, and even groups of students.

A summative view of hard data is not without limitations, particularly because it gives us only a look at what has already occurred. Reeves (2005) compares summative data to an autopsy, where the patient is already dead and now it's a matter of determining what killed him. On the other hand, formative assessments are akin to physical examinations, "an uncomfortable ordeal, but… preferable to and less intrusive than autopsies" (p. 53). An important detail inferred in this analogy is that formative data should not become mini-autopsies. In the same way that you would be dismayed if your doctor didn't discuss the results of your physical to recommend healthy changes to your diet and lifestyle, so should you be if colleagues are only using formative data to evaluate, but not to make changes. You can't fatten pigs by weighing them (Jones, Carr, & Ataya, 2007). Formative use of hard data should inform us so that we can make thoughtful adjustments to maximize results.

Nonacademic data such as attendance, office referral, and suspension rates are not generally regarded as formative, but they should be. As we described at the beginning of this chapter, these data can be analyzed for trends and patterns. If there are outlier class periods, subject matter, or times of day when attendance either significantly spikes or decreases, data can be further analyzed to inform instruction and other school programs. Referrals are often overlooked in terms of informing instruction; however, if students are routinely removed from certain classes for behavior reasons then school leaders must not only examine student behavior but also classroom environment, environmental fit, and teacher preparedness. Much like referrals, suspension data can give us information about instruction. Unfortunately, when the instruction does not meet the needs of the students or if the instruction fails to acknowledge the contributions of all students, then they may become disengaged or even disruptive.

Likewise, demographic data should be reviewed each year to determine shifts in student subgroups and are imperative for improving instructional strategies and curriculum. Although demographic information is relatively stable from one year to the next, there are exceptions. For instance, after Hurricane Katrina devastated southern Louisiana and Mississippi in 2005, some school districts found themselves hosting significant numbers of traumatized children and families. They experienced a sudden onset of students who needed significant academic and psychological supports. An analysis of middle and high school students who fled the destruction found that the students who did best were those who transferred to schools that "created a milieu of cooperation… and neither went overboard trying to help nor ignored their special needs" (Barrett, Ausbrooks, & Martinez-Cosio, 2012, p. 7). Without question, these schools looked closely and frequently at their data to achieve this careful balance.

In addition to guiding how we modify or improve instruction and intervention, formative applications of hard data provide feedback to groups or individuals. A good feedback system clarifies the goals and is specific and descriptive of the work and its progress (Varlas, 2012). Assessment measures that can be used to provide feedback include baseline assessments, progress monitoring data, and benchmark assessments. Nonacademic hard data that can be used to inform instruction include attendance, referrals, and suspensions by class or period, teacher, school, or district. For example, a high school tracked tardy slips and office referrals over a two-year period as it switched to a later start time and found a significant reduction in both. The hard data were persuasive feedback

for families and community members who had opposed the change to a later start time and were doubtful that it would be effective.

It has become common practice to gather frequent monitoring data to determine student progress toward a predetermined set of goals, especially through a response to instruction and intervention process (Fisher & Frey, 2010a). These data often include a record of reading fluency scores or timed mathematics calculations. In practice, however, these formative assessments are unidirectional; that is, they are used to view student progress, but more rarely to evaluate instruction. Thus, a student's lack of progress is described just that way—it is the *student's* failure. But formative use of hard data must also be used to assess the effectiveness of the instruction. Using a formative lens, the record or log not only gauges student progress over very short periods of time but also becomes a tool for determining how instruction should be refined. Using the record is much like using the diagnostic data used in the medical profession to identify specific areas of need. With this information, a treatment plan is prescribed and monitored. A failure of treatment is seen as just that—it is not a failure of the patient, but rather that the treatment needs to be adjusted. Our teaching should be understood the same way, but it requires us to be willing to look at hard data as a mirror of our instruction, and not simply as a description of the inevitable (Dimmock & Walker, 2005). The use of data to inform instruction has proven to be a boon to students and teachers. Black and Wiliam's (1998) review of the formative use of data to improve instruction stated that the achievement gains associated with it are "among the largest ever reported for educational interventions" (p. 61).

In the same way that we use formative assessments to determine what should occur next in the instructional cycle, so should formative hard data be used to inform what should occur next in the school improvement cycle. A decline in absence rates merits a closer examination of what is proving to be effective for improving school attendance. A rise in disciplinary referrals should prompt further investigation into root causes.

Making the Data Useful

Schools that are successful—and know why they are—use achievement and nonacademic data throughout the year to guide and refine instructional practices, programs, and student placement. They pay attention to achievement by subgroups and use formative assessments to address the needs of each student. When state assessment results

are published in late summer, these schools have no surprises. Results can be predicted within a few data points. These successful schools use assessments to triangulate their findings; moreover, in such schools the process of using hard and soft data to inform instruction is cyclical.

