Part Four
MOVING TO BETTER PRACTICE

CHAPTER 13

Data Use and School Improvement: Challenges and Prospects

MARGARET HERITAGE AND RAYMOND YEAGLEY

In 1922 Edward Thorndike wrote, "The task of education is to make changes in human beings. For mastery in this task, we need definite and exact knowledge of what changes are made and what ought to be made" (p. 2). Seventy-nine years after he wrote these words, the 2001 No Child Left Behind (NCLB) legislation stated, "Schools must have clear, measurable goals focused on basic skills and essential knowledge. Annual testing in every grade gives teachers, parents and policy makers the information they need to ensure that every child will reach academic success" (p. 7). While the language of the two may be different, the message is the same: schools need accurate and actionable information about what students know and can do so that they can plan effectively for student learning.

Margaret Heritage is the Assistant Director for Professional Development at the UCLA National Center for Research Evaluation, Standards, and Student Testing (CRESST). Raymond Yeagley is the superintendent of the Rochester School Department in Rochester, New York.

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Education Sciences (IES), and the U.S. Department of Education. The findings and opinions expressed in this report are those of the authors and do not necessarily reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences (IES), or the U.S. Department of Education.

However, there is a real difference in the context in which each of these two statements was made. In 1922, the measurement movement in education was in its infancy. Thorndike noted, "The first steps to establish units of education products and to devise instruments to measure them with reasonable precision were taken about a dozen years ago" (Thorndike, 1922, p. 3). The years since have brought both a significant expansion in educational measurement and an associated demand for accountability. To ensure that all students meet the standards that have been set for them, federal, state, and district accountability mandates require schools to measure student performance against established standards, with sanctions for schools that do not make adequate progress to close the gap between low- and high-performing students. These sanctions mean that high stakes are now attached to educational measures, and it has become imperative for educational practitioners to understand measures of achievement and to take action to improve student learning based on what the achievement data tell them.

However, despite the contemporary prominence of achievement data as vehicles to school improvement and research evidence showing that data use is a characteristic of effective schools and districts (Cawelti & Protheroe, 2001; Doolittle, Herlily, & Snipes, 2002; Teddlie & Reynolds, 2000; Tognieri & Anderson, 2003), there remain substantial challenges to the meaningful integration of data use into the everyday practices of schools. This chapter will examine characteristics of effective data use, outline challenges in the implementation of school-based data analysis, and offer perspectives as to how best practice in data use can become a widespread reality in schools.

District and School Inquiry

The starting point for any form of data analysis by education practitioners is to decide what questions they want the data to answer. Herman and Gribbons (2001, p. 5) identified three basic questions that are foundational to district and school inquiry:

- How are we doing?
- Are we well serving all students?
- What are our relative strengths and weaknesses?

And these questions can lead in a more action-oriented direction:

- Why are things the way they are?
- How can we make them better?

These questions are not, of course, confined to schools and districts: classroom teachers will need to answer the very same questions about their students to make plans for student learning. Finding answers to these questions will depend on the degree to which practitioners, at each level of the system, from classroom to district, have a set of useful data to query.

Essential Characteristics of Data to Guide School Improvement

There are essential features that achievement data must have if practitioners are to use these data to make decisions about school improvement. In brief, achievement data must be: (1) aligned; (2) valid and reliable; and (3) sensitive to differences.

Alignment

Alignment is the match between the expectations for students and the assessments used to measure whether or not students are meeting the expectations (Herman, Webb, & Zuniga, 2003; Webb, 1997). As accountability demands increase, the alignment between assessments and state standards for learning remains essential (Webb, 1997). Without strongly aligned assessments and standards, educators will be "chasing the wrong goal and policy makers will not have a good gauge of whether schools are producing the desired results" (American Educational Research Association [AERA], 2003, p. 1). The closer the alignment between standards and assessments, the greater the likelihood that teachers will focus on the desired content and that students will have a fair chance to demonstrate what they know (AERA, 2003). Recent studies have demonstrated that there is a significant need to improve the alignment between standards and assessment (Porter, 2002; Rothman, Slattery, Vranek, & Resnick, 2002; Webb, 1999).

Assessment Quality

Accurate conclusions from assessment results that can be used as the basis for sound decision making about student achievement depend on assessment quality. There are two key concepts to determining assessment quality: validity and reliability. Validity is the extent to which a test measures the underlying quality it is intended to measure. It is evaluated by considering the degree to which accumulated evidence and theory support specific interpretations of test scores entailed by the proposed use of the test. Reliability is the degree to which test scores for a group of test takers are consistent over repeated applications of a

measurement procedure. Simply put, the validity of an assessment is dependent on the purpose for which it is being used, while reliability refers to the consistency of scores over time and across raters. For example, an assessment that focuses on multiplication may give highly consistent results but remains an invalid measure of students' math problem-solving skills (Herman, Aschbacher, & Winters, 1992).

