January 2005

center on ¦ **reinventing** ¦ public education

BURIED TREASURE

Developing a Management Guide From Mountains of School Data

Mary Beth Celio
and
James Harvey

UNIVERSITY of WASHINGTON
Daniel J. Evans
School of Public Affairs

"Using the guide, district analysts will be required to provide not simply what they can measure, but what the board and superintendent need to know."

"Buried Treasure," page 28.

Table of Contents

Executive Summary

This report provides a practical discussion of what is required to develop a school district "management guide," along with an actual guide built on evidence-based indicators. It begins with an imaginary discussion at Rebel Valley School District, during which a new superintendent leads his board through the guide.

Indicators are discrete pieces of information, like water temperature or the Dow Jones Average, designed to alert leaders and members of the public about what is going on in large, complex systems. They provide warnings and hints about how well complex systems are functioning. They are, therefore, capable of alerting leaders to potential problems. Although they can help identify problems, they cannot provide solutions.

To be effective, indicators need to be very powerful in terms of the quality of data, the utility of the information they provide, and their ability to communicate something important and meaningful.

The seven **indicators** of interest in the management system described in this report are:

- Achievement (reading and mathematics).
- Elimination of the achievement gap.
- Student attraction (school ability to attract students).
- Student engagement with the school.
- Student retention/completion.
- Teacher attraction and retention.
- Funding equity.

Several **implications** flow from the analysis contained in this report:

Less may be more. School systems are now awash in data and information. The human capacity to absorb information is of necessity limited. Indicator systems should respect that reality.

The principles of parsimony and power should be respected. The temptation to develop 17 indicators, or even 127 different pieces of information capable of satisfying everyone with a question about anything in every individual school, must be avoided The key to success will lie in parsimoniously selecting a few indicators and judging them against the standards of data, proxy, and communications power.

Current status data is necessary but not sufficient. Without a perspective grounded over time, the public may be confused by year-to-year pronouncements about how well (or poorly) things are going.

Smart use of data holds the potential of dramatically altering the tone and quality of board-superintendent relationships. Data sets that identify problems and promise to "get at" real issues on a school-by-school basis offer district leaders what all of them want—the opportunity to target scarce resources where they can do the most good.

Targeting resources where they can do the most good requires better funding indicators. If district leaders genuinely seek strategic use of limited resources, it is essential that they stop "resourcing" schools and start thinking about the real dollar amounts spent in each of them.

Currently, teacher attraction and retention are the best proxies we have for teacher effectiveness. Reliable indicators on teacher and principal quality are hard to come by. The lack of such indicators greatly hampers our ability to measure the impact of teachers and principals in any given school.

The seven school-level indicators discussed in this paper are a solid jumping off point for any district. Although these indicators are well grounded in research and experience, each district will have to decide for itself the extent to which any (or all) of these seven measures fit its particular needs and circumstances.

Professional development and technical assistance will be required. In recent years, leadership sophistication about data usage has increased dramatically. Still, effective use of data as a management tool will undoubtedly require additional professional development or technical assistance.

State leaders have a significant role to play. The role of state leaders becomes the role of leadership everywhere: pointing people in the right direction, providing political cover, and helping districts move along.

In many ways, indicator development moves beyond bottom-line assessment systems to encourage new ways of thinking about accountability, while doing the right thing and taking the time to do it right. The work outlined in this report suggests that educators and leaders can find a better way, and then provides a concrete example of how this might be done.

Preface

"Depend upon it, sir," said Samuel Johnson, one of the great wits of the eighteenth century, "when a man knows he is to be hanged in a fortnight, it concentrates his mind wonderfully." Johnson would have had no trouble understanding the recent response of school districts across the country to the many government carrots and sticks before them. The accountability provisions of enactments such as the No Child Left Behind Act (NCLB) have concentrated education's mind wonderfully.

Now that federal money is tied to "adequate yearly progress" and parents can transfer their children from a "failing" school, there is major focus on quantitative rating of school and district level success. This is revolutionary. Instead of concentrating primarily on the "inputs" into the school system (dollars, classroom size, classes offered, size of library, number and quality of teachers), school leaders are now focusing on the "outputs": whether students learn when they are in the classroom and whether they leave the classroom at the appropriate time and with the appropriate credentials in hand. These kinds of data are now being used to make judgments about the value of individual schools and districts, and there is substantial criticism of this new reality.

What is often overlooked is that the "outputs" can also be used by system leaders to make decisions about where they should intervene and what they should emphasize as they seek to assist struggling schools and students in them. In other words, information that some critics think is only used in an arbitrary way to judge schools can, instead, be used to help leaders make critical decisions about what to do and when to do it.

Accountability, in short, can be diagnostic and constructive; it does not have to be punitive or destructive.

One challenge is that school leaders are overwhelmed by educational minutiae in this new environment. There is probably treasure in there somewhere, but, buried as it is in mountains of ill-understood data, it is hard to discern the shape or potential value of these gems. While there is undoubted interest in gathering and reporting both input and output data on schools, much of what is on district websites and district-produced school guides is scattershot and unfocused. Numbers collected and presented in this way provide masses of data but little guidance about what the information means, how to use it, or what to do with it. The result, as one study of district data needs and uses puts it, is that "Most urban cities lack the strategic information to successfully identify and implement a district reform strategy."[1]

This report sets out to define and construct a working model of a school management guide that will help school officials, school leaders, and community members make sense of the mountains of data they are generating. It is not so much a map to buried treasure as it is a guide to action. It does not say "X marks the spot," but it does provide some general guidance on where to start digging and how to begin.

This guide grew from an extensive study of school district leaders and how they use data in making decisions. It is informed by analyses of how leaders in other areas of community life try to understand how their public institutions measure up against public institutions elsewhere. It is designed to assist leaders—of individual schools, school districts, and state education systems—improve school management. One way it attempts to do this is by describing how leaders can use key indicators to understand and use the data they already have available. Through carefully selected indicators, school leaders can examine how well their schools are doing in relation to comparable schools elsewhere, as well as how things are changing over time. The foundation of the guide is work completed over the last six years at the University of Washington's Center on Reinventing Public Education (CRPE), much of it supported by The Wallace Foundation. This work focused on the achievement gap as measured by Washington Assessment of Student Learning (WASL)—the state's standardized testing system, national and statewide dropout statistics, national studies of superintendents and principals, and extensive work on school and school district reform, including school finance.

At the heart of the management guide is a new educational indicator system designed to provide focus without adding new layers of data collection and analysis. Indicators are discrete pieces of information that tell us something about what is going on in a larger system. More technically, an indicator is "simply a set of rules of gathering and organizing data so they can be assigned meaning."[2] Indicators are often single items or indices of data that provide information about an underlying characteristic. The readings on automobile speedometers and gas gauges are indicators. A fever thermometer reading is an indicator. New factory orders and housing starts are indicators, as are rates of unemployment and hospital morbidity. The point is that whenever we are unable to view a large system in its totality—whether an automobile, the human body, the national economy, a local community, a school system, or a hospital—indicators can provide a general sense of how well the system is functioning. The trick is to find indicators that both have meaning and are easy to read. If the engine is overheating, the driver needs to be able to tell that at a glance.

In the case of education, the requirement that indicators have meaning and be easy to read pre-supposes that measures have been validated by research as related to student learning. Perhaps even more difficult where educators are concerned, it pre-supposes that indicators are capable of being presented

in a comprehensible graph or chart, or presented in a page or two at most, rather than in a volume.

Many school leaders have the skill to sift through reams of data to see how the system is performing (as a whole and in its parts), yet they seldom have the time or the training to do so. Those outside district leadership—principals, teachers, parents, and the public—generally have neither the time, the skill, nor the inclination. Two unattractive developments potentially arise from this state of affairs. The first is that, faced with piles of confusing and sometimes contradictory data, all shrug and throw up their hands. The second is that one or two discrete and readily understandable pieces of data will be seized upon as definitive evidence that the schools are "failing" or, alternatively, "turning the corner." The management guide presented here provides indicators that respect both the complexities of the institution they are describing and the subtleties of the data, while being meaningful and straightforward.

This report is the fifth developed at the Center on Reinventing Public Education for The Wallace Foundation. Earlier, the Center produced:

- *A Matter of Definition: Is There Truly a Shortage of School Principals?* (Marguerite Roza et al., January 2003.)
- *An Impossible Job?: The View from the Urban Superintendent's Chair.* (Howard Fuller et al., July 2003.)
- *Making Sense of Leading Schools: A Study of the School Principalship.* (Bradley Portin et al., September 2003.)
- *From Bystander to Ally: Transforming the District Human Resources Department.* (Christine Campbell et al., April 2004.)

Buried Treasure is divided into four chapters. Following the summary and this preface, the first chapter provides an example of how the management guide would look and could be used in a typical (i.e., highly complex) urban school district. It does so by reproducing an imaginary transcript of a school district board meeting where the management guide is introduced. The second chapter provides a short review of indicator systems, a description of the criteria used for selection of the indicators used here, and a commentary on the potential sources of the data. Chapter 3 provides a detailed explanation of each element in the guide, discussing the reasons for using the particular indicator and how it would be displayed for individual schools in a district. These charts and graphs could, potentially, be used as part of an individual school report card that is then aggregated into a district report card. Chapter 4 concludes with implications for state and district leaders.

In providing an example of how the guide would look and could be used in practice—at both the individual school and district levels—the guide is unique among indicator systems presently available. In most cases, the literature provides only prescription (i.e., telling leaders what data they should

gather and how it might be used) without actually spelling out practical implications. Alternatively, the literature provides only description (e.g., dashboards and other tools that provide shells for school or district data) without an explanation of why particular indicators are used or what they mean in practice.[3]

The actual management guide presented in the text (and the individual school reports that make it up) uses real school and school district data, although the data for the different schools that are described are a composite of different schools in several districts and a couple of states. They are representative of the data and school situations that exist in the real world, but the guide is not meant to describe an actual school or school district. The authors present actual school data in the management guide to illustrate how districts can use information they already have to focus attention either on particular schools or on cross-school trends that call for additional study and intervention.

The fact that the authors used data readily available was merely a function of time and convenience. Still, the ability to do so in this study has important implications across the United States. Any state or district with a standardized testing system and a minimum of key data points in hand should be able to use the approach outlined in this report. Large school systems with some analytical capability in-house can probably do so with minimal outside assistance. Smaller districts, with less internal analytical capability, may well need consulting help or technical assistance from the state.

---

[1] Marguerite Roza, "Chapter 8: Rethinking Data Capacity" in *Making School Reform Work: New Partnerships for Real Change*, Paul T. Hill and James Harvey eds. (Washington, D.C.: Brookings Institution Press, 2004, p. 105).

[2] Judith Eleanor Innes, *Knowledge and Public Policy: The Search for Meaningful Indicators*. 2nd expanded edition. (New Brunswick, NJ: Transaction Publishers, 1990, p. 5). Also, see Tyler Norris, Alan AtKisson et al., *Community Indicators Handbook: Measuring Progress Toward Healthy and Sustainable Communities.* (Oakland, CA: Redefining Progress, 1997) ; J. E. Innes and David E. Booher, "Indicators for Sustainable Communities: A Strategy Building on Complexity Theory and Distributed Intelligence" in *Planning Theory and Practice*, Vols. 1 and 2. (2000, 173-186).

[3] Some representative examples of proposed indicator systems in a number of different fields are found in the following: Jason M. Fields and Kristin E. Smith, *Poverty, Family Structure, and Child Well-Being: Indicators from the SIPP*. (US Census: Population Division Working Paper No. 23, 1998); National Civic League, *The Civic Index: Measuring Your Community's Civic Health*, 2nd. ed. (National Civic League, 1999); Eva Gold, Elaine Simon and Chris Brown, *Strong Neighborhoods Strong Schools: The Indicators Project on Education Organizing.* (Chicago: Cross City Campaign for Urban School Reform, March 2002); *Social and Health Indicators for King County: Selection of Core Indicators*, www.communitiescount.org/GreenReport.PDF. (King County, December 1998). See also: David S. Sawicki and P. Flynn, "Neighborhood indicators: a review of the literature and an assessment of conceptual and methodological issues," *Journal of the America Planning Association*, V62n2, pp 165-184. (Spring 1996); European Commission Directorate-General for Education and Culture, *European Report on the Quality of School Education: Sixteen Quality Indicators*. (Luxembourg: Office for Official Publications of the European Communities, 2001); Deborah Gibbs and Brett Brown, *Community-Level Indicators for Understanding Health and Human Services Issues,* http://aspe.hhs.gov/progsys/Community (Office of the Assistant Secretary for Planning and Evaluation, September 2000).

Chapter 1   The Superintendent Introduces
a New Management Guide

Most superintendents and school boards are trying to make decisions while buried in mounds of data that many have trouble interpreting. This creates a situation ripe for poor decisions, as policymakers focus on the dross of data instead of the nuggets within it. It is easy to seize on one or two pieces of information without fully understanding what lies behind them. But there's buried treasure within the mountain of numbers facing local leaders. The challenge lies in knowing where to dig.

Let's imagine how the situation might be improved by watching what happens when new superintendent Chris Hernandez arrives at mythical Rebel Valley School District and sets out to impose order on the information chaos he confronts.

●   ●   ●

When Chris Hernandez arrived for his first official day as superintendent in Rebel Valley, he was already aware that he was facing an all but impossible challenge[1]: overseeing 82 regular and alternative schools serving over 18,000 students of all types. Enrollment included a volatile urban mixture of children from highly educated professional families, low-income minority communities, and first-generation immigrant families from Eastern Europe, Mexico and the Philippines. There were state criterion-referenced tests given in six areas at three grades and increasing accountability pressures at the district, state, and federal levels. Bond measures had received mixed responses from voters in the previous four years: one up, two down. And the district had had two superintendents in the previous six years, both coming in with high expectations and both leaving (one with a contract buy-out) with most of those expectations unfulfilled.

The size and complexity of the challenge was made concrete for Hernandez when he found on his work table 82 three-inch binders, carefully and creatively prepared as a welcoming gift from the schools under his purview. Each contained the usual demographic information and test score results (in great detail), plus the kind of high-context material that could make a school come alive: examples of student work, pictures of the students at work and play, programs from school plays, and letters of support from parents and members of the community. In addition, there were several loose-leaf volumes of district and state assessment data compar-

ing individual schools with others within and outside the district. Chris was also presented with minutes from the meetings of the school board going back 10 years.

Hernandez was immediately reminded of his college-age daughter's response when he gave her a three-inch stack of printouts and reports on new cars when she was buying her first vehicle: "Too much data!!!" Every piece of information was important. Some pieces were critical. But there are only so many hours in the day and Chris faced an initial school board retreat in a month, where his new employers expected him to set directions for the future.

Unlike many superintendents in his position, Hernandez had a secret weapon: a former colleague, now at the local university, who had been working for years trying to make sense of the data overload facing leaders in individual schools and school systems. The colleague had rashly claimed that looking at school district data from a "portfolio management" perspective could help superintendents make sense of the data they had. Chris immediately offered her this challenge: "Help me make sense of these 82+ binders so I can get off to the right start. And do it so that a Ph.D. isn't required to understand the results or maintain the system." He was hoping, of course, for the educational equivalent of a single-number credit rating or school GPA, a summary measure that would say that the Michael Collins High School was a B+ (3.21 on a 4.00 scale), while DeValera was a D (1.42).

What Hernandez received was somewhat more complex than that, but much more helpful. In a month, he was ready to go to the board retreat with less than 100 pieces of paper or, for those who preferred laptops, three introductory screens and their links.

Indicators as a Management Guide

The heart of the management guide that Chris Hernandez provides to the Rebel Valley school board is built around emerging theory and practice regarding "social indicators." As described in Chapter 2, indicators are discrete pieces of information, like water temperature or the Dow Jones Average, designed to alert leaders and members of the public to what is going on in large, complex systems. To be effective, indicators need to be very powerful in terms of the quality of the data, the value of the information they provide, and their ability to communicate something important and meaningful.

The seven **indicators** of interest in the management system described in this report are:

- **Achievement** (reading and mathematics)
- **Elimination of the achievement gap**
- **Student attraction** (school ability to attract students)
- **Student engagement with the school**
- **Student retention/completion by school level**
- **Teacher attraction and retention**
- **Funding equity**

Hernandez uses several of these indicators in his discussion with the school board. All are more fully described in Chapter 3.

Below is a transcript of his presentation and the response it received from school board members.

Superintendent Hernandez: I'm going to suggest that what we examine during this retreat is a snapshot of what all of our

schools look like right now from a management point of view. We are going to focus on where the schools are relative to each other, relative to the district, relative to the state, and relative to our standards. Before we turn to the data, I want to say that what I am going to show you is organized around the kind of information that can really help us. But it is not a silver bullet, something that can lay out a work plan or do our work for us. I won't call it a report card because that word has been so overused. What I want to consider today is what I think of as a management tool, something we can use to provide leadership for our schools.