Triangulation is the use of three or more data sources to arrive or converge on the same findings. It is a means of determining targeted information from multiple data sets. For an architect, triangulation means discerning with precision key load factors and points in space and time from other reliable and predictable data (White, 2011). If themes or findings are established by converging several sources of data, then the process of triangulating data sources adds to the validity of the findings (Creswell, 2008).

Steps to Inform Instructional Improvement

We provide a step-by-step process to make data useful in informing instruction. These steps can be used to transform ideas into action.

Start early. Data teams, leadership teams, and school leaders start the process of looking at school data at the beginning of each school year. But schools that use data to improve instruction don't stop there. They continue to review both hard and soft data throughout the year, paying close attention to everything from progress monitoring reports to quarterly benchmarks, and from classroom observations to annual state standards tests. The overall goal is to improve instruction by closely monitoring achievement results, nonacademic indicators, and classroom practices. Highly effective schools use both hard and soft data to refine programs, design teacher supports, and respond to intervention needs.

Look for links between practice and results. Identifying contributing factors that increase student achievement starts at the beginning of the school year. The instructional data team should analyze schoolwide state test results to determine how the school, grades, content areas, subgroups, and individual students performed. It can compare the results with the state assessments from the previous year. The team will then begin to ask questions about overall growth. Were there changes by subgroup, by grade? Were there content clusters that were outliers because the results were comparatively high or low? If so, the data team can review the previous year's formative assessment data to see whether the summative results could have been predicted earlier. Keep in mind Reeves's (2005) caution to avoid turning a data review into an autopsy. If the only outcome is to identify indicators, but no action is taken, then one shouldn't expect the future to be

much different from the past. The team's goal is to find changes and consider their causes in order to develop an action plan.

Expect the unexpected. If the results from the initial analysis are unexpected, the data team might then review and compare the alignment of formative and summative assessment measures. Use the additional data to calibrate the type of data the team is collecting and analyzing to see if they are in fact the right measures. The more aligned the state assessment is to district benchmark assessments, as well as teacher-created measures, the more accurately these predictors can inform practice.

Make the findings public and encourage speculation. The data team next presents these findings to the entire school and its stakeholders, including families and students, for further discussion. The data team displays the data in a way that is comprehensible and encourages discussion about other plausible reasons for the outcomes. Information is presented in the form of texts, tables, charts, and graphs to paint a picture of the results. (We will discuss visual displays of data in Chapter 4.) The format selected to present each piece of information should be carefully considered to ensure ease of understanding and to highlight key data. The same information is also presented in different ways to provide an analysis through various lenses:

- Schoolwide by content area and grade
- By content area, grade, and subgroup
- By content area and subgroup
- By content area, grade, and gender
- By content area and gender
- By content area, grade, gender, and language proficiency levels
- By cohort bands, content area, and proficiency levels
- By test year, content area, and subgroup
- By test year, content area, and grade
- By test year, content area, and gender

Drill deeper to examine classroom data. At the beginning of the school year, teachers should analyze their prior year state test results and end-of-year summative assessment to determine their own teaching strengths and challenges. By reviewing content subcluster areas for their prior year's students, teachers can determine where their instruction was effective and where it needs to grow or change. Strong content subcluster area results are one indicator of effective instruction, but do not stand alone. Teachers should review

these subcluster results along with classroom observations and nonacademic data to glean information about student performance as well as to self-assess their performance.

Anticipate the needs of new students. Rearview data from the previous year can be further leveraged to identify the needs of incoming students. Teachers should expect to have state standards assessment results for their new students to determine where needs might be. It is especially important to be proactive by identifying students who have already amassed a record of academic or social struggle. No child should have to wait until the first reporting period is over before being noticed as a struggling student. It is much easier to begin the year with anticipated personal, curricular, or technological supports in place than to try to make up lost ground after 25 percent of the school year is over. Teachers across a grade level or discipline can meet to look for larger patterns and trends to identify whether more intensive supports are warranted. For example, in our school, the 11th grade team matched incoming students' test results from the previous year to locate students who needed coordinated academic support. Traditionally these hard data would have remained at the individual teacher level, and the chemistry teacher would not have realized until much later that a struggling student was also having trouble in math. Instead, the grade level team identified 24 students in need of academic support before the school year even began. The team was able to build the students' schedules so they could access the school's academic recovery resources from the first day.

At the elementary school where Cheryl was principal, grade level teams of teachers met at the beginning of the school year to analyze state test data. Students who scored in the Below Basic and Far Below Basic range on the California Content Standards tests were identified for diagnostic assessments in an effort to identify their specific areas of need. Students were then clustered in fluid groups by need for intervention during the school day, and in some cases had additional intervention time before or after school. The specific information from the test was also used for unit planning to determine when to slow down instruction and when to increase the pace.