Sensitivity to Differences

If administrators and teachers are going to make sound decisions about improving learning, they need accurate information about the performance of *all* their students. However, there are challenges associated with deriving accurate information from outcome data for certain populations. Historically, English language learners (ELLs) perform lower than other students (Abedi & Dietel, 2004). Research shows that the language demands of tests negatively influence the accurate measurement of ELL students (see Abedi, chapter 8, this volume). Rather than an accurate measure of content knowledge, tests become measures of both achievement and language ability (Bailey & Butler, 2004). One way to address the issue of accurate measurement of ELL student achievement is to reduce unnecessary linguistic complexity, simplifying the language of the test without diminishing the rigor of the test (Abedi & Lord, 2001). Providing extra time or dictionaries for ELL students are further test accommodations that can be made for more accurate assessment results. Another way to improve the accuracy of assessments for ELL students on content tests is to ensure that their English proficiency is of a type and level to handle the complexity of language used on the content tests by developing academic English assessments (Bailey & Butler, 2004). Similar issues arise for students with disabilities (see Pullin, chapter 9, this volume).

Types of Data for School Improvement

What kinds of achievement data are potentially available to districts, schools, and teachers to guide school improvement? We will address four main types of data, presenting at very different levels of granularity and detail in the system.

Large-Scale Achievement Tests

Current accountability systems rely on large-scale achievement tests as a lever in standards-based reform. Test results are expected to inform educators about student performance with regard to state standards and

on how to improve student academic achievement. Large-scale assessments can provide general information on student achievement and comparative data for stakeholders to make judgments about the quality and effectiveness of programs, schools, and districts. Whatever the value of these tests in providing a framework for macro-level decision making, their value in providing the kind of real-time guidance necessary for curriculum or pedagogic adjustment that is the fulcrum of school improvement is quite limited (Baker, 2001; Herman & Gribbons, 2001; Shepard, 2004). In particular, the infrequency with which these tests are administered and the large period of instruction they cover significantly limit their effectiveness. Practitioners need timely, accurate, detailed, and comprehensive information to provide guidance for ongoing teaching and learning and to steer school improvement efforts. The time lag between the administration of the tests and the publication of their results vitiates their value to teachers and school administrators as far as the cohort of tested students is concerned. Moreover, as is widely recognized, the range of curriculum and instructional practices embraced by this form of testing is large. Hence, these tests unavoidably yield results reflecting aspects of learning at a coarse-grained level, and do not give information about the detailed nature of student learning needed by school administrators and teachers to provide a positive direction for improvement. For this reason, educational policy researchers increasingly advocate multiple probes as an adjunct to annual assessment (see, e.g., Baker, Linn, Herman, & Koretz, 2002). Indeed, the NCLB legislation itself makes provision for the use of multiple forms of assessment, including assessments that are diagnostically useful.

Moreover, it is well recognized that, in the realm of educational measurement, "One assessment does not fit all" (National Research Council [NRC], 2001, p. 220), and that the validity of inferences drawn from a measure depends on the purpose for which that measure is intended (AERA, 2000; AERA, APA, & NCME, 1999). To draw valid inferences about overall school and individual performance, and to provide a sufficiently detailed picture of student achievement to make decisions about school improvement, it is essential for educational practitioners to use evidence from a range of measures. Finally, while assessment data give information about the *level* of student performance, data from other sources can be used to examine the factors that *contribute* to student performance. Understanding the context of student achievement can be just as central to school improvement as knowing the parameters of test performance (Baker et al., 2002).

Benchmark Assessments

In an effort to give principals and teachers more timely and detailed assessment information, many districts are now providing administrators and teachers with benchmark assessment data. Benchmark assessments provide multiple occasions of measurement throughout the year and the results are intended to guide instruction and to identify areas of improvement. For example, in Rochester School District NH, teachers administer benchmark assessments that are aligned to the state standards up to four times each year. The superintendent (one of the authors of this chapter) reports that some teachers in the district are beginning to take advantage of the multiple administration option for instructional guidance, as well as to identify students about whom they need more diagnostic information. At the district level these data are being used as a predictive measure for the state test. Recent studies have highlighted the potential benefits of benchmark assessments when they are used by teachers and principals who have received training in interpreting these data (Snipes, Doolittle, & Herlihy, 2002).