What I am going to show you is a picture of the schools for which we are responsible: how they are doing now in key areas and how they have changed over the past five years. I'm going to first give you an overview and then focus on a couple of middle schools to illustrate how we can use this tool.

Board Chairwoman Anne Barnes: Superintendent, you probably know that we have already had some bitter fights about labeling schools as failing on the basis of a random test score for a single year. We are dealing with a community that seems to be losing faith in our public schools. We really don't want another downer.

Superintendent Hernandez: I couldn't agree with you more, Anne. As a teacher and a principal I learned how unfair it was to blame everything on the school or the teacher without seeing the whole picture. I know how a single incident or test report can focus negative attention on a school. For this reason, the tool I'm going to show you uses a variety of indicators that research has shown to be directly related to student achievement. No single indicator is enough, but all together can provide a picture that we, as leaders, can use to decide where we should focus attention and resources. I want us to think like system managers, looking at the whole picture and trying to decide how we can manage the portfolio of resources available to us. Those resources include families and teachers and principals and school buildings and programs.

I'm certainly not saying that the indicators I'm presenting today are the only ones or even the best ones, but they are the ones that research appears to support at this time. I have prepared a more technical report that explains the history and development of each indicator, and that should answer many of your more detailed questions. But right now, I'd like to go through this exercise and encourage you to challenge me at any point. This is our work in progress.

Board Chairwoman Barnes: I'm going to reserve judgment until I've seen what you have, but I'm certainly happy to hear that you aren't planning to impose something we aren't comfortable with. I, for one, am tired of having other people's ideas foisted on our district.

Superintendent Hernandez: The first chart I'm going to show you will remind you of the kind of *Consumer Reports* charts you are probably all familiar with. Let's look at Figure 1.1. Across the top of this chart you find categories of schools we will be considering, in this case our district's three school grade spans. These are compared to all schools in the state at the same grade spans, as well as with schools in urban districts most similar to Rebel Valley. Down the side are the indicators I mentioned. On the upper half of the left-hand side are the status indicators. These items answer the question: "Where are we now?" These rows contain the most recent information we have about the schools in terms of test scores, the achievement gap, and so forth. On the bottom half are change indicators that answer the question "How have things changed?" These indicators look at things like test scores and the achievement gap, and indicate how they've changed over the past five years.

On this chart, grey is used to indicate that our schools are doing better than other schools in the state. Black is used to indicate that our schools are not doing as well as other schools. Solid black circles indicate one extreme, considerably worse than other schools—while donut grey circles indicate the other extreme, considerably beter. The empty circles mean that the average scores of our schools on that measure are within the middle third of scores in the comparison group as a whole. In other words, the students in these schools earned scores in a range of about 15% above or below the mean. As I said, they're in the middle. And, in fact, if you look at Figure 1.1, that's where our high schools are in math. In terms of math achievement, our high schools are about in the middle of the pack in the entire state. I think you've probably had a sense of that, but I'm not sure the data have ever been boiled down this way before.

If you look down that same high school column, you'll see a half-grey circle for "Elimination of achievement gap." That's pretty good. We should all be happy with that. That means that our high schools are doing better than about two-thirds of the schools in the state in closing the achievement gap between white and minority students. Even more exciting is the donut shaped grey circle in the elementary school column. It shows that our elementary

Figure 1.1
District and State Comparison Chart

Legend for symbols used in the table:
- ● = In bottom 10% of comparison group
- ◒ = In bottom third of comparison group, but > bottom 10% (circle, bottom half black)
- ○ = Within 15% (+/-) of comparison group
- ◓ = In top third of comparison group, but < top 10% (circle, top half gray)
- ◉ = In top 10% of comparison group
- ∗ = Not available for comparison group

Indicators			Rebel Valley Elementary Schools		Rebel Valley Middle Schools		Rebel Valley High Schools	
			Compared to state	Compared to other urban districts	Compared to state	Compared to other urban districts	Compared to state	Compared to other urban districts
Status	Achievement	Math	◒	◒	●	○	○	◒
		Reading	◒	◒	●	●	◒	◓
	Elimination of achievement gap	Math	○	◉	◓	○	◓	◓
		Reading	○	◓	◓	◓	◓	○
	Student attraction		○	○	◒	◒	○	○
	Engagement with school		○	○	◒	◒	○	○
	Student retention/completion		◓	◓	◒	◉	◒	○
	Teacher attraction and retention		◒	◒	●	○	○	◒
	Funding equity		∗	∗	∗	∗	∗	∗
Change	Achievement, change from 1999	Math	◓	◓	◒	○	◒	◓
		Reading	◓	◓	◒	○	◓	◓
	Reduction in achievement gap, change from 1999	Math	◒	○	◒	○	◒	○
		Reading	○	○	◒	○	○	○
	Student attraction, change from 1999		○	○	◓	◓	◒	○
	Student retention/completion, change from 1999		◒	◓	◒	◒	○	◓
	Engagement with school, change from 1999		◒	◓	◒	○	◒	◓
	Teacher attraction and retention, change from 1999		●	◓	◓	◉	○	◓
	Funding equity, change from 1999		∗	∗	∗	∗	∗	∗

Worse ← ● ◒ ○ ◓ ◉ → Better

● = In bottom 10% of comparison group
◒ = In bottom third of comparison group, but > bottom 10%
○ = Within 15% (+/-) of comparison group
◓ = In top third of comparison group, but < top 10%
◉ = In top 10% of comparison group
∗ = Not available for comparison group

schools are doing better than about 90% of the schools in the state in eliminating the achievement gap. That's a pretty phenomenal accomplishment. The board and the community have been doing a lot of good things at the elementary level.

Unfortunately, we also have some black circles. The black circles are at the other end of the scale. A half black circle, like the one for elementary math achievement, means that the math scores in our elementary schools are in the bottom third of the scores across the state—two thirds or more of the schools in the state get better average math scores than our elementary schools. In a few cases, the comparison is even more serious. Middle schools seem to be our real challenge. Our middle schools fall in the lowest 10% of the scores in the state in both math and reading.[2]

Board Member Carl Chipson: One question before you continue, Chris. The state and now the US Department of Education are requiring us to report scores in several different areas for three different grades. Why are you only showing us the reading and math achievement?

Superintendent Hernandez: First of all, we need to remind ourselves that we aren't replacing the state or national accountability requirements here. We might want to talk about what that would mean at some point, but it's not what this is about.

We are trying to use the data we have to get a handle on the challenges in front of us. One of the reasons some indicator systems failed in the past, before they were even given a chance, was that they tried to put all possible data onto the table.[3] We all tend to get into overload mode immediately when this happens.

To avoid burying us in data, this system focuses on a few key indicators. The research shows that math and reading are key subjects; we can know a great deal about a school or a student by knowing those two scores. We can't know everything, of course, but we can get a pretty strong indication of what is going on by knowing those two. However, I'm open to suggestions here. If we find that these two areas aren't enough or aren't the right ones, we can certainly look into alternatives.

I'm not going to go into detail on each of these indicators yet, but I want to point out something right away. I think we can be proud that our schools match up very well with the schools across the state and with other urban systems in a number of areas. The achievement gap between white and minority students shown in

test scores is at or below the levels found in other urban districts across the state. That's something very positive to build on. We also seem to be doing well in retaining our students. However, there's always another shoe. You can also see that there is a concentration of black circles in the reading achievement row and in teacher retention. These indicators can't tell us what is going on in these areas or what to do about it, but these may be red flag areas, areas to which we want to pay particular attention this year. This chart can provide a wake-up call in these areas.

Board Member Chipson: This looks pretty helpful. It doesn't tell us what to do, but I can see going into my community and explaining why we're doing certain things with the help of a figure like this.

Superintendent Hernandez: Yes. That's certainly one way this can be used. But there are a lot of others—and a lot of other useful information. Let's look a little more closely at Figure 1.1.

What I want you to examine particularly are the groups of schools organized by grade span. As you can see, the middle schools stand out from the elementary and high schools because both the status and the change indicators tend more toward black than grey. That's not good. It's for this reason that I'm going to use our middle schools to demonstrate how this management tool can be used here in Rebel Valley. Thus, although there are areas across all grade spans that we might want to focus on this year and next—areas like teacher attraction and retention—for now I would like to look at the middle schools so I can walk you through the process.

So, let's turn to Figure 1.2.

Figure 1.2
Rebel Valley Middle Schools Comparison Chart

	Indicators		Middle Schools					
			Guy Fawkes	D.B. Cooper	Monmouth	Troy Memorial	Edsel United	Crispus Atticus
Status	Achievement	Math	◑	◐	●	○	○	◐
		Reading	◑	○	●	○	◐	◉
	Elimination of achievement gap	Math	◉	◐	●	○	○	○
		Reading	◑	◐	●	○	○	○
	Student attraction		○	◐	◐	◑	●	◉
	Engagement with school		◑	○	○	◐	◉	◑
	Student retention/completion		◉	◐	○	○	●	○
	Teacher attraction and retention		◑	○	○	◑	●	◐
	Funding equity		○	◉	○	◉	○	●
Change	Achievement, change from 1999	Math	◐	◉	●	○	◐	○
		Reading	○	◑	◐	○	●	◉
	Reduction in achievement gap, change from 1999	Math	◐	◑	○	●	○	○
		Reading	●	○	◐	○	○	◉
	Student attraction, change from 1999		◉	◑	○	●	○	○
	Student retention/completion, change from 1999		◐	◑	●	○	○	○
	Engagement with school, change from 1999		○	◉	◐	◑	○	●
	Teacher attraction and retention, change from 1999		◐	◑	○	○	◉	●
	Funding equity, change from 1999		✳	✳	✳	✳	✳	✳

● ◐ ○ ◑ ◉

Worse ⬅———➡ **Better**

● = In bottom 10% of comparison group

◐ = In bottom third of comparison group, but > bottom 10%

○ = Within 15% (+/-) of comparison group

◑ = In top third of comparison group, but < top 10%

◉ = In top 10% of comparison group

✳ = Not available for comparison group

Superintendent Hernandez: I'd like to start by focusing on Guy Fawkes Middle School. If you look only at the first half of the chart, the status indicators, everything looks fine. Most of the circles are grey, in whole or in part, indicating that the achievement gap is below the district average, and that the school is attracting both students and teachers. In fact, scanning the status of the other five traditional middle schools, it looks as if Guy Fawkes is doing a good job overall.

But the change indicators in the bottom half of the chart tell a different story. This is a school that might be losing the confidence of teachers and parents. Even if it looks OK today, it is a school that may be entering a period of trouble. There has been less improvement in all measures than is true in the rest of the district and the number of applications from both students and teachers has dropped.

Board Chairwoman Barnes: Wow, hold on. Are you sure we have that right? I'm surprised by these black circles. You must know that Guy Fawkes is one of our most successful schools. It's in one of the nicest areas of the district. Why, Guy Fawkes was even given a national award recently! I was lucky enough to be there at the ceremony with the Assistant Secretary of Education.

Superintendent Hernandez: Good point, Anne. This surprised me, too, when I first saw it, so I'm glad you asked about this. But I think we got it right; there's nothing wrong with the data.

The management guide can't tell us what's going on below the surface, but it can give us a "heads up" to look more closely at a particular school or a particular program. Each of the rows on the chart corresponds to a set of data and graphs that can provide more information on what's going on in the individual school. I'm now going to project on the screen the backup data (Figure 1.3) on current math and reading achievement for Guy Fawkes. The summary chart was based on these data.

As you saw in the status achievement data, test scores at Fawkes are at the upper end of the district ranges in both reading and math What this graph shows in addition, however, is how the actual scale scores are distributed compared to the district as a whole.

This figure demonstrates an interesting and not altogether unusual pattern. As you can see, last year about 20% of the students from Guy Fawkes who took the 7th grade math and reading tests scored in the top 10% of the district scores. This helps account for the

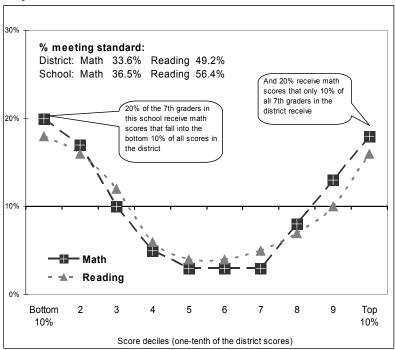fact that the school has higher overall achievement scores than other schools in the district.

Figure 1.3
Guy Fawkes Status: Achievement, 2003

But in a lot of ways, Guy Fawkes' reputation rests on the achievement of its best students. They bring up the average. You'll also notice the odd "U" shape to the distribution of scores. If students at Guy Fawkes earned scores in the same pattern as the district as a whole, the dotted and dashed lines representing math and reading would be approximately straight lines and overlapping the black district line in several places. Instead, it appears that although there are quite a number of students at Fawkes who do very well, an equally large number aren't even close to meeting the standards. At the left side of the graph, you can see that almost 20% of Guy Fawkes students fall into the bottom tenth of scores received across the district.

Something is going on here that needs to be looked at more closely if we are to know how to assist this school.

Member Chipson: We all know that every class has some really smart kids and some kids who struggle. That may have happened last year at Fawkes. Is this pattern somehow different from other

schools? You know, I don't want to go out into the community and start planting doubts about a school that's receiving national awards.

Superintendent Hernandez: We don't have time right now to go into the mathematical details, but we'd usually expect the math and reading lines for a particular school to be relatively close to the district average. That's what you'll see if you look at the backup data and graphs for most of the other middle schools. Here at Guy Fawkes, more students than normal are at the polar ends of the scale: doing very well or very poorly. It is the odd shape of the distribution that should alert us to a possible problem. The next chart for Guy Fawkes, the one about the achievement gap, may be able to provide some hint about where the problem may lie.

Figure 1.4 is what I call a "gap figure." In this case, the gap is about student achievement, precisely what we are all worried about in terms of state assessment and No Child Left Behind.

Figure 1.4
Guy Fawkes Status: Achievement Gap, 2003

In Figure 1.4, the distance between the black line and the dotted and dashed lines represents the difference between the distribution of the scores earned by white students at Guy Fawkes and

the scores earned by minority race students. It helps you visualize the nature of the gap.

Here you can see that the non-white scores are concentrated in the bottom half of the distribution of white scores in the school. About 40% of the minority students get the same low scores as 20% of the white students. Thus, while in reading there are a number of minority students in the upper levels of the distribution of scores, minority students tend to "top off" in math at a point where15% of white students are doing math more proficiently.

Figure 1.4 gives us very useful information. It reveals graphically that if we are to close the achievement gap, we can't do it by focusing only on those students who fall just below the cutoff point on the state assessments. That strategy will look good in the short run, but to genuinely close the gap, we're going to have to focus like lasers on the students who are falling into the lowest achievement groups.

Although these data don't prove conclusively that the odd achievement pattern we saw in the last graph was caused by a racial achievement gap, it is probable that this is what is happening. More research will need to be done at Guy Fawkes to see what is going on, but we can see some clues here that can direct our search.

Member Rostowloski: But that's just 2003. Couldn't this just be that particular class? We all know that some classes are quite different from others. I'd hate to see a school be targeted because of the peculiarities of a single group of kids.

Superintendent Hernandez: That's why we need to look at trends and not just a one-year snapshot of the school. In fact, the trend data reveal that this is probably not a one-year phenomenon. They show some troublesome changes at Fawkes . Look at Figure 1.5. The chart shown on the screen right now displays the change in achievement over the past five years at Guy Fawkes. The lines display the change in the average reading and math scale scores from a "0" point of 1999.
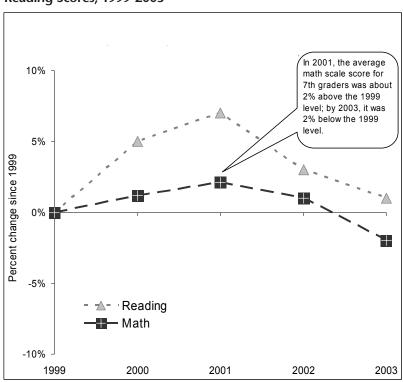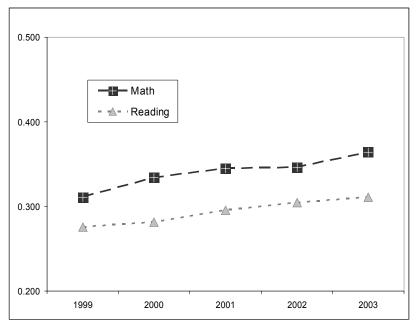
Figure 1.5
Guy Fawkes Trends: 7th Grade Standardized Math and
Reading Scores, 1999-2003

What this figure shows is that both the reading and the math scores increased by between 3 and 5% from 1999 levels up to 2001 but have dropped gradually since then. Thus, although school scores are still above the district averages, there appears to be a negative trend. There are a number of reasons why this might happen and these data can't tell us anything for sure about those reasons. It could be a demographic change, a new curriculum, a change in faculty composition or orientation, or something else entirely. All we know from this is that something is happening that is negatively influencing the achievement of children in the school. It warrants attention.