Pay close attention to historically underserved subgroups. Analyzing state assessment data by subgroups helps to ensure that students who may have been historically overlooked receive instruction that meets their needs. Districts, schools, and individual teachers must examine the content delivered, the method of delivery, and the cultural proficiency of those delivering the instruction (Lindsey, Robins, & Terrell, 2009). This point about subgroups was brought to our attention by the 9th grade team members, who were concerned that advanced students were not being sufficiently challenged. They

identified specific students and assigned each student a mentor teacher from their team. Throughout the year, they met regularly with identified students to hold academic discussions. Word about these mentoring meetings soon got out, and the 9th grade team began hosting "power lunches" that were open to all students. These power lunches featured guest speakers from various professions. Although the individual mentoring meetings still continued, the intervention for advanced students broadened to meet the interests of a wide array of students.

Plan regularly scheduled dates to analyze interval data. After the first benchmarks are administered, the data team will reconvene to analyze the new hard data and to develop charts, tables, and graphs to share the data with the rest of the staff. Once again, the team will analyze these data across content areas, grades, subgroups, cohorts, teachers, and gender. The results will be compared to state test results and the previous year's first benchmark results to gauge growth over time. This process will again be repeated at the end of the second and third benchmarks with more information being added for discussion.

Instructional Improvement in Action

Whether they are new to a school site or returning for their tenth year, principals begin by querying the data. They must analyze their students' results from both state and local assessments. Principals have to analyze data for trends and patterns in student outcomes to determine academic strengths and weaknesses. In addition, they review classroom observation data and outcomes from learning walks, requiring an in-depth exploration of the connection between instruction and student performance. To ensure that the information is correct, they triangulate it with other data sets to ensure accuracy.

For example, Principal Mario Marcos begins a new elementary school assignment in early August, when the state test scores are just released. He reviews those scores along with the end-of-year district course assessments. In his district, students take summative assessments in math, writing, and reading to determine student growth over the school year. Mr. Marcos also consults the attendance records, reviews suspension and expulsion rates, and reads the classroom observations that had been conducted by the former principal.

Mr. Marcos knows that an isolated review of the data is inadequate to his school's needs and that the data team is a rich resource, especially to someone new to the school.

The data team can offer the context of the last school year, especially as it relates to the school's goals and its ongoing professional development. He invites the team to meet with him to learn about the data together. Mr. Marcos posts questions around the room to guide their exploration of the numbers.

- How did students do overall in math, reading, and writing?
- Were there specific strands that ran across all grade levels in which students were notably strong or weak?
- Are there patterns we see across subgroups?

Mr. Marcos and the data team begin to plug in the numbers on chart paper in order to see if any patterns are emerging. After posting the state standards-based test results with the district end-of-course data, they compare the two. One member of the team remarks, "According to the district assessment, we performed well in math, but that's not what the state test results say." A discussion of possible explanations ensues, and Mr. Marcos listens carefully. At this early stage of analysis, he is aware that data teams often want to move prematurely to solutions. "Those are interesting possibilities," he says, "but let's keep digging."

He and the team examine additional hard data. They view last year's suspension and expulsion rates, and though they find them to be low compared to the district average, they are the highest they have been in five years. "That's really disturbing," another data team member observed. "I wonder what happened last year?" Mr. Marcos reinforces her question. "That's a great question, and that's the point of this data dig. We want to pull the questions to the surface." He adds her question to another chart labeled "Questioning Our Data." As he writes he says, "We've got two now. 'Why are there discrepancies in math performance between our state test results and those on the district assessment?' And the second is, 'What caused our suspension and expulsion rate to rise so much last year?' We'll use these questions to guide our next steps." By the end of the morning, Mr. Marcos and the data team have reviewed student performance data, nonacademic measures, and last year's professional development goals. In addition, they have generated six questions and are beginning to see that it's too early to propose solutions. "We just don't have enough information yet," a team member reasoned. "We need to look at the soft data, too." Mr. Marcos smiled and said, "Let's have lunch first, then we'll take that on this afternoon."

Quality Assurance

Throughout this chapter we have discussed the types of hard data that are available to administrators and data teams. The data that come to mind first have to do with student achievement, especially on state standards-based tests. Although these provide necessary information, they are limited in scope as they highlight only one aspect of student performance. Other hard data include safe schools reports, attendance records, suspension and expulsion incidences, and even health records. Another limitation is the way in which hard data are conventionally used. They are most often a look backward at what has already occurred. It is informative to know where you have been. But in the same way that you can't drive a car by only looking in the rearview mirror, you can't steer a school toward improvement by only looking at last year's results. These data are made more useful when they are examined for trends and compared to data across several years. In addition, hard data are often right under our noses, waiting for us to spot them. Other than end-of-year assessments, assessments can be analyzed formatively throughout the school year to make timely adjustments and to design new interventions. In all cases, the goal is to make data results widely available to stakeholders who can nudge students, educators, and families toward action. As you review the hard data results at your school, keep these quality assurance questions in mind:

- What do the achievement results report?
- Are these achievement results consistent across subgroups?
- Are there subgroups who are underserved?
- What nonacademic data are available to us?
- Are we examining the hard data at regular intervals?
- Are we examining the hard data across reports as well as within?
- Do our observations result in action?