Benchmark assessment has the potential to provide achievement data at a level that is intermediate between large-scale and classroom-based assessment. Most benchmark assessment is intended to be aligned with state standards and can, in principle, provide practitioners with feedback about student performance and prospective achievement, together with some guidance to adjust curriculum and instruction. For these reasons, it is likely that benchmark assessment will be a context in which data use will expand in the next several years. This expansion, however, will require a significant investment in training practitioners to interpret results and to use those results to effect improvement.

Formative Assessment

Black and Wiliam (1998, 2004) stress the importance of formative assessment above other kinds of assessments as sources of information for teaching and learning. An assessment is formative "when the evidence is used to adapt the teaching work to meet the learning needs" (Black, Harrison, Lee, Marshall, & Wiliam, 2003, p. 2). In their landmark meta-analysis of formative assessment, Black and Wiliam (1998) established a very strong body of evidence that improving formative assessment can raise the standards of student performance. Moreover, scholars are increasingly advocating the inclusion of classroom-based formative assessments in accountability systems (Wilson, 2004). However, the use of classroom-based formative assessment as a tool for school improvement is not widespread.

A central problem in the use of classroom-based assessments for accountability and for instructional decision making is that most available classroom assessments will likely not have met the same standards of validity and reliability as external assessments used for accountability purposes (Salinger, 2001; Shepard, 2000). One of the reasons for this is that a considerable imbalance exists between the amount of funding given to the development of classroom-based measures and the resources channeled into large-scale assessment (NRC, 2001; Stiggins, 2002). This imbalance is indicative of the traditional emphasis on large-scale assessment in the educational system. A redress in this imbalance could result in more classroom-based assessments, linked to the instructional goals of teachers, which could be integrated with information from the annual tests to provide a fuller and more detailed picture of student achievement. Indeed, Stiggins (2002) proposes that every dollar invested in large-scale assessments should be matched with another dollar devoted to the development of classroom assessments.

Grading

Grades are without question the most common measure of student achievement in schools. However, Cross and Frary (1999) note that classroom grading has been a source of controversy since the beginning of the 19th century. The basis of all the controversy lies primarily in three areas: teachers use nonachievement factors like behavior and attendance when they factor grades; they weight assessments differently; and they use a single score to represent student performance on a range of skills and abilities. The result of these grading practices is a system that is unreliable and potentially invalid as a measure of achievement (Marzano, 2000). Teacher grades as they are presently conceived do not provide particularly useful data on which to make decisions about school improvement.

However, teachers are the ones who have the most direct knowledge of students and they make judgments about students' learning every day. If current grading practices are an unreliable means of including teachers' judgments in school improvement efforts, how could the system be improved to make use of teacher knowledge? One approach is to situate teacher knowledge in a community of judgment—an interpretive system that connects teacher judgments to student work through a process of moderation (Wilson, 2004). A system of moderation involves two components: *assessment moderation* and *verification*. Assessment moderation concentrates on judging student work. Teachers meet regularly to rate student responses using scoring guidelines

and apply consistent and fair grading practices in the assessments. Verification of teachers' ratings is the second part of the moderation system; rated student work is either sampled and rerated by external experts, or a technique of statistical moderation is used (for a full discussion of this technique, see Wilson, 2004). Such a system of moderation would enable teacher knowledge of their students to play a significant role in school improvement efforts and would serve, additionally, as an instrument of professional development.

Going beyond Assessment Data

While assessment data give information about the *level* of student performance, data from other sources can be used to examine the factors that *contribute* to student performance.

In addition to achievement data, Bernhardt (1998) has identified three domains of data that can provide contextual information: (1) demographic data (e.g., grade level, ethnicity, language spoken at home); (2) perception data, which can reveal student, teacher, and parent attitudes about learning, teaching, and school programs—usually obtained through surveys of stakeholders; and (3) school processes, which include information about curriculum, teaching strategies, student discipline, parent communication, teacher qualifications, professional development, and any other aspect of school operation that can have an impact on student learning.