Board Chairwoman Barnes: What happened to the achievement gap during those same years? I noticed in the overall chart that Fawkes didn't measure up well against other schools where change in the achievement gap was concerned. Can you tell us anything about that?

Superintendent Hernandez: Luckily, there is a way to summarize the achievement gap so we can illustrate changes over time. I won't go into the technicalities here, but Figure 1.6 shows the changes in the gap "index" between 1999 and 2003. In this chart, the higher the index number, the greater the gap. And, unfortunately, the gap increased at Guy Fawkes Middle School between 1999 and 2003, both in math and reading.

Figure 1.6
Guy Fawkes Trends: Gap Index, 1999-2003

Member Dan Rostowloski: I may start sounding like a broken record here, and I can see that test scores are dropping somewhat and the gap is increasing somewhat, but these tests aren't being given to the same students each year. Couldn't the trends we see here be a result of demographic changes in the school's population? You might have real changes in the school population that can account for the apparent drop, and we might be reading more into the results than the reality justifies.

Superintendent Hernandez: You're right, Dan. This is one of the biggest drawbacks of the "snapshot" testing most states do, where we see our students' achievement at only one point and don't follow the same kids over time. Obviously, if we decided that this school warrants more attention from our oversight team, there would need to be particular attention directed at this issue. The drop may be due to a demographic change in the school

population linked to employment or housing changes in the area. Of course, understanding the gap doesn't justify the gap. Our legal and moral obligation is to eliminate it. Whatever lies behind the gap, it bears watching to assure that we are providing the school with the resources it needs to make sure all of the students are achieving.

Member Rostowloski: Whatever is happening, my sense from talking with families in the community is that it appears something is affecting the ways families are choosing this school.

Superintendent Hernandez: You've provided a perfect segue into the next indicators, Dan. Figures 1.7 and 1.8 are charts on attraction. They indicate that Guy Fawkes still attracts a proportionate number of students and teacher applicants. Here I'd like you to look at the charts showing the changes in attraction over time at Guy Fawkes for both students and teachers. You can see that the trend lines for the school are somewhat negative, indicating that there may be something going on that is causing parents and teachers to ask hard questions about the school. Since the district levels have remained much the same over the past five years, it is likely that the drops in attraction are due to something specific to the school rather than demographic or policy changes affecting the district as a whole.

Figure 1.7
Guy Fawkes Change: Percent of Students Choosing School as First Choice, 1999-2003
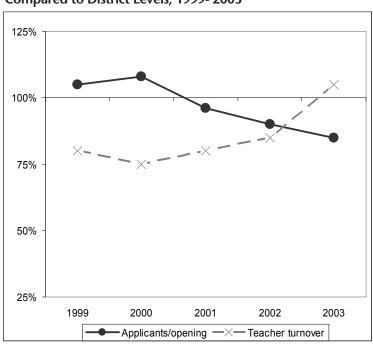
Figure 1.8
Guy Fawkes Change: Teacher Applicants and Turnover Compared to District Levels, 1999- 2003

Member Chipson: Chris, you've shown us several outcome measures like achievement, but you haven't said much about the resources available to the school. Isn't it important to know about the quality of the teachers and other things that might account for what we are seeing in these charts and graphs?

Superintendent Hernandez: As you know from your experience on the board, Carl, the quality of leadership and faculty in the school are key to the success of the students there. We are hampered, however, in not having a reliable measure of teacher or principal quality. Years of training or service aren't proven indicators of quality, and most of the available measures of teacher or principal effectiveness are based on the test scores earned by the students in a given classroom or school and not on any personal attributes we can track.
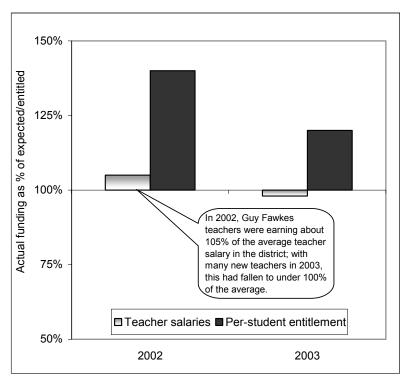
In preparing this overview of schools, we have tried to use only indicators that have research to support their use. There are a number of research efforts underway right now trying to determine what resources and what level of support are necessary to provide a quality education. There is certainly no consensus right

now on an overall measure of funding equity. However, I know from reading the minutes of Rebel Valley's school board meetings over the past several years that, as a board, you recognized the importance of looking at the resources available to the schools. Wisely, you also asked the university to help set up a system of tracking the data.

That's all been very helpful. My staff and I have used these data to develop two measures of equity. You can see both in Figure 1.9 for both Guy Fawkes and Monmouth. The first column in each figure addresses teacher salaries, while the second is a measure of how a school's funding compares to what it should be when we consider the students in that particular school.

For teacher salaries, the base line is the amount the district has been budgeting for that school based on the number of teachers assigned to the school and the overall average district teacher salary. As you probably realize, some teachers within any given school may be earning at the bottom of the pay scale, while others may actually have topped out.

Figure 1.9A
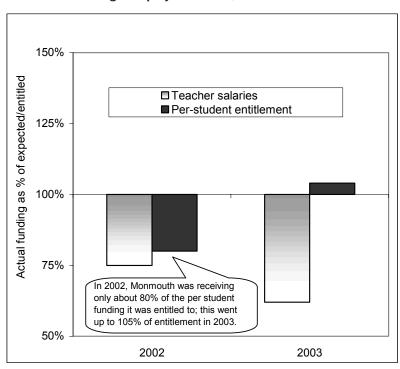**Guy Fawkes Change: Equity Measures, 2002-2003**

Figure 1.9B
Monmouth Change: Equity Measures, 2002-2003

We looked at each individual school's situation: how much has been budgeted vs. how much is actually spent in the school. You may be surprised to hear that the real salary expenditures in some schools are over 50% higher than the district average, while in others the salary expenditures are less than 60% of the district average. These are enormous differences. Although the salary paid to a teacher is not a fair indicator of his or her capability, it can certainly be used as an indicator of the teacher's experience and training. As with the other indicators, this one isn't perfect, but it can provide insights from which we can work to address disparities.

The second index, based on per-student funding, looks at what a school should be receiving if funds were allocated to the school according to the purpose of the funding, like Title I or special education, or the values established by the community. Rebel Valley has been ahead of most school districts in the country in establishing a policy of "student-based budgeting." Thus, support is attached to students rather than to buildings. A school that serves a larger number of low-income, educationally vulnerable students is supposed to receive more funding than one that serves children of upper-income, professional families because the former group of students needs more assistance than the latter.

The data gathered by the university last year were used to determine how much funding each school in the district should receive based on the students it serves. We then looked at the actual amount the school received last year in terms of teacher salaries, compensatory education, bilingual and gifted/talented resources, and so forth.

Ladies and gentlemen, what we have here is disquieting information about inequities in school funding. Some of the schools received less than 70% of what they were entitled to, while others received as much as 400% of their entitlement. Guy Fawkes is one of those schools that has received more money for teachers and for other student expenditures than average for the district. Other schools, like Monmouth, have received much less.

Member Sharron Tanner: I can see what you mean about possible inequities in funding, but I do want to make sure you understand that none of these apparent discrepancies between entitlement and actual spending were intended.

Superintendent Hernandez: I am certain of that. I also know, from the minutes of the meetings, that when these data were first presented to you, warning bells went off in all of your heads.

Many of the accounting practices in school districts across the country were inherited from different times and we may need to be more conscious of how old practices of average-salary budgeting might result in some schools having all new (and thus less expensive) teachers and some having all more experienced (and more expensive) faculty. The numbers of teachers might be the same in each school, but the use of average teacher salary to determine the per-student cost hides the inequities. Right now, it is important to look at the index as one more indicator of the health of the individual schools.

Member Chipson: Have things gotten any better on this measure? I know we just got the information last year and just started to address the issues before you were hired. Any indications of improvement?

Superintendent Hernandez: You may have noticed in Figure 1.1 that there were asterisks in the row next to the funding equity trends and the same thing for comparison to the state as a whole and other urban districts. That's because, as you noted, we have only begun to look at this problem and only a few other districts in the state have started doing this. As we get more information

each year, we'll be able to show changes over time. Meanwhile, this gives us a starting point for our own district. We also hope that other districts will begin paying attention to this challenge so we can have comparative data.

Member Chipson: I can see why many of the findings you've discussed so far might provide an early warning signal about Guy Fawkes, but what about Monmouth Middle? All the status indicators are black!! Shouldn't we be concentrating some attention there?

Superintendent Hernandez: No question about it. I'm just going to spend a couple of minutes on Monmouth's detailed backup information so you can get an idea of the flexibility of the indicators we're using. We'll be looking at each school in more detail later in this retreat.

Let's look back at Figure 1.2. For now, you can see that there are few bright signs in either the summary profile or the backup information for Monmouth. This school appears to have settled to the bottom of the district, with few signs of hope. This may be one of those triage situations where we need to authorize a thorough look at the culture and resources of the school and make a decision fairly quickly.

There could be any number of reasons for the profile we see here, but it is clear that the children at Monmouth are not succeeding at the school and are not being adequately prepared to take on the next level of education. This is a school that may warrant a specific improvement strategy or the introduction of a new model or approach to the problems of the students. Unlike Fawkes, where things haven't reached a critical point, at Monmouth we should probably consider ways to intervene now. But we can't ignore Fawkes while we salvage Monmouth.

Member Wilfred Cleveland: It's very interesting to see how Guy Fawkes and Monmouth face different challenges and probably need different approaches. But they have quite distinctive patterns. Are there other patterns? What do other schools look like?

Superintendent Hernandez: Just to provide an illustration of a different pattern we might face and to show that not all changes are negative, I'm now going to briefly highlight a third school that presents a profile different from either Fawkes or Monmouth. Let's look at D.B. Cooper Middle School in Figure 1.2.

In terms of status, Cooper looks worse than Fawkes: all of the test score measures and the achievement gap are more negative than the district as a whole. In addition, both attraction and retention are, at best, right at the district average.

However, unlike Fawkes, D.B. Cooper appears to be a school on the rise. When there have been job openings, they seem to be getting more applicants than other schools. Parent confidence also appears to be growing. Something positive is going on here that teachers and parents have heard about. We should watch the achievement figures over the next few years, but for now it is a school that needs to be supported on its way. Fawkes, on the other hand may be a school going down. Left alone it may get worse. That's a school we may need to look at more closely right away.

Member Cleveland: For years we have been hearing complaints about Cooper. I think we've talked about Cooper several times in previous meetings.

Member Chipson: Yes, but remember in 2001 we suggested that the superintendent transfer Ms. Kanasaka there. She had been sent into other troubled schools in the past and had a record for making positive changes and attracting good teachers. Maybe this strategy is working.

Superintendent Hernandez: As you can see, we have just touched the surface in looking at a single grade span using this management tool. It is clear that as managers we have a portfolio of quite different middle schools. Each requires a particular approach. One or two might need immediate study and intervention; some will warrant careful study over a period of time, with no great sense of urgency. Finally, one or two may simply need congratulations and continued encouragement to do what they're doing already. A district-wide initiative like professional development for all teachers or even salary increases or class size reductions everywhere would not respond to the patterns we see on this simple chart. There is more to know about schools and the district, but I wanted you to see how the right information can focus us on the right questions. This is just a first step in our work together.

● ● ●

This initial school board retreat in a mythical school district did not end with any decisions about where to go in the district or what to do about individual schools and their challenges. However, the members found a starting place based on both the current situation and the trends in the schools.

The board members now know that there are a group of schools with particularly difficult and widespread challenges (the middle schools) and that a couple of those school merit immediate and concerted attention. They also know that a couple of district-wide issues may need to be addressed fairly soon because district trends at the middle school level are declining relative to districts of the same kind in the rest of the state (Figure 1.1). The new management guide doesn't tell the superintendent or school board members what to do about the district's inability to attract enough teachers, but it does alert them to the fact that this is a problem they need to address. Decisions must be made. Action is essential.

The good news is that the district's leadership team now possesses a management guide to help map out the road ahead. The information overload represented by dozens of three-inch binders and reams of computer printouts is reduced to a handful of tables. The guide offers a place to start.

The guide offers something else as well. With these indicators as their map, district leaders have been empowered to define what is important. Using the guide, district analysts will be required to provide not simply what they are able to measure, but what the board and superintendent need to know.

---

[1] Howard L. Fuller with Christine Campbell, Mary Beth Celio, James Harvey, John Immerwahr, and Abigail Winger, *An Impossible Job? The View from the Urban Superintendent's Chair* (Seattle, WA: Center on Reinventing Public Education, 2003).

[2] The divisions in this and later indicators are based on standard deviations and effect sizes. A further discussion of this is provided in Chapter 3.

[3] Detailed explanations of each of the status and change indicators and how they are calculated and displayed are provided in Chapter 3 of this report and a discussion of their implications is provided in Chapter 4.

Chapter 2

A Short History of Indicators, the Foundation of the Management Guide

The management guide used in the mythical Rebel Valley School District is made up of seven indicators looked at in terms of status (what the individual schools and the district look like right now on this indicator) and trends (how the schools have changed over time in terms of the given indicator).

The indicators themselves were selected to act as discrete pieces of information because they tell leaders something about what is going on in the larger system. No single indicator, nor all combined, can provide all the information that might be necessary to understand a system. That would take masses of data and intensive study. Even then, a coherent picture might not emerge. The point of indicators is to get a fix on the larger system. Whenever we are unable to view a large system in its totality — whether an automobile, the human body, the national economy, a local community, a school system, or a hospital — indicators can provide a general sense of how well the system is functioning.

The need for indicators in school systems is highlighted by recent experience with the No Child Left Behind statute. In 2001 this legislation imposed a national definition of a set of indicators on American schools. The NCLB definition rests on test results for all students in 3rd through 10th grades in reading and mathematics, combined with development of indices of "adequate yearly progress" disaggregated by race and ethnicity.[1] But the reams of paper involving dozens of comparisons generated by these requirements confound the use of these data as "indicators." The data are likely to overwhelm the recipient, while conceivably balkanizing schools. If the point is to understand system performance, what is one to do with data indicating that 3rd grade reading scores are going up while 4th grade scores are declining in one school, and that the picture is reversed in a school in the next neighborhood? Or that one school is experiencing what appears to be a mass exodus while another, a mile away, is steadily gaining enrollment? There must be a way to assess how the system and individual schools in it are performing without drowning districts in meaningless detail.

The unacknowledged goal of all contemporary attempts to build indicators or indicator systems is to reproduce the effectiveness and utility of the Cost of Living Index, a national indicator that has become so deeply embedded in the national psyche that it is almost never explained or justified. This exemplar of all indicators was developed in 1963 by Mollie Orshansky, a modest, middle-aged statistician in the Social Security Administration. As the food stamp program developed, she set out to answer how much it

cost a family of four to feed itself a minimally nutritious diet. She used that piece of information (based on Department of Agriculture estimates of a low-cost adequate diet) to develop a cost-of-living estimate for families of different sizes and composition. The latter calculation was based on another Agriculture survey revealing that families of three or more persons spent about one third of their after-tax income on food in 1955. For her efforts, Ms. Orshansky received the federal Distinguished Service Award. She had created the first nationally accepted measure of income adequacy and applied it to public policy.

Orshansky differentiated her thresholds not only by family size but also by farm/nonfarm status, by the sex of the family head, by the number of family members who were children, and by aged/non-aged status (for one- and two-person families). The result was a detailed matrix of 124 poverty thresholds. (The figures generally cited are weighted average thresholds by family size.) Although her poverty thresholds had been calculated on the basis of after-tax income, they were applied to before-tax income because that was the only data set on hand. Orshansky was aware of the inconsistency and reasoned that the result yielded "a conservative underestimate" of poverty.[2]

The "Orshansky Index" is one of the most powerful social indicators ever developed. Many educators are unaware of Orshansky, but her index underlies the formula distributing funds under Title I of the Elementary and Secondary Education Act (now known as No Child Left Behind). The Orshansky Index permitted policymakers for the first time to make accurate estimates of the number of families, adults and children living in poverty within their jurisdictions. It also arrived at a time when the federal government was launching a "war on poverty," and was looking for a measure to allocate funds where they were most needed, and would, presumably, do the most good. Her invention provided a reasonable sense of the scale and scope of poverty in the United States and within individual states and communities. Its value is that even non-experts could understand the numbers the index produced, even if they knew nothing about the Department of Agriculture's four different food plans or its 1955 Household Food Consumption Survey—much less Orshansky's reservations about the use of before-tax income.