All these data can be integrated to deepen practitioners' knowledge base and provide a broader understanding of what is, and what is not, working in their schools. For example, a Los Angeles high school was concerned about the math performance of Latino students, which was considerably lower than all other student groups, including other minorities. Further inquiry determined that a subgroup of Latino students and a subgroup of African-American students, in particular, were struggling. Both groups, as it turned out, were being bussed from the same distant neighborhood as part of the district's mandated desegregation plan. Parents indicated that their children were too tired after a full day, including two long bus rides, to spend sufficient time on homework. After-school programs, successful for other students, were unavailable because of the immediate bus departure. The decision was made to place a tutor on the bus to help students complete their homework during the bus ride home. While this single intervention did not resolve all of the academic challenges for these students, there was a meaningful improvement in performance from both groups, and the program was considered a success. The program was a result of com-

bining assessment data and academic marks with demographic data that went beyond simply racial classification. This was further supplemented with perception data from a parent survey, and resulted in a change in school processes that was subsequently found to be effective (D. Mitchell, personal communication, August 15, 2004).

To reap the benefits of using data in this way requires practitioners to "think outside the box." NCLB requirements have led to a greater focus on data but in the authors' experience, many practitioners are currently stuck on analyzing student characteristics in relation to achievement data. They may know whether certain ethnic groups are performing better than others on achievement tests, but they do not know *what* to measure to search for the reasons for the differential performance levels. They may want to move beyond this level of analysis but do not understand the possibilities for creating data elements and indicators. It is our belief that the more practitioners are exposed to examples of *combined* data use that encourage unique solutions to problems, the more they will be able to collect and use other types of data in conjunction with achievement data for school improvement.

Integrating Data Use into School Practices

Integrated Assessment System

Earlier we highlighted the imbalance that currently exists between the amount of funding allocated to large-scale versus intermediate and classroom-based assessments. However, increasing funding for classroom assessments to strike a better balance will not necessarily result in a system that satisfies the needs of both policymakers and practitioners. In *Knowing What Students Know* (NRC, 2001), a committee of the National Research Council laid out an ambitious vision for a coordinated system for assessment. The committee outlined three characteristics of such a system:

- *Comprehensiveness.* A system that includes a range of measurement approaches to provide the evidence for educational decision making
- *Coherence.* A system that combines large-scale and classroom-based assessments built on the same underlying model of learning with consistent constructs for both levels of assessments
- *Continuity.* A system that includes measures of students' progress over time to provide a continuous stream of evidence about performance

The committee advocated that these properties would be aligned along three dimensions: vertically across levels of the education system; horizontally across assessment, curriculum, and instruction; and temporally, across the course of a student's education. These proposals highlight the undesirability of a disorganized and unplanned aggregation of ad hoc assessments. In the current context of high-stakes achievement tests, teachers tend to allocate more time to teaching what is on the test at the expense of the broader content of the subject (Stecher & Barron, 2002). The result is a narrowing of the curriculum and an increased focus on the less demanding aspects of the standards (Rothman et al., 2002; Stecher & Hamilton, 2002). The benefits of a coordinated system of assessment are clear: policymakers and the public will have access to the information they need for monitoring and accountability, while practitioners all the way down to the classroom level will be able to take full advantage of assessment information, which embraces the breadth and depth of curricula content, for educational decision making.

Interpreting Assessment Information

Traditionally, neither administrators nor teachers have received formal training in how to assess students or how to make use of assessment information. As Stiggins (2002, p. 5) notes, U.S. educators remain "a national faculty unschooled in the principles of sound assessment." A primary consideration for pre-service and in-service training is that practitioners need to have the skills to gather relevant information and to interpret it to maximize achievement.

Without knowledge of assessment principles, many practitioners will likely not be able to judge the degree to which standards and assessment are aligned, nor to evaluate the validity and reliability of assessments. For example, even if teachers are using a math assessment that is closely aligned to standards and covers the content taught, they may not realize that it is an invalid measure of math achievement for *all* students if it includes vocabulary and a text level that is beyond some.

In an effort to increase assessment literacy across the country, Stiggins (2002) advocates a number of actions. These include establishing a comprehensive program at the national, state, and local levels to foster literacy in classroom assessments, together with a similar program in effective large-scale and classroom assessment use for state, district, and building administrators. Also, he advocates changes in the licensing requirements of all educators to include an expectation of competence in assessment. These are important steps that need to be taken, because

without them, even if the integrated assessment system described earlier existed, practitioners will not have the competence to interpret strengths and weaknesses in student performance or to design appropriate interventions.

Data Tools

School data often exist in disparate forms and locations, making it difficult to organize efficiently and to retrieve quickly (Thorn, 2001). This situation prevents educators from easily storing and retrieving data to use for school improvement. Data tools provide a solution to this problem and can permit easy access to a wide range of data.

Most schools use an electronic student information system (SIS) that assists in creation of the master schedule and organizes individual student attendance, discipline, and academic records. However, for effective data use in school improvement, schools and districts need a data warehouse and data analysis tools beyond their SIS. A June 2004 survey revealed that only 20% of New Hampshire school districts were using any recognized data warehouse and analysis tool. An interesting finding from the survey was that several districts believed that they *were* using such a tool, citing only their SIS and/or spreadsheet as the tools used for this crucial function. Current SIS products will not track student achievement from year to year for analysis purposes and do not have any strong analysis capabilities.

Increasingly, there are data tools that allow administrators and teachers significantly enhanced access to data. Although the features of data storage and analysis tools vary considerably, currently available tools permit educators to store, access, and analyze data effectively (Wayman, Stringfield, & Yakimowski, 2004). These tools can be categorized not so much by their data management and analysis capacity as by the way educators access them and the level of outside support available to the school or district.

At the highest level of support are tools in which the vendor acts as the technical arm of the school or district by providing assistance in: (1) identifying data elements and indicators; (2) cleaning, importing, and storing data; (3) developing pertinent questions for analysis; (4) conducting the analysis and generating reports; and (5) assisting the school, if desired, in interpreting the results.

In a more common model, the vendor will serve as an application service provider, working with the school initially to identify data elements and to custom design the data map and preformatted reports.

These vendors may assist in data cleaning and import, and usually will provide further customization as the school's data needs and sophistication increase. But generally, these vendors will not be deeply involved in ongoing training of school personnel as regards question design and interpretation of reports.

A third model provides a tool housed on a local server, with technical training in using and operating the software, training for users in research design and data use, and ongoing technical support for the database. School and district personnel are responsible for their inquiry and analysis to inform educational decisions.

A fourth model, which is now emerging, is an add-on to the student information system that permits some analysis and maintenance of data elements not historically stored in the SIS.

Most of these tools now have reporting capabilities specific to NCLB, and all permit a variety of analyses that can meet some of the demands of accountability while informing instructional planning and school improvement. (For more details on these models, see Wayman et al., 2004.)

Whatever the model adopted and whatever the level of support available, several tasks remain before practitioners can begin data inquiry. School data can exist in many different places, from a main district data store to excel spreadsheets to loose papers in files. One task is to take an inventory to identify data sources and locations and from this inventory determine which data will be useful for analysis and will be imported into the data tool (Wayman et al., 2004). Another task is to make sure that the data are "clean." Inaccurate and incomplete data will present problems for data analysis, and so educators will need to assess the quality of the existing data and take the necessary steps to ensure clean data. Finally, it will be important to identify the resources for any data cleaning and for importing data into the system. Initially, this can represent a considerable time commitment and not all schools and districts will be able to afford to have vendors or dedicated personnel for this task. A study undertaken at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California at Los Angeles (UCLA) on the implementation of a data analysis tool, the Quality School Portfolio (QSP), revealed a possible solution to the personnel problem. Several district consortia pooled resources and were able to dedicate personnel to data cleaning and importing, making data analysis possible for all their schools (Chen, Heritage, La Torre, & Lee, 2005).

Data Analysis Practices

The availability of data analysis tools represents a step forward in using data for school improvement. However, educators frequently lack the skills for making effective use of data (Baker, 2003; Choppin, 2002; Cizek, 2000; Cromey, 2000). In the CRESST study of QSP implementation, a majority of administrators, most with advanced degrees, reported that they were not experienced in data analysis (Chen, Heritage, Danish, Choi, & Lee, 2003). Engaging in a systematic process that involves the skills of defining questions, making meaning from the data, taking action, and evaluating the action can provide the framework necessary for effective data use.

Defining Questions

A frequently asked question in schools might be, "Is the new standards-based math program more effective than our previous math program?" This question certainly identifies something of value for a school or district to know, but it provides little focus for data collection and inquiry. Asking such an overly broad question is a common error made by untrained K-12 practitioners. Unfocused and incomplete questions of this kind will leave practitioners scratching their heads about how to identify and analyze the necessary data, and about where the questions will ultimately lead. CRESST has developed a structure for questions that helps to focus analytical efforts and drive more productive inquiry. The structure involves six components useful in focusing questions for local data-based inquiry: (1) What is the purpose of the investigation? (2) What will be measured? (3) How will it be measured? (4) When will it be measured? (5) Who will be assessed? (6) How will the results be interpreted? (Heritage, 2003).

Consequently, a more effective question to examine and compare the results from the standards-based and the traditional math programs, structured with the CRESST model, might be: What is the magnitude and direction of the difference in Rasch Unit score growth from fall to spring 2004–2005 on the Northwest Evaluation Association math assessment for fourth-grade students in classes piloting a standards-based math program and students in fourth-grade classes using the district's previously adopted math program?