The same utility can be seen in other widely used indicators developed over the years:

- Almost nobody can define what is in the "basket of goods" that the Bureau of Labor Statistics has used for years to measure the Consumer Price Index, but practically every adult in the United States understands the "inflation rates" and cost-of-living indices developed from the CPI, and their utility.
- The Dow Jones Average is an indicator that is followed religiously by millions every day as they track their portfolios, but

most would be hard pressed to identify any of the 30 companies whose stocks make up the average.

- Although many governments pay attention to indicators of human rights contained in periodic UN assessments, few citizens understand that these indicators incorporate assessments of practices related to trade, sweatshops, child labor, and routine denial of voting and civil rights, including torture and abuse.

Although these indicators are far from perfect, the best of them are highly credible and have demonstrated their reliability over decades as a reasonably reliable picture of what is going on.[3]

Because of the value of national and international indicators, many communities have tried to develop indicators related to local quality of life. This is likely to be a trend that accelerates as magazines (such as *Money*) and television networks (such as CNN) develop issues and story lines oriented around lists of the "most livable cities" or "best places to raise children." The ranking of colleges and universities conducted annually by *U.S. News and World Report* is another example of how indicators are used to make highly influential judgments about the quality of an institution within a community. In some cases, the indicators used to come up with the final "best" and "worst" lists are not clearly identified and, when they are, they can be controversial. In all cases, however, the indicators exist. The magazine or network can and often does provide the details on how the ranking was done, but it is clear that the goal is to make headlines—and hence attract viewers or readers.

For communities and institutions wishing to identify quality indicators, however, the goal is not publicity but to know where they are and where they should be going. They may be using similar data and aspire to similar clarity in rankings, but the purpose is to permit leaders to understand and address the challenges confronting them.

Interest in indicators is not limited to communities. Across the nation, public and private institutions, from colleges and universities to hospitals and corporations, have been busy trying to develop indicator systems—sometimes as much for the benefit of management and executives as for the local public. For example:

- In 1999, North Carolina State University developed a planning process that incorporates strategic indicators (e.g., membership in prestigious associations and those *U.S. News* rankings), dashboard indicators (graduation and placement rates), and program indicators (passing rates on licensing examinations).[4]
- The Franklin Community Health Network in Farmington, Maine, provides quality indicators to its board and administrators to track patient satisfaction, re-admissions within 31 days, documented patient falls, in-patient mortality, and medication errors.[5]

- Private firms such as Visual Mining have developed very sophisticated software programs that make it possible to develop "executive dashboards," customizable charts and gauges that graphically represent complicated quantitative data.[6]

What seems clear is that a wide variety of public and private entities view indicators as a promising tool for monitoring the performance of large, complex systems.

In recent years, communities as diverse as Baltimore, Cleveland, Jacksonville, Silicon Valley, Seattle, St. Louis, and St. Paul have all launched ambitious indicator projects.[7] These efforts have revealed a number of methodological challenges. Among them:

1. **Top-down vs. grass-roots development.** Many indicator efforts collapse following a highly inclusive consultative process that produces laundry lists of desirable data, with no way of sorting out what is important from what is interesting but peripheral. Almost invariably, expert opinion favors detailed data on processes, while public opinion focuses on results but then concedes that everything the experts seek is desirable also. The conflict between top down and grass roots methodologies[8] often results in a muddled mixture of both.

   Experts believe that social indicators as developed in the past will continue to fail "if the production of social indicators is confined to collecting statistical data, publishing chartbooks and reports, and publicizing findings."[9] This conclusion carries with it clear implications for the utility of producing school report cards of the sort encouraged by No Child Left Behind.

2. **Snapshots or trends.** Snapshots tell you what just happened, and they are relatively inexpensive. This is what most states' criterion-referenced tests do: capture the performance of this year's 4th graders in math as of a given month, and then capture the performance of the next year's 4th graders a year later. Trends put together such snap-shots over a period of time and thus provide a moving picture of a community's performance over time. Such trend analysis is more complex and costs more, but it avoids the danger of making decisions on the basis of a single, and possibly unrepresentative, year's data. Jacksonville's effort, with an emphasis on quality-of-life indicators, tracked such trends as graduation rates, resident-infant deaths per 100,000 births, and public perceptions of elected officials.

3. **Comparison vs. measurement against preset standards.** Comparative analysis was the approach of choice in greater Baltimore. Here, the interest was not so much in how well the community was doing in absolute terms but how it stacked up with similar communities elsewhere. Seeking "best-in-class" examples to emulate, Baltimore compared itself to 20 comparable metropolitan areas in terms of population diversity, income, unemployment rates, and tourism. Other comparisons are also possible. In Connecticut, schools are compared within Educational Reference Groups, collections of schools organized according to their similar demographics and student composition.

   On the other hand, some states, like Texas, rate schools according to how they meet specific standards: a school is termed "exemplary" if it meets all key requirements for that rating. In the first type of comparison, the "standard" is the average value of the indicator in the state, district, Educational Reference Group, or other comparison group. The second kind of measurement is defined against specific standards. It is not clear that either is better than the other. Both kinds of comparisons may be measuring and judging by criteria that are entirely appropriate. On the other hand, each may be judging by inappropriate standards or by criteria that have little relationship to high-quality schooling.

4. **Parsimony vs. complexity.** Most cities and school districts that have undertaken indicator development have struggled to find a balance between the "one best indicator" (a single number or designation that would summarize the standing of a school or other organization/city) and complexity. An example of the "one best" approach is found in the way many states have interpreted NCLB Adequate Yearly Progress provisions. The different determinants are combined into a single meaningful dichotomy: failing or not failing. This is parsimonious, but is frequently attacked as overly simplistic and experientially unfair. An example of a highly complex indicator system is found in Cleveland's use of "benchmarking" that combines both trend and comparative analysis. Cleveland benchmarked itself against 13 comparison regions on 114 comparative indicators and 46 trend indicators and could therefore assess itself against the competition on everything from environmental contamination to office vacancy rates. On several critical issues, people in Cleveland were able to measure their progress relative to where they began (e.g., college completion or number of

families on food stamps). Whether such complexity can yield effective action has yet to be seen.

The management guide outlined in this report addresses each of these challenges. The indicators illustrated in Chapter 1 and described more fully in Chapter 3 were selected specifically so that they are few and clear. They were also designed so that it would be possible to look at both status (the snapshot) and trends, not only in comparison to other schools and districts, but also against benchmarks set by the district or state.

How did we decide which indicators to use in school systems? Three criteria were used throughout: data power, proxy power, and communications power.

- **Data power.** Here, the question is how the indicator meets the basic criteria of measurement. Are the indicators selected valid and reliable? Is the measurement accurate and does it meet the normal data standards of validity and generalization? Are the data clearly defined and collected in the same way from one year to the next? If someone else used the same data, would they get the same results? (That is, is it computationally transparent?) And are the data timely and available? Many of the indicator systems recommended over the years have called for extensive surveys of parents, detailed review of student histories, and the collection of sensitive information about the households in which students live. The data derived from such forces would likely be powerful if available, but often they are not available.

  Defective temperature sensors send the wrong signals to the automobile dashboard. Schools can't rely on defective sensors. The data employed in a school indicator system must be powerful and reliable.

- **Proxy power.** Here, we are interested in whether the indicator acts as an effective proxy for what it is trying to measure. Does the indicator "get at" what most people think is important? Achievement test results contain great proxy power. They focus on something that everyone considers important. But most people also believe several other things are important, too, including indicators of the achievement gap, adequate school funding, and children's engagement in school. The question is: are achievement test results adequate proxies for school effectiveness? Are measures of attendance and school participation adequate proxies for school engagement? In short, is the collection of indicators able to provide appropri-

ate information about the major concerns people have about schools?

- **Communications power.** Does the indicator describe something most people comprehend? People understand body temperature, unemployment rates, the cost of living, and student achievement. What indicators in education might appeal to a wide audience, engaging a large cross-section of people in terms they can appreciate? Here it is extremely important that the indicator be able to be graphed—displayed graphically and tracked over time. The National Center for Education Statistics has struggled for several years to develop ways of reporting NAEP (National Assessment of Educational Progress) results in a way that can be understood by the media and conveyed to the general public. They admit that they have yet to find ways that both convey the complexity of the data and can be readily understood by the general public.

It is clear from this short review that indicator systems are an accepted part of the American experience and that school leaders could benefit from systems incorporating the successes (and avoiding the pitfalls) of other indicator systems. The management guide presented in the first chapter was designed to provide such a system. But much of the utility of the management guide will be determined by two factors. First, appropriate data on schools in the district and state must be available. Second, the issue of the capacity of district personnel to analyze and present the data needs attention. The management guide is flexible enough to be adjusted to both varying data sources and analytic capacities. The following section discusses the minimum requirements for the local development and use of a management guide such as the one presented here. The next chapter provides a more complete explanation of each of the seven indicators used in the guide, along with examples of how status and trends can be displayed for each.

1. **Achievement (reading and mathematics).**

   Data requirement: Scale scores from a criterion-referenced test (like Washington State's WASL) or norm-referenced test (like Iowa Test of Basic Skills) taken by all students at a particular grade level/grade span within the district. In order to compare a district's schools to other schools within a state, it would be necessary to have scale scores on the same test(s) for all or most of the schools of a particular grade level/grade span in the state. It is necessary to have scale scores for at least one grade level within each grade span, although it is not necessary to have the same test at each level/grade span. For trend analysis, in the best of all possible worlds, the indicator system would track individual students through the grades, using linked tests for each

grade level. Without the availability of such data, trends would be established by looking at the same grade level in successive years.

Analytic requirement: An internal statistician or external consultant able to work with SPSS, SAS or other statistical software to generate cumulative distribution functions (CDF) is essential. Most of the charts and graphs can be drawn in Excel, but much of the data used to generate the graphs would have to be manipulated by means of statistical software. Many larger school districts, most state departments of education and most colleges and universities have people trained in the use of such software.

2. **Elimination of the achievement gap.**

Data requirement: Scale scores used in the development of the achievement indicator disaggregated by race/ethnicity and other student characteristics, if possible. Thus, if free/reduced lunch eligibility (or some other measure of demographic difference) is available, this could be used either in addition to, or instead of, the disaggregation by race/ethnicity. Again, the best possible measure of changes in the gap would come from looking at each cohort of students as it progresses through the grades. Without this, it is necessary to look at the gap in the same grade in successive years.

Analytic requirement: Again, an internal statistician or external consultant would be needed, at least initially, to set up the software to produce the distribution functions that are used to compare two or more groups of students disaggregated by race, income, and the like. SPSS, SAS and other software packages have these capabilities. Excel or one of the statistical packages can produce the graphs.

3. **Student attraction (school ability to attract students).**

Data requirement: As presented in the management guide, this indicator assumes that parents/students within the district are able to make some choices about what schools to attend from within the district or a catchment area within the district. If such a situation exists, then what is required is simply the number of students designating the school as "first choice" in a given school year. The capacity/expectation would be computed from the number of seats available in the school as a proportion of all seats available at a given grade level. When no choice is available to parents, a substitute might be found in the percentage of all children of a particular age who live within the geographic

catchment area of the school who attend the school. This data is available from the U.S. Census every 10 years and can be estimated by state and local planning offices for intervening years.

Analytic requirement: The analysis of the "choice" status and trend data can be done entirely in a basic spreadsheet such as Excel. However, if it is necessary to use data from the U.S. Census or local planning offices, more sophisticated tools may be necessary. These alternatives have not yet been worked out for the management guide.

4. **Student engagement with the school.**

Data requirement: There are several possible sources of information here, depending on the grade level and grade span. For all levels/spans, the most important data will be average daily attendance (ADA) since this has been found in the research to have the highest correlation with disciplinary actions, school grades, and retention in school. Other sources of information that could be added to ADA (and incorporated into an index, which may be developed if needed) would be student disciplinary actions and student involvement in extra-curricular activities. Although these are correlated with ADA, their combination in an index could provide valuable information to school leaders.

Analytic requirement: Once collected, these data (with the exception of the development of an index) can be analyzed using a standard spreadsheet. Graphs can be produced either from the spreadsheet or a statistical software program such as SPSS.

5. **Student retention/completion by school level.**

Data requirement: The data used in this indicator are typically collected at the beginning and end of each school year. The status report is simply a graphic presentation of the grade level enrollment in the school, by race/ethnicity, at a single point in the school year, usually at the beginning. The trend report gathers this information for each cohort and year. Thus, the current year's graduating class is looked at as a cohort, even though many individual students may have come and gone since the beginning of the cohort's first year in school. For example, in a middle school there would be four data points for each cohort: at the beginning of the 6th, 7th and 8th grades and at the end of the 8th grade. This would be done for five cohorts so that it would be possible to see whether there are major differences within cohorts (e.g., "the 2001 cohort lost a disproportion-

ate number of students, but the other cohorts have been very stable") or within grade levels (e.g., "there seems to be a pattern of dropping out/not returning between the 7th and 8th grade for all cohorts").

Analytic requirement: No advanced training is required to analyze or present the data. Some manipulation of the data is required for the trend analysis, but it can be done in Excel or a similar spreadsheet. Detailed instructions on how to set up the cohort analysis and graphs may be necessary.

6. **Teacher attraction and retention.**

Data requirement: The data necessary to present this indicator are almost certainly available, but often not easily accesible. Two pieces of information would be necessary for each year for each school within a district: (1) the number of teachers who apply for each opening in the school and (2) the number of teachers who leave the school during or at the end of each school year. The second item really represents the number of openings a school has in a given year that are generated by leave-taking rather than position creation. These data are not always kept at the district level and it may be necessary to collect the information from each school individually. Once started, however, such data collection should be routine and shouldn't require excessive paperwork.

Analytic requirement: Again, analysis and presentation of these data is fairly simple and straightforward. It should require nothing more than detailed instructions for setting up the tables and graphs and a standard spreadsheet to generate the indicator and its graphs.

7. **Funding equity.**

Data requirement: This is the indicator that has the most complex data requirements. The CRPE researchers who developed this indicator spent many months in districts helping central office and school staff collect and/or categorize data on individual school allocations and spending. At a minimum, it would be necessary to collect and report the budgeted and actual salaries of all teachers in every school over a period of years. To assess whether a school is receiving the per-student support that would be implied by funding formulas, even more work might be necessary, setting up what might be a different but apparently more useful accounting system. Chapter 3 presents a more complete discussion of what this indicator requires in

terms of data and analysis. Here, it is enough to know that the information required is usually available, even if deeply buried, in district records.

<u>Analytic requirement:</u> The analysis required for this indicator is fairly complex and will almost certainly require outside support at first. Once the data are routinely collected and the analysis software set up, only minimal training will be necessary to maintain the data and present this indicator.

As is clear from the above, most of the indicators will require initial set up of data, either identification of existing data or collection of minimal new data. Some analytic sophistication will always be required when the indicators are first analyzed and incorporated into the management guide. However, as with the data requirements, once these processes are established, there should be little need for outside support. The development of the *Consumer Report*-type tables will require, again, some initial help from internal or external statisticians familiar with comparison of one set of data with another—in this case, district vs. state data, or school vs. district data. The processes will probably require use of a statistical software package such as SPSS or SAS but, once the processes are established, ongoing maintenance of the system should be fairly straightforward. This is one indicator that would benefit from assistance provided by technical support such as that provided at the State Office of Education.

After experimenting with this guide, school leaders might suggest additional or alternative indicators. However, any alternative indicators selected should meet the criteria set in this chapter for the guide itself. Indicators should be reliable (i.e., they should possess data power); they should provide information beyond their own limits (i.e., they should demonstrate proxy power); and they should be able to be used to inform policy decisions critical to the improvement of public education (i.e., they must possess considerable communication power).

---

[1] An extensive discussion of the accountability requirements of the No Child Left Behind Act are provided by the Council of Chief State School Officers in *Making Valid and Reliable Decisions in Determining Adequate Yearly Progress*. (Washington, DC: CCSSO, December 2002.) This report points out that, under NCLB, adequate yearly progress would be based on up to 37 "determinations" of student performance in each of the grade spans 3-5, 6-9 and 10-12—in at least reading and mathematics.