Deepening the Inquiry

The use of descriptive data has become a national mandate through NCLB, which requires description of academic performance account-

ability test data by race/ethnicity, gender, English proficiency, migrant status, economic disadvantage, and disability. While these categories are useful for highlighting if some subgroups are achieving less well than others, to be truly effective, disaggregation needs to go deeper than a single characteristic like race or gender. Deeper levels of disaggregation are essential for determining the effectiveness of school practices and the equity of services for different populations within a school and district, and educators should look at subgroups within the major categories. For example, academic achievement over time may be better for ELLs who have received English language instruction in one program than for those in another. Further, the program may be more effective for students whose native language is Spanish than for those whose native language is Chinese. Disaggregation of subgroups within the major categories could be used to examine whether the introduction of a new math curriculum is effective in closing the achievement gap and increasing the participation of underrepresented groups in algebra and higher-level math courses.

Descriptive data alone will have limitations as tools of school improvement. If practitioners only examine test scores and student characteristics, they are unlikely to discover the reasons for differences among groups. Deeper understandings mean developing analytical skills that go beyond simple bivariate relationships. Techniques that involve multivariate analysis can be used to examine the relative effectiveness of programs, instructional techniques, staff and resource allocation, and other factors that may have an impact on student learning.

While data are valuable for informing decisions in schools, relying on data alone, without the wisdom of experience and the caution of thoughtfulness, can lead to disaster. In a recent meeting attended by one of the authors, a state official cited research correlating mothers' educational attainment with their children's relative success in school, and pointing to poorly educated mothers as a major cause of poorly performing schools, suggested that getting more future mothers into college would solve most of the problems! Misinterpreting data or relying on a single, often unreliable, data point to make crucial decisions may be even more detrimental to a school and its students than having no data at all.

Every data-based conclusion that will have a major impact on educational decisions needs to be viewed through the microscope of common sense and reasonableness. Data can inform human judgment but should not replace it (Jamentz, 2001; Secada, 2001). Is the sample of students large enough to support a solid evidence-based conclusion?

Does a strong correlation between two variables really suggest causation, or are both being driven by a third factor not measured? Will an apparently obvious solution have unforeseen public support ramifications that may ultimately be detrimental to the school? Should local data that vary substantially from known national results be trusted or retested? To make intelligent use of data, it is important to know how to differentiate between useful and useless data, gold and garbage.

Taking Action and Evaluating Results

Once an analysis of the data is completed, the next step is to decide what action to take for school improvement. This requires practitioners to identify priorities, establish goals, set targets to achieve the goal, and determine strategies to reach the targets. As important as data analysis skills are, any possibilities for school improvement will be thwarted if practitioners do not have the necessary skills to translate analysis into action. Furthermore, practitioners will need to recognize that data use for school improvement is a continuous process. Any change in curriculum, instruction, and organization, and any program interventions designed to improve student achievement, require monitoring and evaluation. This involves determining what data will be collected for evaluation purposes, the frequency with which these data will be gathered, and the kind of analysis that will be undertaken to determine the effectiveness of the interventions.

Data Culture

Although data analysis tools and skills are essential elements of increased data capacity in schools, they will remain largely ineffective if teachers and administrators are unwilling or unable to use them. For practitioners to develop a commitment to data use, district and school cultures that trust data and support high-quality data use must be nurtured.

The culture of a school consists of the expectations, beliefs, and behaviors that constitute the norm for the school and the district (DuFour & Eaker, 1998). A culture that supports data use incorporates an expectation that judicious use of evidence should drive the day-to-day practices of administrators and teachers, as well as the collective decision making in the school (Jamentz, 2001). Stakeholders believe that data should inform human judgment and that planned and targeted data use can provide the necessary indicators to improving student learning. A data culture is one in which teachers and administrators work together in a community of practice—trusting data, focusing on

results, and engaging in using data for systematic reflection and planning. Simply put, in a culture that supports data use educators will say, "Using data is the way we do things around here."

In this era of high-stakes accountability educators can harbor suspicions of data, lacking trust in its validity and suspecting that data will be used against them. Rather than viewing data as a vehicle for school improvement, they can see data as penalizing and punitive (Cromey, 2000). Counteracting this view and instilling trust in data involve building a school culture that is supportive of data use and that makes data use transparent, open, and inclusive, enabling the educators to have confidence in the value of data use for improving student achievement (Katz, Sutherland, & Earl, 2002).