[2] Orshansky applied her calculations to Agriculture's "economy food plan"—the cheapest of four food plans developed by the Department of Agriculture. She also developed a second set of poverty thresholds based on the Agriculture Department's somewhat less stringent low-cost food plan, but relatively little use was ever made of the higher thresholds.

[3] One of the best sources for understanding the principles behind and history of indicator systems is found in Judith Eleanor Innes, *Knowledge and Public Policy: The Search for Meaningful Indicators*. 2nd Expanded Edition. (New Brunswick: Transaction Publishers, 1990).

[4] www.2.acs.ncsu.edu/UPA/compactplan/compact99/indicators_performance.htm

[5] www.fchn.org/fmh/quality/dashboard.asp

[6] www.visualmining.com

[7] See: *Greater Baltimore: State of the Region Report, 1998* (Greater Baltimore Committee & Greater Baltimore Alliance); *Rating the Region, 1997* (Citizens League Research Institute, Cleveland, Ohio); *Quality of Life in Jacksonville*, November 1997 (Jacksonville Community Council, Inc.); *Index of Silicon Valley*, 1997 (Joint Ventures); *Indicators of Sustainable Community*, 1998 (Sustainable Seattle); Where We Stand, 1996 (St. Louis Region East-West Gateway Council); and *Maintaining Our Competitive Edge for the 21st Century*, 1998 (Metropolitan Council of St. Paul).

[8] A well-reasoned plea for the grass-roots approach which ends up advocating for top-down imposition of the resulting recommendations is contained in a paper by two British educators who have developed a system for monitoring schools that is now used in England, Scotland, Australia, and elsewhere. Carol Taylor Fitz-Gibbon and Peter Tymms, "Technical and Ethical Issues in Indicator Systems: Doing Things Right and Doing Wrong Things," *Education Policy Analysis Archives*, Vol 10, N. 6, January 16, 2002.

[9] Clifford Cobb and Craig Rixford, *Competing Paradigms in the Development of Social and Economic Indicators*, a paper prepared for the Centre for the Study of Living Standards Conference, Ottawa, Ontario, October 1998, p. 1.

Chapter 3   School-Level Indicators and Their Role in the Management Guide

Outside medicine, few fields are the subject of such intense public analysis as education. Given the sheer volume of data about schools and the hundreds of articles published each year trumpeting evidence of school effectiveness, it should be possible to develop a parsimonious set of educational indicators that contain great power in terms of data, proxy value, and communications utility.[1] The first two chapters of this report provide a concrete example of a set of educational indicators that meet these criteria. In short, the indicators selected for the management guide are designed for use in monitoring the general health of the educational enterprise on a school-by-school basis within a district or a state.

The indicators selected for the guide can't tell school, district and state personnel everything about the system, but they serve as a mechanism for providing feedback about systems that might otherwise be too large and cumbersome to understand. Like the unemployment rate, the poverty index and the Dow Jones Average, they provide insight into complex modern systems, offering leverage points for thinking about what large systems need when they're in distress. They also offer a center of gravity for educators and citizens faced with mountains of data. These indicators can't diagnose problems or prescribe solutions. They won't tell superintendents, board members, or other leaders what is wrong, but they will instantly tell when something is wrong and offer those in leadership positions some preliminary information about where to begin and what to examine.

As described by Superintendent Chris Hernandez, the management guide provides both status indicators and trend indicators in seven areas. Each of the indicators tells part of the story, but even all taken together cannot possibly tell the whole story. However, these indicators are based on what research tells us about school and student characteristics associated with improved educational outcomes. Some are more thoroughly researched and powerful than others, but each provides a unique piece of the story that can act, cumulatively, as either a wake-up call (to shock, enlighten, and jumpstart) or a guide (to provide the goal or standard to be attained), or both.

The seven indicators are the following:

1. **Achievement** (reading and mathematics).
2. **Elimination of the achievement gap** in reading and mathematics between subgroups of students by race, economic status, English language facility, etc. (where there are adequate numbers within a subgroup for comparison).
3. **Student attraction** (ability of the school to attract students where there are opportunities for choice among parents/students).
4. **Student engagement with school** (index of measures of school engagement, including attendance, tardiness, and involvement in school activities).
5. **Student retention/completion** (depending on the level of the school: elementary, middle school, high school).
6. **Teacher attraction and retention** (number of applications for teacher openings; proportion of teachers leaving the school for reasons other than scheduled retirement).
7. **Funding equity** (measure of whether the school receives the funding that would be predicted given the composition of the student body).

The reason for and brief explanation of each of the indicators is provided below.

Achievement

Standards-based test scores have become something of a lightning rod in contemporary education. Some experts see them as necessary measures of the effectiveness of a school or school system, while others view them as a force that limits the creativity of educators and pupils and pushes students out in the end. Whether they are bane or boon in general, they are essential to any indicator system.[2] Two items need brief discussion here: the use of test scale scores instead of the ubiquitous "percent meeting standard" and the reporting of only math and reading scores.

Although there is a satisfying directness in the use of a single number to characterize a given school or group of students (i.e., percent meeting standard or classified as "proficient"), such an approach ignores the fact that scores below or above the "cutoff" may be distributed in very different ways. If most of the "below standard" scores are clustered close to the cutoff point, the approach to raising achievement would be quite different from

the approach required if the "below standard" scores were found primarily at the bottom end of the test-score distribution. Richard Rothstein made an impassioned plea in the *New York Times*[3] for using scale scores in reporting criterion-referenced test performance, noting that the cut-points used to determine the standard are simply a pre-determined point on the scale score distribution, not a magic number. Thus, moving the cutoff point one direction or another could make a radical difference in the percent "meeting standards." Rothstein noted, "Criterion-referenced reporting can't detect growth except when a student passes one of only a few fixed points on a scale."[4]

A "scale score" is neither the raw score a student earns (i.e., the number of correct answers) nor a percentage of correct answers. It is a number on a scale that is derived from the raw score but takes into account differences in the form of the test students take. A well-known example of a scale score is the Scholastic Assessment Test (SAT). For both verbal and math portions of the test, the scale runs from 200 to 800. Raw scores are converted to that common scale, even though test forms vary. Most states use scale scores and research on what different scale scores mean in terms of acquisition of required knowledge and skill, to set two or more cutoff points along their scale, with the most important division being between those who are "proficient" and those who are not. In Washington State, the two lowest categories ("not proficient") were originally titled "well below standard" and "below standard" but are now called "below standard" and "approaching standard." The two highest classifications ("proficient"), are labeled "meets standard" and "exceeds standard." States differ in cutoff points terminology used and implications of meeting or not meeting the standards.

For leaders to see and understand how students in a school are actually performing, it is not enough simply to know how many students fall to one side or another of an arbitrarily drawn bar. Using scale scores, permits educational leaders to understand where their students stand as they monitor efforts to improve achievement or close the achievement gap. "Percent meeting standard" provides no such guidance. Using scale scores, also permits leaders to detect change over the entire range of scores. So, for example, an annual increase of 2% or 3% in the proportion of students meeting standard is certainly cause for celebration, no matter how it is achieved. But if that 2% or 3% represents students already close to the standard who were levered over the bar, that is not nearly as impressive an accomplishment as if some portion of the newly successful students came from the bottom of the distribution barrel. Indeed, districts congratulating themselves on annual increases of 2% or 3% in those meeting standard might find themselves with tougher challenges ahead—if most of the students remaining below standard are substantially below the bar.

The academic subjects used in the management guide are math and reading, generally accepted as the two "basic" skills without which a student is

unlikely to do well on other criterion-referenced tests like writing, sciences, social sciences, and so forth. The correlation among the scores is very high and statistically significant.[5] It is conceivable that one of the two scores might be used by itself. However, reading and math, together, are the scores most generally accepted as meaningful.

Achievement Status: The black or grey circles on the management guide (Figures 1.1 and 1.2), in the Achievement row are based on scale score data from schools and on what is known as relative distribution/density analysis. This method of analysis and presentation is explained in detail by Handcock and Morris[6] and developed specifically to study and report on "gaps" among different groups in society, with special attention to variations in income. Handcock and Morris wanted to provide a full picture of the distribution of different measures, rather than simply summary measures like means, modes, or "percent meeting standard." Since the method was specifically developed to show the relationship of one group to another (e.g., scores of Hispanic students and white students, earned income of male and female employees) it does not rely on a hypothetical population as represented by the standard bell-shaped curve.

The straight line (reference line) shown as a solid black line in Figure 3.1 can belong to any reference population: e.g., all 7th graders in the district, all 7th graders in the state, or all 7th graders in urban schools. In the case of the example given in Chapter 1, the reference group is all 7th graders in the district, with the distribution of math and reading scale scores for a given school presented in relation to the district distribution.

Two examples of achievement status data are provided below in Figures 3.1A and 3.1B. The first is of a school with a fairly high proportion of students meeting standard, but a bi-polar distribution where students tend to fall at either the lowest or highest ends of the district distribution. The second is of a school with lower than average achievement scores and a large proportion of students in the 7th grade scoring in the bottom deciles of the district distribution. The two patterns of scores tell very different stories about the schools in question. At the first school (called Guy Fawkes in this example), there are students at both extremes of the score scale; at Monmouth Middle School, a majority of students are very poor readers and there are almost no students in the highest achievement levels.

In short, what Figures 3.1A and 3.1B reveal is that school leaders seeking to improve achievement at Guy Fawkes and Monmouth face quite different challenges. Guy Fawkes requires two strategies: renewed commitment to bring students in the middle deciles (deciles 4-7 particularly) up to standard, combined with a far more intensive effort to address the problems of large numbers of students in the bottom three deciles. What works for students in deciles 6 and 7 is likely to be of little value to students in deciles 1 and 2.

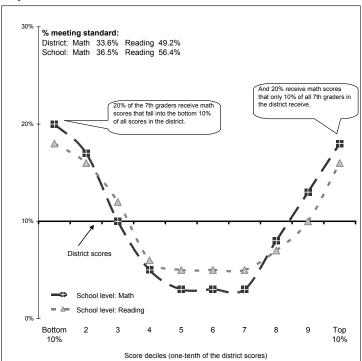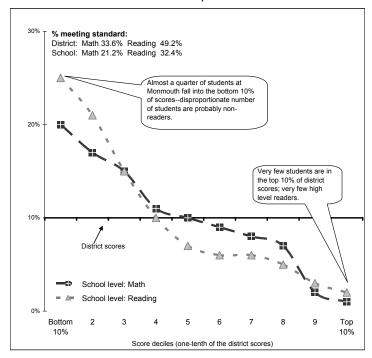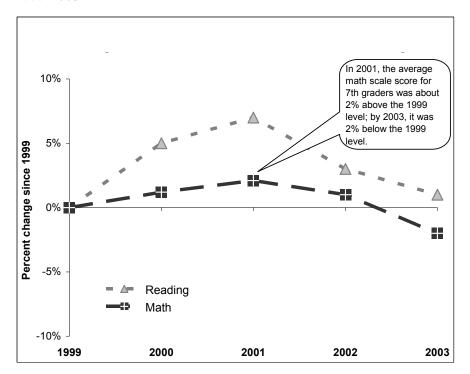**Figure 3.1A**
Guy Fawkes Status: Achievement, 2003

Figure 3.1B
Monmouth Status: Achievement, 2003

Monmouth faces the need for remediation almost across the board. Although there is a need to dramatically increase the number of high-level readers and dramatically decrease the number of students at very low reading levels, achievement is very low overall and a variety of approaches need to be considered to address these problems.

Achievement Change: Rate of change analyses provide information on how scores have changed over time. Changes from year-to-year are likely to be highly unstable but potentially indicative of progress toward academic achievement across the spectrum of students.[7] Achievement change graphs are not cohort charts, which would show progress of the same group of students as they move through school. Still, they provide a picture of what is happening within a given school building, at a particular grade level, from year to year. The first of the five years is the base year, shown as a "0" in Figure 3.2. Each subsequent year of data used in the management guide shows the percent change from the base year. Figures 3.2A and 3.2B show the patterns for our two hypothetical target schools, Fawkes and Monmouth.

Figure 3.2A
Guy Fawkes Trends: 7th Grade Standardized Math and Reading Scores, 1999-2003

Figure 3.2B
Monmouth Trends: 7th Grade Standardized Math and Reading Scores,
1999-2003

Perhaps not surprisingly, here again we find different patterns at work. Guy Fawkes shows improvement for two years and then a troubling drop. Monmouth, meanwhile, demonstrates stability, frequently considered to be a desirable state of affairs. But stability here is no virtue since growth is urgently needed. It is not clear what lies behind either situation. New staff, new curriculum, new students, or new attitudes may explain what is going on at Guy Fawkes. The indicator does not diagnose the challenge; it identifies it. Beyond random variation due to different cohorts of students, the challenge at Monmouth is that nothing much seems to be happening. This may be a case where the indicator both identifies and provides a partial diagnosis of the problem.

Achievement Gap

One of the great accomplishments of the accountability movement of recent years has been the insistence that data on average student achievement be "disaggregated" so that low achievement among particular subgroups (e.g., ethnic and low-income groups) is not concealed within overall averages. In the last five years, the importance of closing the achievement gap has taken on an urgency never seen before in the United States. There is no doubt that, in education, the achievement gap is the major learning issue.

Educational leaders need to know what is happening here, both overall and within individual schools.

The achievement gap discussed in this report is the gap between or among racial/ethnic groups. Those descriptors were the only student level descriptors available in the mythical school district being analyzed. However, some districts also collect information on free/reduced lunch status and family composition (e.g., single parent households) that could be used to analyze student achievement. The same approach can be taken, independent of the gap being considered.

Gap Status: The management guide uses an analytic/graphic method based on the cumulative distribution of individual student scores—a method of presentation often used in medical research, marketing, insurance, and elsewhere. This approach has been suggested to the Educational Testing Service for use with data from the National Assessment of Educational Progress (NAEP) in the sponsoring organization's long search for effective ways to present NAEP results to the public.[8] Cumulative distribution (also known as Lorenz) graphs display test scores across the entire range of performance and make any gaps visually evident. In Figures 3.3A and 3.3B the visual gap between the diagonal line (the reference group) and the subgroups of interest is the "gap status." In these graphs the diagonal line represents the cumulative distribution of scores of the base group (usually the white and, in some cases, Asian students). Because the distribution is cumulative (going from 0 to 100%), the line is diagonal rather than horizontal as in the achievement graphs, where scores across the district are divided into ten equal groups.

The other lines on the graph show the cumulative distribution of minority student scores within the school. The distance of the minority student scores from the majority student scores constitutes the "gap," with a larger area indicating a larger gap.

How should Figure 3.3A be understood? It tells us several things about Guy Fawkes. In math (black line), most of the gap exists in the bottom 80 percent of the achievement distribution. Fortunately, the gap between students at the highest math achievement levels seems to be non-existent. In reading, by contrast (grey line), there is a large and relatively consistent gap across all achievement levels, and it seems particularly pronounced in the bottom two deciles. The picture in Monmouth is not as complicated. A fairly uniform gap exists across all achievement levels, in both reading and math.
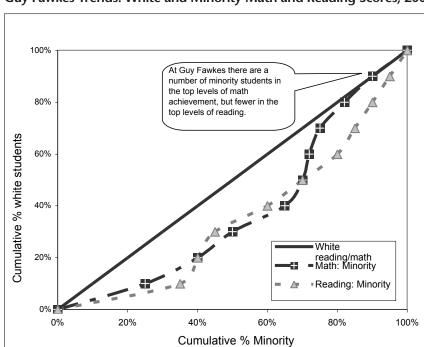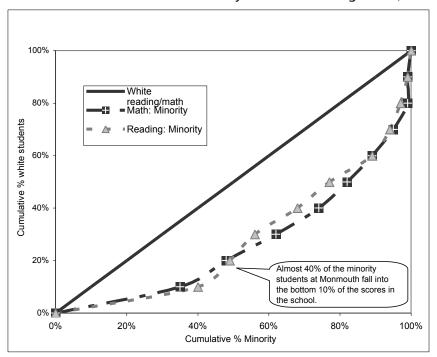
Figure 3.3A
Guy Fawkes Trends: White and Minority Math and Reading Scores, 2003

Figure 3.3B
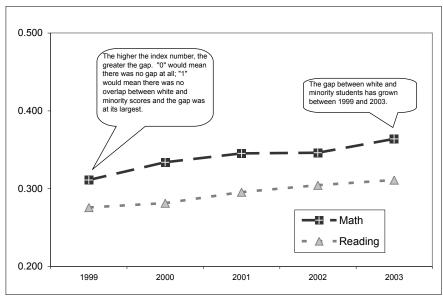**Monmouth Trends: White and Minority Math and Reading Scores, 2003**

Change in the Gap Status: The gap status graphs in Figures 3.3A and 3.3B, known as Lorenz graphs, show a moment in time. Mathematically, the total area between a reference line and the lines for any subgroup on a Lorenz graph can be measured and assigned a single number known as a "Gini Coefficient." The Gini Coefficient is always a number between 0 and 1 (like a correlation coefficient); in the case of Gini Coefficients, the higher the number the larger the gap. If there was no difference in the distribution of scores between the reference and comparison groups, the Gini Coefficient would be 0.

The Gini Coefficient can also be tracked over time, providing an indicator of change in gap status. Figures 3.4A and 3.4B track the changes at Guy Fawkes and Monmouth in terms of this coefficient.

What we see in Figure 3.4A is that, in both reading and math, the Gini Coefficient has increased at Fawkes (one of the district's flagship schools), despite the fact that the scale scores themselves are about the same in 2003 as they were in 1999. At Monmouth (Figure 3.4B), on the other hand, the gap has narrowed somewhat in both math and reading, although the change is not statistically significant and there are sizeable swings from year to year. In the case of both schools, the displays of the changes in this coefficient provide information unavailable from the scores themselves, or even the scores over time.
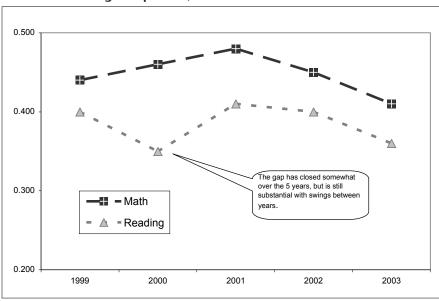
Figure 3.4A
Guy Fawkes Change: Gap Index, 1999-2003

Figure 3.4B
Monmouth Change: Gap Index, 1999-2003

What's Involved When Schools Try to Close the Achievement Gap?

States, districts, and schools are struggling to close the achievement gap, with individual states and the federal government setting various deadlines for the complete elimination of gaps that have existed for generations. It is not surprising, then, that educators under the gun to close the achievement gap have fallen into the habit of hailing relatively small annual signs of improvement between groups as evidence of success. Left unspoken is the inference that these improvements, if continued, will inexorably close the gap. It seems to make sense, and many people of good will believe the achievement gap is well on its way to elimination.[9] Under current trajectories, however, these expectations are likely to be disappointed.

Close analysis of most improvements in this area indicate that small annual improvements are, in fact, tiny baby steps. In Washington State, for example, a major analysis of more than 800,000 statewide assessment results concluded that despite the state's relatively impressive progress in closing the gap, 80% of the gap would remain after five years and 50% after ten.[10] This conclusion was reached despite the fact that the gap was closing in Washington State much faster than national averages. Figure 3.5 provides a graphic reminder to education leaders of the stark challenge that is involved in closing the achievement gap.
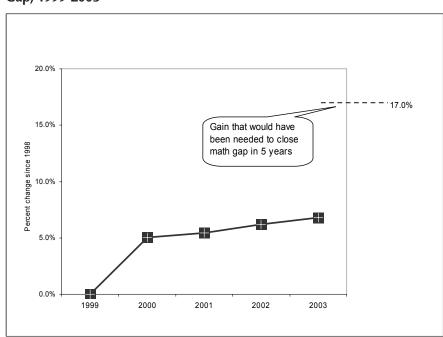
Figure 3.5
Monmouth Change: Achievement Gap With Change Needed to Close Gap, 1999-2003

Figure 3.5 displays a graphic picture of progress made in one school in Washington State, versus what's required. At Monmouth, minority math scores increased over 6% from 1999 levels. However, the improvement would have had to be almost three times as great (17%) over that period of time for the gap to close (as indicated by the dotted line at the top right of the graph). A nearly identical figure could be produced for reading at Monmouth.

Reassured that things are improving by annual increments, most members of the public (and many educators) have little conception of just how drastic an improvement is required to close the gap.[11] The reality is that most states and districts are not closing the gap rapidly enough to get the job done—in five years, ten years, or even twelve. This is a critical issue in No Child Left Behind since the legislation contemplates that schools will have eliminated the achievement gap after twelve years, or lose federal funding.

Student Attraction

Makers of toothpaste, producers of TV shows, and designers of teen clothes all conduct extensive research into what makes a product attractive to the potential consumer. Some public school districts, and many private schools, do similar market research to determine what parents and students are looking for in a school. In the absence of intensive polling, one way of know-

ing whether a school is attractive to its target market is to look at families' choices when choices are possible. Many public school districts now permit a modicum of choice for parents, most often by offering magnet or alternative schools that are open (usually by lottery) to all students in the district or permitting parents to rank their school choices from among any within the system. Where some level of choice is possible, the management guide uses an indicator called "student attraction," which will differ in definition according to district policies.

Attraction Status and Change: In Rebel Valley, parents and students can designate any school in the system as first, second, or third choice, and approximately 90% of the families get one of these three choices. If everything were equal (that is, if every middle school was equally attractive), then the percentage of students would be comparable to the capacity of the school. Therefore, if the school has the capacity to enroll 20% of the 6th graders in the district, and all schools have the same ability to attract students, the district could expect that about 20% of the incoming 6th graders would choose that school. Since Rebel Valley has had this policy for more than five years, trends are available, as indicated in Figures 3.6A and 3.6B for the two schools already introduced.

Figures 3.6A and 3.6B reveal that while Guy Fawkes is able to accommodate approximately 18% of the incoming 6th graders in the district, it attracted more than its share in the years between 1999 and 2001. After that time, however, the proportion of all students designating Fawkes as first choice has dropped below its capacity. In other words, either Guy Fawkes is no longer seen as a highly attractive choice by a number of parents, or other schools have become more attractive. Monmouth was substantially below capacity and expectation in 1999 and continues at about the same levels after that. In both schools, something seems to be at work to lessen the schools' attractiveness.

What does this have to do with districts in which school choice is limited or not available at all? Even here, proxies can be developed. Large urban districts without an established policy of choice among available public schools can learn a great deal by comparing the number of children living within a school's catchment area with the number of students actually enrolling in the school. U.S. Census Bureau data, available down to the block level, are a good place to start. A large number of "missing" students may indicate that parents have taken an exit strategy such as private school enrollment to find an appropriate school for their child.

Figure 3.6A
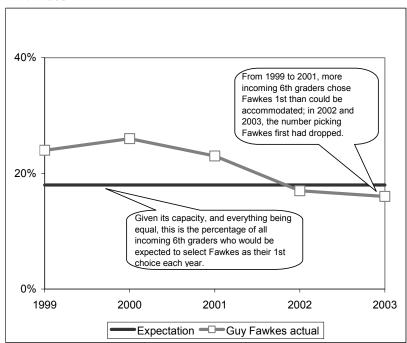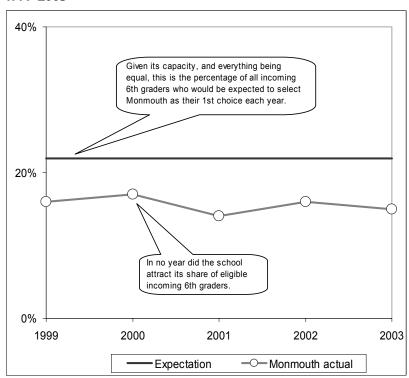**Guy Fawkes: Difference Between Expectation/Capacity and Choice, 1999-2003**

From 1999 to 2001, more incoming 6th graders chose Fawkes 1st than could be accommodated; in 2002 and 2003, the number picking Fawkes first had dropped.

Given its capacity, and everything being equal, this is the percentage of all incoming 6th graders who would be expected to select Fawkes as their 1st choice each year.

Figure 3.6B
Monmouth: Difference Between Expectation/Capacity and Choice, 1999-2003

Given its capacity, and everything being equal, this is the percentage of all incoming 6th graders who would be expected to select Monmouth as their 1st choice each year.

In no year did the school attract its share of eligible incoming 6th graders.

Engagement with School

One of the most potent behavioral predictors of failure in school and subsequent dropping out is simple attendance.[12] A summary of national research on the issue reported simply that "the Los Angeles County Office of Education identifies truancy as the most powerful predictor of delinquency... When Van Nuys, California, officials conducted a three-week sweep for truants on the streets, shoplifting arrests dropped by 60 percent. Absenteeism is detrimental to students' achievement, promotion, graduation, self-esteem, and employment potential. Clearly, students who miss school fall behind their peers in the classroom. This, in turn, leads to low self-esteem and increases the likelihood that at-risk students will drop out of school."[13]

On the other side of the coin, a potent predictor of persistence and success in school is engagement with the school, defined as involvement in school clubs, sports, and other extracurricular activities. Engagement has been viewed as both a cause and an effect of other risk factors, but there is clear evidence that the school itself has a strong influence on student participation and sense of belonging.[14] An international study conducted in 2000 was designed to assess the sense of belonging and participation of students in school.[15] The term "engagement", as used in the international research, and this management guide, refers to "the extent to which students identify with and value schooling outcomes, and participate in academic and non-academic school activities."[16]

Engagement Status and Change: As with most assessments in this area, the primary source of data for the international study cited above was self-report surveys of students. This is the ideal and has been used extensively by the Chicago Consortium in tracking the progress of educational reform in the Chicago Public Schools. However, few school districts have such data, especially data that make it possible to identify trends over several years. The management guide, therefore, relies on the use of behavioral data (attendance/absenteeism, tardiness, and membership in school-sponsored activities) to build this indicator. The information available includes average daily attendance, average class attendance (at high school level), percent tardiness, and percent of students belonging to school-sponsored activities. For the purposes of this report, only average daily attendance and activity participation are displayed in Figures 3.7A and 3.7B, although additional data can be added when available.

Figure 3.7A
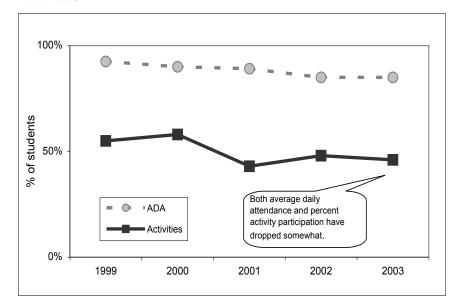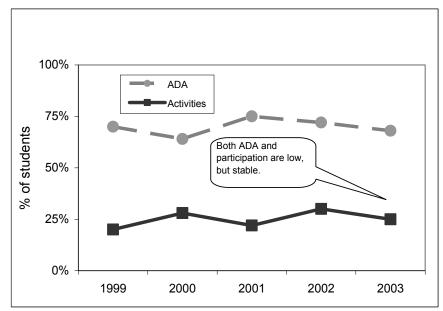**Guy Fawkes Trend: Average Daily Attendance and Activity Participation, 1999-2003**

Figure 3.7B
Monmouth Trend: Average Daily Attendance and Activity Participation, 1999-2003

What we see in Figure 3.7A (Guy Fawkes) is that the trendlines in this apparently successful school are fairly healthy in terms of attendance and student involvement in activities, but both indicators seem to be on the decline. At Monmouth, by contrast, despite some annual variation, the trendlines in both areas are essentially flat. Monmouth's problem is that each of these indicators demonstrates anemic performance; although not declining, both attendance and participation are lower at Monmouth than at Guy Fawkes. One can conclude, based on these indicators, that engagement at Monmouth seems to be stable and low. Meanwhile, engagement at Guy Fawkes, although healthier, is showing some weaknesses.

Student Retention/Completion

The retention/completion indicator assesses "leaks" from the system at each level—elementary, middle, and high school. Here the management guide shows the proportion of students who enter the system at the beginning of a cycle—1st grade, 6th grade, or 9th grade—and are still in school at the end of the cycle. Although this is commonly reported now as "completion rate" for high school students, it can provide valuable information to school leaders at every grade span. If a particular cohort of students begins 6th grade together and only 70% of that number are present at the end of the 8th grade, this may carry a warning to district leaders: for some reason, parents are abandoning the school and students who leave are not being replaced by others, as would normally occur with routine movement across a district. Exploring the reason for these "losses" could inform school leaders about potential problems in the school. Breaking the changes down by subgroups of students could also help illuminate the situation.

Ideally, it should be possible for a school and district to track each student through the system, permitting educational leaders to know whether an individual who started at Guy Fawkes Middle School remained there through three years. However, only 31 of 51 states (including D.C.) have individual student identifiers, and many of these states do not yet have data systems flexible enough to track students easily or economically for the purposes of district-wide monitoring.[17] Thus, the ideal (knowing where students are in the system over their years of schooling) may not be practicable within the immediate future. In the absence of the ideal, what most districts do, at least for high school students, is report dropout rates. There has been considerable controversy around these reports, however, and the way the data are collected and computed makes all the difference to their usefulness as indicators of school health.[18]

As a measure of a school's ability to retain its students, the completion rate used by Jay Greene of the Manhattan Institute has some real advantages over the traditional "dropout rate." Greene has demonstrated that the results of computing the completion rate tend to be straightforward and stark. The Manhattan Institute approach is used in the management guide.

The retention/completion rate measures school completion in the aggregate. It compares the number of students who graduate in a given year in a particular school, district, or state with the number who started at an earlier transition point. Every state conducts some version of an "October count" each year, and some states complete counts at other points during the year. The data are usually broken down by gender and race, but with little additional information on individual students.

It is unlikely that the retention/completion rates can explain much about why a school's population is increasing or decreasing over time. But the data can provide a convenient and highly useful way to determine whether changes are occurring that deserve more attention. If only one or two schools are experiencing significant changes in student enrollment, leadership may need to look closely at what is going on in those few schools. Changes across the district may require a more general response.

Retention/Completion Status: The status charts for retention/completion in the management guide (Figures 3.8A and 3.8B) show current enrollment by grade. They are basically a "snapshot" of the school population in a given year, by grade level; they do not reflect the movement of cohorts through a school. In most schools, we might expect the number of students in a grade to be approximately the same each year. Barring something obvious such as increased housing density in a neighborhood, significant variation in grade-level enrollment might alert leadership to possible changes in the school that might negatively affect its quality and attractiveness.

With regard to Guy Fawkes, Figure 3.8A displays a relatively small difference in enrollment from 6th to 8th grade. However, the figure also reveals larger white enrollment in 8th grade than in 6th. The pattern for minority enrollment is reversed, with more African-American and Hispanic students enrolled in 6th grade than in 8th. It is hard to know what these patterns mean. It is conceivable that retention and completion are relatively stable while the catchment area itself is experiencing modest demographic change.

Monmouth presents quite a different picture. Retention of students seems to be a problem, no matter what the student's ethnicity. Each grade level within the school enrolls fewer students than the grade level below it. There seems to be considerable loss between school years. In many districts, the Monmouth picture might be explained by a rise in the number of families with younger school-aged children in the attendance area. If that possibility can be ruled out, this indicator seems to be a warning signal that Monmouth requires management attention from district leaders.

Figure 3.8A
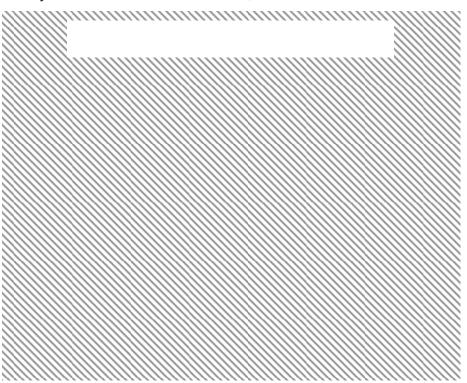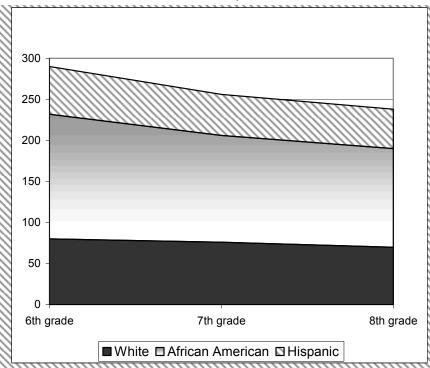Guy Fawkes Status: School Enrollment, 2003

Figure 3.8B
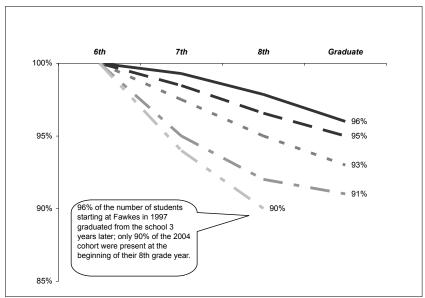Monmouth Status: School Enrollment, 2003

Retention/Completion Change: To see how a school's ability to retain its students changes over time, it is necessary to look at cohorts of students. For example, the Guy Fawkes Class of 2000 is the cohort that began 6th grade together in the fall of 1997. The management guide provides information on four previous cohorts and the cohort currently in the 7th grade. Thus, educators can see at a glance how schools have fared over time and can determine whether a particular situation (e.g., a major drop in a particular cohort or a drop at a particular grade level each year) needs additional investigation. Figures 3.9A and 3.9B provide examples from Guy Fawkes and Monmouth.