Leadership is essential to develop a culture that supports data use (Herman & Gribbons, 2001; Jandris, 2002; Mason, 2001). Marshalling the school's community to a collective sense of purpose about improving student learning, accepting that data use can and will improve learning, aligning data use to school planning and decision-making processes, and creating time for data analysis are key elements of leadership in creating a culture for data use. While school administrators will be pivotal in shaping the culture, leadership does not reside solely with them. Schools can increase leadership capacity by developing the data use expertise of teachers who can provide readily available and site-specific assistance to their colleagues (Feiler, Heritage, & Gallimore, 2000).

Conclusion

No reader of this chapter can fail to be impressed by the magnitude of the task before us if education is truly to become an evidence-based discipline. It is clear that if schools and districts are to make effective use of data a number of far-reaching changes involving a considerable investment of economic and human capital will have to be made.

First, and perhaps least challenging, is the investment in hardware, software, and "pump-priming" data input that will be required before any form of integrated data-driven analysis can get underway. Second, and considerably more challenging, is the investment in human capital required to develop the assessment literacy and data analysis skills that will, ideally, reach from district to classroom level. There can be no question that this is a long-term project that will require considerable changes in in-service and, most importantly, pre-service education. Academic educators have a significant role to play in adjusting the practi-

tioner training curriculum and culture to favor a data literate profession. Policymakers, test developers, and practitioners have an important obligation in the NCLB context, to devise, test, and implement an integrated system of assessment while also ensuring that such a system does not generate perverse incentives to focus instruction on "teaching to the test." The objective of assessment is to provide structure, rather than stricture, for professional practice. Finally, leadership is essential at every level of the system to develop a culture of evidence-based school improvement that will eventually become second nature to all practitioners, regardless of the scope of their responsibilities. Creating the conditions for effective data use is a necessary precursor to the widespread reality of data use as an engine of school improvement.

References

Abedi, J., & Dietel, R. (2004, Winter). *Challenges in the No Child Left Behind Act for English language learners* (CRESST Policy Brief 7). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*, 219–234.

American Educational Research Association (AERA). (2000, July). *AERA position statement concerning high-stakes testing in PreK-12 Education*. Retrieved August 18, 2004, from http://www.aera.net/about/policy/stakes.htm

American Educational Research Association (AERA). (2003, Spring). Standards and tests: Keeping them aligned. *Research Points*, 1(1). Retrieved from http://aera.net/uploadedFiles/Journals_and_Publications/Research_Points/RP_Spring03.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bailey, A.L., & Butler, F.A. (2004). Ethical considerations in the assessment of the language and content knowledge of U.S. school-age English learners. *Language Assessment Quarterly*, *1*(2&3), 177–193.

Baker, E.L. (2001). Testing and assessment: A progress report. *Educational Assessment*, *7*(1), 1–12.

Baker, E.L. (2003). *From usable to useful assessment knowledge: A design problem* (CSE Technical Report 612). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Baker, E.L., Linn, R.L., Herman, J.L., & Koretz, D. (2002, Winter). *Standards for educational accountability systems* (CRESST Policy Brief 5). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bernhardt, V.L. (1998). *Data analysis for comprehensive schoolwide improvement*. Larchmont, NY: Eye on Education.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. New York: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, *5*(1), 7–73.

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability. 103rd yearbook of the National Society for the Study of Education*, Part II (pp. 20–50). Chicago: The National Society for the Study of Education.

Cawelti, G., & Protheroe, N. (2001). *High student achievement: How six school districts changed into high-performance systems*. Arlington, VA: Educational Research Service.

Chen, E., Heritage, M., Danish, J., Choi, S., & Lee, J. (2003, August). *Evaluating the web-based quality school portfolio: Year two research report* (CSE deliverable). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Chen, E., Heritage, M., La Torre, D., & Lee, J. (2005, January). *Evaluating the web-based quality school portfolio: Final Report* (CSE deliverable). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Choppin, J. (2002, April). *Data use in practice: Examples from the school level*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. Retrieved April 22, 2002, from http://www.wcer.wisc.edu/mps/AERA2002/data_use_in_practice.html

Cizek, G.J. (2000). Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practices*, *19*(2), 16–23.

Cromey, A. (2000). Using student assessment data: What can we learn from schools? *Policy Issues, Issue 6*. (ERIC Document Reproduction Service ED452593).

Cross, L.H., & Frary, R.B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, *12*(1), 53–72.