What we see in Figure 3.9A is an accelerating problem at Guy Fawkes. Each of the cohort classes, from 2000 to 2003, is somewhat less successful in retention/completion than the class before it. The changes aren't large in any single year, and Fawkes still has one of the best retention/completion rates in the district. However, unless there is a dramatic reversal in the class of 2004 (still attending Guy Fawkes at the time Figure 3.9A was developed), it will have the worst record of all five classes. This is a problem that seems to be growing gradually worse, and it is unlikely it can be reversed by ignoring it.

Retention and completion at Monmouth are in even worse shape than at Fawkes. Paralleling what was evident in Figure 3.8B, this new figure demonstrates that Monmouth routinely loses between 10 and 15% of its students between 6th and 8th grades. What had been a warning signal earlier is now flashing red.

Figure 3.9A
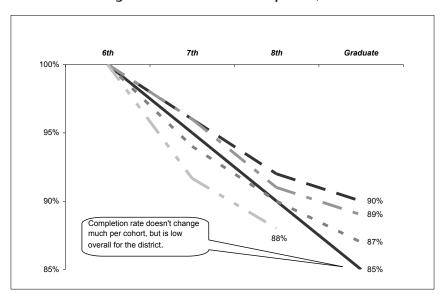**Guy Fawkes Change: Cohort Retention/Completion, 2000-2004**

Figure 3.9B
Monmouth Change: Cohort Retention/Completion, 2000-2004

Teacher Effectiveness

Teacher effectiveness is not an element listed on the management guide because, independent of student performance, there is no way to assess effectiveness at this time. There is a circular quality to many analyses of teacher effectiveness. The best teachers are identified as those whose students gain the most from their time with those teachers. To date, no research has been able to identify the characteristics that help effective teachers produce those student results. No external characteristic of the teacher—not years of teaching experience, type of certification, nor having a major in the subjects taught—has been shown to be reliably related to significantly greater student achievement.[19]

Interest in assessing teacher effectiveness is ubiquitous and increasingly urgent, given the pressures of NCLB. Most efforts have looked retrospectively at teacher performance through the test scores earned by their students. Such approaches are of little help to a district trying to determine the role a particular school's teachers may play in the performance of the totality of the students. The interactions among prior student achievement, school demographics, teacher preparation, and student background are too complex to permit easy disentanglement. At the moment, there is no direct way to measure the effectiveness of all teachers, or specific teachers, within a school or district. If a measure of teacher effectiveness is identified, it can be added to the management guide. Until such a measure develops, teacher attraction and retention are suggested as proxies.

Teacher Attraction and Retention

Although not an ideal indicator of the important role teachers play in school effectiveness, the measurement of teacher attraction and retention is both possible and useful. Just as it would be useful to know what parents are looking for in a school before their children walk in the door, it would be helpful to know how potential employees view the school before they accept positions. It would be valuable, also, to get a better understanding of what draws the most highly qualified teachers to particular schools.

If a school is perceived as unsupported, in trouble or failing, it is unlikely that many teachers—especially those with the most experience and expertise—will be interested in applying for positions there. An unattractive school might be on either pole of several variables: a weak or a domineering principal, uninterested or overly controlling parents, extremely needy or overindulged students. Whatever the cause, the perceptions of teachers can play an important role in determining whether a school will attract the numbers of teachers it needs and the quality it seeks.

It should be possible to identify attractive and unattractive schools by surveying current and prospective teachers. But such an effort is costly. Another approach is to quantify teacher attraction and retention. Roza reports that calling around to a few schools in Los Angeles revealed that some schools received one to three applicants, at best, per position, while others received up to 130 applications. Similar studies in the Seattle Public School District found much the same pattern: some schools have a handful of applicants per opening, while others receive dozens. It should not be a surprise that schools with few applicants wind up doing the best they can in terms of hiring, while schools with many applicants enjoy the ability to select among teachers with the qualifications and qualities the schools seek.[20]

Attraction and retention are not the same thing. It is possible that a school could attract many more applicants per opening than neighboring schools while experiencing higher than average teacher turnover, even for several years in a row. This situation might develop because potential applicants had not yet heard that the school was not a good place to work. There might also be more positive explanations, e.g., a core of experienced and highly qualified teachers retiring or moving around in the system to take on master-teacher or administrative roles. Either way, the management guide can act as an early warning system, alerting leaders to look closely at what is happening in a school. As with other indicators, a particular "rating" on this teacher attraction/retention indicator can call for additional attention from the superintendent and school board. Like the others, this indicator cannot provide a diagnosis or prescription, but it can call attention to this aspect of a school and possibly provide an early warning of problems ahead.

Teacher Attraction and Retention: Status and Change: Figures 3.10A and 3.10B provide a picture of teacher attraction and retention, both status and trends, in a single graph. For each illustrative school, it compares the school's level of attraction and retention to district averages in terms of applicants per opening and the percentages of teachers leaving the school in a given year.

Figure 3.10A
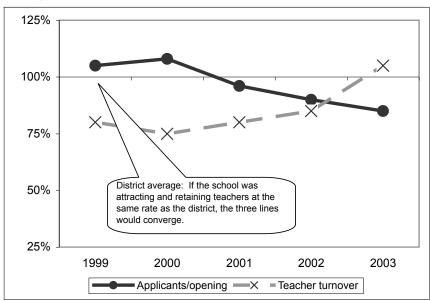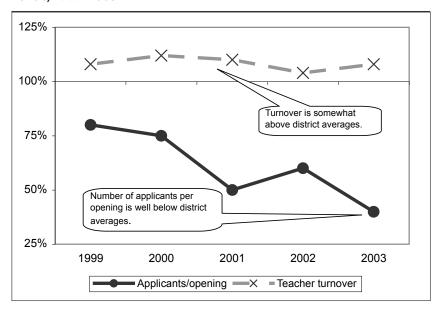**Guy Fawkes: Teacher Applicants and Turnover Compared to District Levels, 1999-2003**

Figure 3.10B
Monmouth: Teacher Applicants and Turnover Compared to District Levels, 1999-2003

Guy Fawkes, once again, seems to be showing some signs of impending problems on these indicators. Since 2000, the school has slipped from exceeding the district on applicants per opening (a positive position) to slipping behind. Meanwhile, where it once lagged behind the district average in teacher turnover (usually a positive situation), by 2003 it was slightly exceeding the district average. This may only indicate that normal retirements or promotions are affecting the teaching force, or it may indicate some turmoil in the school that is leading some teachers to seek positions elsewhere. Although the turnovers themselves may have a significant effect on the school, the reasons for the turnovers are likely to be much more important.

Teacher turnover at Monmouth has hovered around the district average for a number of years, with no significant peaks or valleys. However, the number of applicants per opening at Monmouth lags far behind the district average. Monmouth may be perceived to be a very unattractive place to teach, or a place where new teachers go as a last resort to get experience-by-fire before they are able to leave for more attractive positions. Either way, Monmouth is unlikely to be able to pick and choose among candidates for open positions.

Funding Equity

Although the achievement gaps between groups of students has received a lot of public and professional attention recently, another gap exists with little comment: funding gaps among school districts and even schools within districts. Gaps between districts cannot be addressed by districts themselves, but disparities in funding among schools in the same district can be.

The Center on Reinventing Public Education has conducted detailed studies of budgeting practices in more than half a dozen major school districts over the past six years.[21] This research has revealed that there are substantial differences in the actual (as opposed to the budgeted) funding levels of schools within each of these districts. Surprisingly, these differences are largely invisible, not just to the public eye but to the eyes of district leaders also.

The differences fly beneath the radar of both district leaders and the general public because schools are "resourced," not funded. That is to say, they are provided with so many teachers depending on enrollment; they do not receive the funds to pay for teachers, much less the categorical funding that is intended for particular groups of students. Schools are not what accountants would consider to be "cost centers"; it is, therefore, difficult to determine exactly how much funding is going to each school.

When CRPE researchers looked into this, they found that disparities were related both to the way school districts budget for teachers (the single largest expense for schools) and the way funds do (or do not) follow the students for whom they are intended.[22] The result, in all the districts studied, was

that the most needy schools tended to receive lower funding per student than schools with lesser needs. In effect, schools with the most challenges were subsidizing schools with the fewest.

Funding Equity: Status and Change: The CRPE researchers developed two measures of funding equity that are used in the management guide. The first uses actual vs. budgeted teacher salaries, and the second uses a weighted index of resource allocation to compare expected funding to actual funding among schools.

The first measure (teacher equity) compares what the district budgets for teacher salaries in a particular school with actual teacher salaries in that school. That is to say, for each school this measure compares how the district "budgets" its money for teachers (almost always the district's average teacher salary multiplied by the number of teachers assigned to the school) with how it spends it (the real salaries of the teachers actually in the school). Some of the teachers in the school may make the minimum salary, while others are paid at the top of the scale, but all are budgeted at the same dollar amount. This CRPE-developed measure throws light on a finance fiction—namely that budgeting (or "resourcing") schools on the basis of average teacher salaries actually represents what is spent on teacher salaries in individual schools.

The CRPE researchers found that every district queried about the effects of average vs. actual teacher salaries was convinced that average teacher salary within the schools would closely match the average district salary.[23] That was not the case in even a single district studied. In fact, the disparity among schools within a district in real teacher salaries amounted to as much as $1 million gain in some schools. This could only be made up with corresponding losses to other district schools.

This result might not be of great significance if all teachers were of equal experience and ability. In practice what happens, however, is that very needy schools tend to be staffed largely with new and inexperienced teachers at the bottom of the salary scale. Once they get a few years of experience, they tend to take their increased capability to a more attractive school—or leave teaching altogether. In summarizing the effects, Roza and Hill conclude that "there is good reason to believe that schools with higher average salaries have more capable teachers."[24]

The second measure of funding equity, a weighted index of expected allocation, was developed to look at how student-based budgeting would affect school-level funding.[25] If support is attached to students rather than to buildings, a school that serves a great number of low-income, educationally vulnerable students should receive more funding than one that serves children of upper-income, professional families. The former students not only need more assistance than the latter, but districts can draw on specific sources of state and federal funds (e.g., Title I) to help these students.

Because the two Rebel Valley schools were among those studied by Superintendent Hernandez's academic colleague in Chapter 1, it was possible to identify the amounts these schools received in terms of teacher salaries, compensatory education, bilingual, and gifted/talented resources. On this second weighted index (like the first), there were large and persistent differences among schools, with the neediest schools often receiving less than they were entitled to if categorical program funds had been budgeted around students instead of schools.

For the management guide, both measures are shown together and compared to the "base" line. The "base" line is the amount the school would receive if teacher salaries and other funds are allocated according to actual teachers and actual students in the individual school. Few if any districts will have such data for more than a year or two, so the "trend" data are limited. However, two years of data are shown in Figures 3.11A and 3.11B for our sample schools. The figure clearly reveals the difference between anticipated and actual school funding.

The results can be quite dramatic. Guy Fawkes has a teaching force that appears to be at about the average salary level in the district. There are no self-evident equity problems in terms of teacher salaries at Guy Fawkes.

The situation at Monmouth could hardly be more different. The amount actually going to teacher salaries at Monmouth was only about 75% of the district average in 2002. The following year was even more dismal—Monmouth's salaries averaged only about 60% of the district average. The teachers at Monmouth are clearly the bottom of the salary scale (which is based entirely on experience and education, not performance).

With respect to per-student allocations, Fawkes and Monmouth are also at opposite ends of the spectrum. For any number of reasons, many of them buried in history, Guy Fawkes receives a much higher allocation per student than would be expected given its size and composition. In contrast, Monmouth, the school with greater needs, received only 80% of the funds to which it should have been "entitled" in 2002. Because the district has already begun to address this problem (see Chapter 1), some reallocations have already occurred and Monmouth received slightly more than its entitlement in 2003.

As with other indicators, the funding equity indicator cannot tell how a particular situation came about or how to address it, but the indicator can provide educational leaders with a tool that will give them a meaningful summary of complex data. What they do with these insights is up to them.
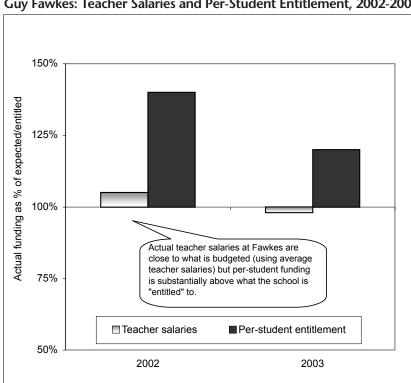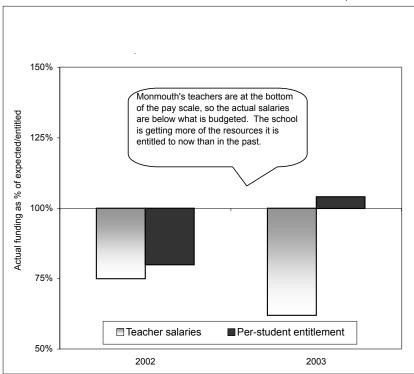
Figure 3.11A
Guy Fawkes: Teacher Salaries and Per-Student Entitlement, 2002-2003

Figure 3.11B
Monmouth: Teacher Salaries and Per-Student Entitlement, 2002-2003

Toward More Effective School Management Tools

The indicators described in this chapter and combined in the management guide described in Chapter 1 were selected because they offer great promise for school and district leaders. They hold out the hope of being able to cut through mountains of near-impenetrable data so that school officials (and members of the public) can find "actionable intelligence" to help improve student learning.

Most school districts are already collecting the data underlying these indicators. Much of the overwhelming volume of information is also available to the public, which is likely to find it even more difficult to comprehend than teachers and principals. The nation's school data problem today is not due to a paucity of data. Quite the contrary. The problem is that educators and parents are awash in data they find hard to understand. The management guide offered here promises to create a center of gravity for data usage, a focal point around which to organize data so as to identify both critical problems and promising opportunities.

---

[1] Robert J. Marzano, *A New Era of School Reform: Going Where the Research Takes Us.* (Aurora, CO: Mid-continent Research for Education and Learning, 2000). Marzano, working under contract with the U.S. Office of Educational Research, summarizes much of the research that has been done in recent years on factors affecting/indicating school effectiveness.

[2] The management guide presented here uses scale scores unadjusted for the racial or economic composition of the student body or the geographic location of the school. Howard Wainer makes a strong argument that using unadjusted scores creates a situation called the Simpson Paradox, where average scores for subgroups may actually be higher than the average for the group as a whole because the subgroups are of different sizes. He presents a detailed explanation of the paradox in several articles, including Howard Wainer, "On the Academic Performance of New Jersey's Public School Children: Fourth and Eighth Grade Mathematics in 1992," *Education Policy Analysis Archives*, Vol 2, #10 (July 11, 1994) and Howard Wainer and Lisa Brown, "Three Statistical Paradoxes in the Interpretation of Group Differences: Illustrated with Medical School Admission and Licensing Data" (paper submitted for publication in *The American Statistician*, 2004, http://www.statlit.org/PDF/2004Wainer_ThreeParadoxes.pdf).

When adjustments are made for the racial/ethnic constitution of the schools in the database used for the report presented here, the effect on average school scores is often considerable, with many of the differences equivalent to an effect size that would be classified as "high." Adjusting average school achievement scores for student body composition can have as much effect on *apparent* achievement as most educational interventions, amounting to as much as 20% of a standard deviation. More attention needs to be paid to this and other such weighted measures, including measures of the "efficiency" of schools suggested by Amy Ellen Schwartz and Leanna Stiefel in "Measuring School Efficiency: Lessons from Economics, Implications for Practice" in David H. Monk, Margaret Wang and Herbert Walberg, *Improving Educational Productivity*, (Greenwich, CT: Information Age Publishing, 2001). Such weighted measures of school effectiveness are still in the development stage and can cause controversy that confuses rather than furthers the efforts of leaders to grasp what is going on in their schools. We hope to address the use of weighted measure as we do further development on the management guide presented in this report.

[3] Richard Rothstein, "Lessons: Testing Reaches a Fork in the Road," *New York Times* (May 22, 2002).

[4] In fact, Washington State recently changed the cutoff points for 4th and 7th grade WASL scores in math and reading because of concerns that they had been set too high. Thus, without any major changes in the actual test performance of students, there was a significant jump in the number and percentage of students who met the standards at both levels and in both subjects. See Linda Shaw, "Improvement on WASL Carries Asterisk," *Seattle Times* (September 02, 2004).

[5] For example, correlations among the scores on the Washington Assessment of Student Learning (WASL) were very high for reading-math (around .76), but much more modest for listening-writing (around .40). Correlations were very similar across the grades tested (4th, 7th, and 10th) and across years of testing (1998 to 2003).

[6] Mark S. Handcock and Martina Morris, *Relative Distribution Methods in the Social Sciences* (New York: Springer-Verlag, 1999).

[7] Thomas J. Kane and Douglas O. Staiger, in *Improving School Accountability Measures*, (National Bureau of Economic Research Working Paper 8156, March 2001), emphasize the imprecision of school-level test score means. They estimate that "28% of the variance in 5th grade reading scores is due to sampling variation and about 10% is due to other non-persistent sources and that less than half of the variance in the mean gain in reading performance between 4th and 5th grade is due to persistent differences between schools." Based on their study, the authors estimate that the confidence interval for the average 5th grade reading score in a school with 60 students per grade level would extend from roughly the 25th to the 75th percentile!

If such volatility exists for other criterion-referenced tests (and there is no reason to believe it doesn't), then relying on a single year or year-to-year data at the school level would be risky at best, and probably misleading. Needless to say, not all statisticians agree with Kane and Staiger's conclusions. One of their major critics is David Rogosa of Stanford. His arguments are presented in a number of publications, including D. R. Rogosa, "Irrelevance of Reliability Coefficients to Accountability Systems: Statistical Disconnect in Kane-Staiger 'Volatility in School Test Scores'." (CRESST, November 2002). Available at www-stat.stanford.edu/~rag/api/kscresst.pdf.

[8] Olson, Lynn. "Testing Experts Develop New Method of Presenting Achievement-Gap Data." *Education Week* (March 13, 2002.) Ms. Olson was reporting on a method of graphical presentation/analysis suggested by Paul W. Holland of the Educational Testing Service. A more thorough discussion of his methods is presented in Paul W Holland, "Two Measures of Change in the Gaps between the CDFs of Test-Score Distributions" (Center for Statistical Theory and Practice, Educational testing Service, January 11, 2002.)

[9] In fact, a Wallace-funded study of urban school superintendents found that a majority of those who see the achievement gap as "a critical and chronic challenge" believe that the programs the district has in place are capable of closing the gap within five years. Howard Fuller, et al., *An Impossible Job? The View from the Urban Superintendent's Chair* (Seattle, WA: Center on Reinventing Public Education, 2003).

[10] Elise M. Huggins and Mary Beth Celio, *Closing the Achievement Gap in Washington State: Holding Schools Accountable for Equity*. (Report for the Washington State Academic Achievement and Accountability Commission, November 5, 2002).

[11] Paul Holland, "How Big is Big when it Comes to Gaps in Scores?" (ETS Occasional Paper, January 2002) and James McMillan, "Standards-Based Accountability: Measuring Yearly Progress" (CEPI Briefings, http://www.edpolicyvcu.org/stanbaseacctpol.htm, 2000).

[12] Mary Beth Celio, *Falling Through the Cracks: A Study of Dropouts in the Seattle Public Schools*. A Report for the Seattle-King County Economic Development Council (September 20, 1989.) National Research Council, *Understanding Dropouts: Statistics, Strategies, and High-Stakes Testing,* Committee on Educational Excellence and Testing Equity (2001). Alexandra Beatty, Ulric Neisser, William T. Trent, and Jay P. Hebert, Editors. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington D.C.: National Academy Press, p. 13. National Association of School Psychologists, National Mental Health and Education Center, "High School Dropouts Cost Everyone Something!", Leslie F. Hale, Ed.D.,(Bethesda, Maryland: NCSP, 1998); Matthew J. Taylor and Philip L. Rodgers, *Increasing Graduation Rates for Minority & Other At-Risk Students: The High School Completion Study*, Center for the School of the Future, Utah State University (February 4. 2002.)

[13] DeKalb, Jay, "Student Truancy." *ERIC Digest*, Number 125 (ED429334 April 1999). See also Eileen Garry, *Truancy: First Step to a Lifetime of Problems*. (Washington, DC: Office of Juvenile Justice and Delinquency Prevention, October 1996. ED 408 666); Lee Robins and Kathryn Ratcliff. *Long Range Outcomes Associated with School Truancy*. (Washington, DC: Public Health Service, 1978. ED 152 893); J. Finn, "Withdrawing from School." *Review of Educational Research* (1989, 59(2): 117-42).

[14] A.S. Bryk and Y. M. Thum, "The effects of high school organization on dropping out: An exploratory investigation." *American Educational Research Journal,* 26(3): 353-83m (1989). M. Fine, "Why urban adolescents drop into and out of public school" *Teachers College Record*, 87:393-409 (1986). J. D. Finn and K. E. Voelkl, "School characteristics related to student engagement." *Journal of Negro Education* 1993, 62(3). M. K. Johnson, R. Crosnoe and G. H. Elder, "Students' attachment and academic engagement: The role of race and ethnicity." *Sociology of Education* 74: 318-40 (2001).

[15] Jon Douglas Willms, *Student Engagement at School: A Sense of Belonging and Participation*. (Results from Programme for International Student Assessment 2000). Report from the Organisation for Economic Co-Operation and Development, 2003. The two measures used to assess belonging in the PISA were based on responses to six items describing such things as the students' personal feelings about sense of belonging, acceptance by peers, and support from teachers along with frequency of absence, class-skipping and late arrival at school during the two weeks prior to the survey. Measures used in other studies include time spent on homework, participation in classroom discussions and involvement in sports and other extra-curricular activities. Because of the need for cross-cultural applicability, these potentially more sensitive/meaningful measures were not used and absenteeism was used as being "the most important aspect of participation" (p. 18). http://www.pisa.oecd.org/Docs/download/StudentEngagement.pdf.

[16] Willms, 2003, p. 8.

[17] A complete description of each state's student accountability system is provided in Mary Bourguignon and Mary Beth Celio, *Dropout & Graduation Rates and Academic Accountability in Washington State: A Report to the Washington State Academic Achievement & Accountability Commission*, September 2002. http://www.k12.wa.us/accountability/Reports/GRADUATION%20FINAL%20092402.pdf.

[18] See Linda Seebach, "Graduation rates may be worse than agency reports," *Detroit News* (February 24, 2002); Christopher B. Swanson, *Keeping Count and Losing Count: Calculating Graduation Rates for All Students under NCLB Accountability* (Washington, D.C.: The Urban Institute, Education Policy Center, 2003); Jay P. Green, *High School Graduation Rates in the United States* (Report prepared for the Black Alliance for Educational Options, Center for Civic Innovation at the Manhattan Institute, Revised April 2002).

[19] No external characteristic of the teacher—not years of teaching experience, type of certification, or having a major in the subjects taught—has been shown to be consistently or reliably related to significantly greater student achievement. Such a conclusion has been reached through meta-analyses of a large number of teacher effectiveness studies conducted by researchers including Kate Walsh, *Teacher Certification Reconsidered: Stumbling for Quality* (Baltimore, Md.: The Abell Foundation, 2001. Available at http://www.abell.org) and Dan Goldhaber with Emily Anthony, *Teacher Quality and Student Achievement* , ERIC Clearinghouse on Urban Education, Urban Diversity Series No. 115, May 2003. Goldhaber and Anthony's most recent research in this area, *Can Teacher Quality Be Effectively Assessed?* (Center on Reinventing Public Education and the Urban Institute, April 7, 2004. Available at: www.crpe.org/workingpapers.shtml#quality) found small but statistically significant advantages to students taught by teachers with a certificate from the National Board for Professional Teaching Standards, but while the research assures that the NBPTS is certifying the applicants who are somewhat more effective, this research does not provide a basis for judging teacher effectiveness apart from the scores earned by students. A recent issue of the *Journal of Educational and Behavioral Statistics,* Volume 29, Number 1 (Spring 2004), devoted entirely to value-added assessment as it is used to measure teacher quality, revealed both the complexity of assessing teacher quality from student achievement gains and the level of controversy surrounding almost every aspect of the process. Education researchers are far from identifying or agreeing on a measure of teacher effectiveness. However, if such a measure becomes available, it will certainly be incorporated into this management guide. Until that occurs, however, teacher attraction and retention are suggested as surrogates.

[20] Marguerite Roza, "Chapter 8: Rethinking Data Capacity" in *Making School Reform Work: New Partnerships for Real Change*, Paul T. Hill and James Harvey eds. (Washington, D.C.: Brookings Institution Press, 2004, p. 105).

[21] Publications from these efforts include, among others, Marguerite Roza with Karen Hawley Miles, *A New Look at Inequities in School Funding*, (CRPE, May 2002) and Marguerite Roza and Paul Hill, *How Within-District Spending Inequities Help Some Schools to Fail* (A Draft Conference Paper for the Brookings Conference on The Teachers We Need, CRPE, May 2003).

[22] The CRPE researchers found that few districts have developed the capacity to track real-dollar spending on per-pupil basis, using real teacher salaries, but they noted that the necessary data management and computational methods have been published in a toolkit by the Annenberg Task Force for School Communities that Work. This toolkit is designed to help analyze district data and is not overwhelmingly technical.

[23] Roza and Hill, 2003, p. 10.

[24] Ibid, p. 9.

[25] Karen Hawley Miles, Kathleen Ware, and Marguerite Roza, "Leveling the Playing Field: Creating Funding Equity Through Student-Based Budgeting" *Phi Delta Kappan*, Vol 85, no 2, (October 2003 (www.pdkintl.org/kappan/k0310mil.htm)); Larry Miller, Marguerite Roza and Claudine Swartz, *A Cost Allocation Model for Shared District Resources: A Means for Comparing Spending Across Schools* (CRPE working paper, September 2004); Marguerite Roza and Karen Hawley Miles, *Understanding Student-based Budgeting as a Means to Greater School Resource Equity* (CRPE working paper, July 2004).

Chapter 4   Summary and Implications

This report began with Samuel Johnson's observation about the effectiveness of the hangman in concentrating the attention of the potential hangee. Dr. Johnson also opined that "Example is always more efficacious than precept." It is this principle upon which this work is based. Precepts, exhortations, and advice, even if based on research and presented with maximum conviction, will not be as effective in changing minds or systems as an example of how the precept can be put into action. The availability of masses of detailed information about every aspect of schools, collected in the hopes of providing useful direction for education leaders has, all too often, led to paralysis by data. A time-bound and curriculum-driven school system is now facing dramatic demands for reconstruction under the pressures of accountability, and data collectors and analysts have risen to the opportunities. Evaluation is here to stay. Educators need to find a way to make these pressures and the data useful to them and to the students they serve.

In their search for easily digestible assessments of school quality, states and federal agencies have reduced all the wealth of data to a data point or two. Crude, bottom-line comparisons are rarely helpful and frequently unsettling. With indicators that go beyond simplistic comparisons, school leaders can pinpoint problems and develop concrete action plans tailored to the individual challenges of individual schools. Properly employed, assessment of the system and its components can become more diagnostic and useful, less unsettling and punitive.

Several implications stand out from the work outlined in this paper:

Less may be more. Pressure to respond to every possible question and meet all potential attacks from public or press has often led districts to pull together every conceivable piece of information for policy consideration. But it is clear that one of the major reasons indicator systems fail before they are given a chance (in schools and elsewhere) is the volume and complexity of the data developed. The human capacity to absorb information is, of necessity, limited. Indicator systems should respect that reality.

The principles of parsimony and power should be respected. The temptation to develop 17 indicators, or 127 different pieces of information capable of satisfying everyone with a question about anything in every individual school, will be nearly irresistible. Like all temptations, it probably should be resisted. The indicator system must focus on a parsimonious set of powerful indicators of school performance if it is to act as a management

guide. Different data systems are probably required to provide audit trails or to respond to parental requests for information. If the indicator system is expanded to serve other needs, it will likely founder. The key to success will lie in parsimoniously selecting a few indicators and judging them against the standards of data, proxy, and communications power.

Current status data is necessary but not sufficient. If the intent of leaders is to pinpoint potential areas requiring policy and management attention, it is essential to look at trend data, as well. Quite apart from the diagnostic value of trend data, the public's interest in accuracy, credibility, and fairness is also better served with information grounded in time. Without such a perspective, the public may be easily confused by year-to-year pronouncements about how well (or how poorly) things are going.

Smart use of data holds the potential for dramatically altering the tone and quality of board-superintendent relationships. It is no secret that in many districts, much of the dialog between the central office and the board is characterized by wariness and skepticism, if not distrust, about central planning issues. In this environment, broad assertions of the nature of the challenges faced by local schools often substitute for careful analysis. Data sets that identify problems and promise to "get at" real issues on a school-by-school basis offer district leaders what all of them want—the opportunity to target scarce resources where they can do the most good.

Targeting resources where they can do the most good requires better funding indicators. Such indicators are being developed. Most large districts probably have the data required to produce school-by-school funding indicators. They may not know how to get at it, how to analyze it, or how to use it, but they have the raw data. If district leaders genuinely seek strategic use of limited resources, it is essential that they stop "resourcing" schools and start thinking about the real dollar amounts spent in each of them.

Currently, teacher attraction and retention are the best proxies we have for teacher effectiveness. The same might be said for school principals. Reliable indicators on teacher and principal quality are hard to come by, and this lack greatly hampers our ability to measure their impact in any given school. Research is continuing in this vital area, but proxies will have to be used until better measures are available.

The seven school-level indicators outlined in this paper are a solid jumping off point for any district. They are well grounded in research and experience. As with similar metrics, each district will have to decide for itself the extent to which any or all of these seven measures fit its particular needs and circumstances and the extent to which the district has the capacity to gather and analyze the data required for the measures. Grounded as they are in research and experience, these indicators are a good deal more than simple suggestions, but they do not represent commandments engraved in

stone for all districts in every circumstance. In fact, any suggestion that they do would effectively pre-empt the very important internal district discussion required to help districts understand the significance of the indicators and the value and utility of metrics tied to time and progress.

Professional development and technical assistance around data usage will be required. In recent years, leadership sophistication about data usage has increased dramatically. Still, effective use of data as a management tool will undoubtedly require additional professional development or technical assistance. Districts need to ask themselves a set of questions about their capacity for this work:

- What set of parsimonious indicators can we use? (One set is suggested here; others are possible.)
- What data are required?
- Where can we find them?
- Who will do this work? (Can we find the expertise we need internally? Do we turn to local universities? What technical assistance can the state provide?)
- How can these indicators be incorporated into the consultation and decision-making processes of the district?

State leaders have a significant role to play. In considering the development and use of an indicator system such as the one described here, the role of state leaders becomes the role of leadership everywhere: pointing people in the right direction, providing political cover, and helping districts move along. This implies a need for state leadership action that will (1) support the development of actionable intelligence in the form of indicators; (2) educate the public and federal officials about the traps and snares that lie in wait for punitive, "just the facts" accountability schemes; (3) develop model "management guides" to help local districts; (4) provide technical assistance to districts trying to develop their own management guides; and (5) continue to press for appropriate forms of assessment for individuals, schools, and systems.

Going the Extra Mile

Formal systems theory distinguishes between "single-loop" and "double-loop" thinking.[1] Single-loop thinking typically involves doing something routine, doing it properly, and fixing problems quickly. Double-loop thinking is designed to get at more fundamental challenges within systems and worries more about defining new ways of operating, doing the proper thing, and taking the time to do it right.

Despite an abundance of promising assessment and accountability models, in many ways state and local school systems are in danger of being trapped in

a single loop. Everyone understands the dangers of quick fixes, particularly around assessment, but public pressures force just about everyone into the pattern. Why? Because more fundamental solutions, double-loop solutions, require investment, time, and attention. Most of all, double-loop solutions are slower to produce results; we cannot be sure of the results, and we cannot endure the delay while waiting for the results to improve.

By relentlessly focusing on results, providing measures of status, and tracking developments over time, the indicator system outlined in this report may help educators go the extra mile required for more fundamental solutions. Unless educators are willing to put in the hard work required to make these indicators real, they are likely to continue to receive criticism on the basis of measures developed to satisfy the urge for quick fixes. The work outlined in this report suggests we can do better.

---

[1] See, for example, Peter M. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization*. (New York: Doubleday, 1990). and Nelda Cambron-McCabe et al., *The Superintendent's Fieldbook: A Guide for Leaders of Learning* (Corwin Press, 2004).

Acknowledgments

The authors want to acknowledge the contributions of many people who made our work possible.