Doolittle, F., Herlily, F., & Snipes, J. (2002). *Foundations for success: How urban school systems improve student achievement*. New York: MDRC.

DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Bloomington, IN: National Education Service.

Feiler, R., Heritage, M., & Gallimore, R. (2000, April). Teachers leading teachers. *Educational Leadership*, *57*(7), 66–69.

Heritage, H.M. (2003). *Web-based training for the Quality School Portfolio: Module 3*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Herman, J., & Gribbons, B. (2001). *Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation* (CSE Technical Report 535). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Herman, J.L., Webb, N., & Zuniga, S. (2003). *Alignment and college admissions: The match of expectations, assessments, and educator perspectives* (CSE Technical Report 593). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Jamentz, K. (2001). Beyond data-mania. *Leadership Magazine*, *31*(2). Retrieved August 24, 2004, from http://www.acsa.org/publications/pub_detail.cfm?leadershipPubID=1023

Jandris, T.P. (2002). *Data-based decision-making: Essentials for principals*. Alexandria, VA: National Association of Elementary School Principals.

Katz, S., Sutherland, S., & Earl, L. (2002). Developing an evaluation habit of mind. *The Canadian Journal of Program Evaluation*, *17*(2), 103–119.

Marzano, R.J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Mason, S.A. (2001, Spring). Turning data into knowledge: Lessons from six Milwaukee public schools. *Using Data for Educational Decision-Making*, *6*(1), 3–6.

National Research Council (NRC). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

No Child Left Behind. Archived information. (n.d.) Retrieved August 10, 2004, from http://www.ed.gov/nclb/overview/intro/presidentplan/proposal.pdf

Porter, A.C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*(7), 3–14.

Rothman, R., Slattery, J.B., Vranek, J.L., & Resnick, L.B. (2002, May). *Benchmarking and alignment of standards and testing* (CSE Technical Report 566). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Salinger, T. (2001). Assessing the literacy of young children: The case for multiple forms of evidence. In S.B. Neuman & D.K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 390–418). New York: Guilford Press.

Secada, W.G. (2001, Spring). From the director: Data/decision making. *Using Data for Educational Decision-Making*, *6*(1), 1–2.

Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14.

Shepard, L.A. (2004). Curricular coherence in assessment design. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability. 103rd yearbook of the*

National Society for the Study of Education, Part II (pp. 239–249). Chicago: National Society for the Study of Education.

Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: Council of the Great City Schools. Retrieved January 18, 2005, from http://www.cgcs.org/pdfs/Foundations.pdf

Stecher, B.M., & Barron, S. (2002). Unintended consequences of test-based accountability when testing in "milepost" grades. *Educational Assessment*, *7*(4), 259–282.

Stecher, B.M., & Hamilton, L.S. (2002). Putting theory to the test: Systems of "educational accountability" should be held accountable. *RAND Review*, *26*(1), 17–23.

Stiggins, R.J. (2002, June). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10), 758–765. Retrieved August 25, 2004, from http://www.pdkintl.org/kappan/k0206sti.htm

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer.

Thorn, C.A. (2001, November 19). Knowledge management for educational information systems: What is the state of the field? *Education Policy Analysis Archives*, *9*(47). Retrieved August 25, 2004, from http://epaa.asu.edu/epaa/v9n47/

Thorndike, E.L. (1922). *Intelligence tests and their use: The nature, history, and general principles of intelligence testing. The twenty-first yearbook of the National Society for the Study of Education*, Part I (pp. 1–9). Bloomington, IL: Public School Publishing Company.

Tognieri, W., & Anderson, S.E. (2003). *Beyond islands of excellence: What districts can do to improve instruction and achievement in all schools*. Washington, DC: Learning First Alliance. Retrieved from http://www.learningfirst.org/publications/districts/

Wayman, J.C., Stringfield, S., & Yakimowski, M. (2004). *Software enabling school improvement through analysis of student data* (Report 67). Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk (CRESPAR).

Webb, N.L. (1997). Determining alignment of expectations and assessments in mathematics and science education. *NISE Brief*, *1*(2). Retrieved from http://www.wcer.wisc.edu/nise/Publications/Briefs/Vol_1_No_2/

Webb, N.L. (1999). *Alignment of science and mathematics standards and assessments in four states*. (Research Monograph 18). Madison: University of Wisconsin-Madison, National Institute for Science Education.

Wilson, M. (Ed.) (2004). *Towards coherence between classroom assessment and accountability: 103rd yearbook of the National Society for the Study of Education*, Part II (pp. 1–19). Chicago: National Society for the Study of Education.