Chapter 3

Seeking the Magic Metric: Using Evidence to Identify and Track School System Quality

MARY BETH CELIO
Northwest Decision Resources

ABSTRACT: *With the increasing demands for accountability, many educators are searching for a single measure that will inform them about how well their system is serving students-a magic metric. Unfortunately, there is no magic metric that can guide educators decision- making. This chapter reviews three common measurement systems that have developed over the last decade and describes what they can do for educators and their limitations. A set of indicators that can guide educators through the myriad of data that schools collect is described. Finally, an indicator system is proposed that measures the important features of an educational system. The data from this system can be used to guide educator decision-making as they seek to improve educational services for all students*

The problems confronting educational institutions today, at a time of massive budget cuts, are myriad and complex. Increasing pressure for accountability adds to the challenges facing school superintendents, school boards, governors, and the millions of parents who are concerned about the educational future of their children.

Amid the chaos that often characterizes debates about public education, the natural desire for a silver bullet—a single intervention or a policy or program that can trigger all the necessary changes that will improve the system for all children—is strong. Clearly, some politicians and educators thought No Child Left Behind (NCLB) might offer a version of the silver bullet: that by setting ambitious goals, offering incentives and consequences, and tying these efforts to standards-based testing, the federal government could focus local efforts and effect change. Few people who have any familiarity with NCLB retain these early hopes.

Although NCLB seems to have failed in many of its objectives, it has brought into sharp focus a gap in the conversation about educational reform: There is no universally accepted measuring rod for gauging success that can be used to compare individuals, classrooms, schools, districts, and states. What is missing is a metric similar to those in use in other sectors of society.

In the business world, a metric is any type of measurement used to gauge some quantifiable component of a company's performance. No one metric is universally accepted across industries, but some metrics are ubiquitous because they tell managers and stockholders about the health and performance of companies from small shops to major conglomerates: return on investment, employee and customer churn rates, revenue versus debt. Other metrics summarize important economic factors: the unemployment rate, the cost of living index, a borrower's credit score, a country's GNP. What all of these metrics have in common is that they summarize key aspects of a entity (person, state, industry, company, or nation) in a single number that lets the observer know the relative success of that entity on a scale.

The established metric for NCLB—adequate yearly progress—quickly lost its credibility and value. A school that didn't meet the criterion could be put on probation, restructured, or even closed permanently. Another commonly used metric, graduation or dropout rates, has the virtue of measuring a commonly accepted benchmark, but such rates are a summary measure, not a progress report. State-mandated standards-based tests at various grades can work as a metric within a given state, but those tests differ significantly from state to state. The National Assessment of Educational Progress (NAEP) allows for state-by-state comparisons, but they aren't available below the district level.

As is true for health care and other industry leaders, education leaders at all levels want an effective metric that can measure how well they are doing and monitor their improvement across time: a metric that captures the key elements of an institution in a concise and compelling way and points toward a goal that is generally accepted, if not perfectly articulated. Even more, they want what might be termed a "magic metric"—a universally accepted metric that is brief, understandable, and measurable, but one that can be used with units in a system of similar units (e.g., retail stores or restaurants in a chain, hospitals in a network, high schools in a district) as well as the system as a whole.

Three examples of magic metrics currently in use are presented here in order to highlight the advantages and drawbacks of each. Two are state-specific metrics and the third was developed for a national study. The first state-specific index of school quality and improvement, one that has been in use for about a decade, is the California Academic Performance Index, or API. The second state-specific metric is the more recently initiated Washington Achievement Index. The third metric is used in a national study called Return on Educational Investment.

API calculates a single score for each school in California using that school's test results on the state's learning standards as they compare to prior year test results. The state's goal is a score of 800 for each school, but the actual scores in 2011 ranged from 200 to 1,000. Although API definitely has the virtue of being universally applied and concise, one parent information website, GreatSchools, noted that "Educators and parents alike struggle to understand where the API comes from, how it's calculated and what exactly it means." In fact, all it can convey is the relative position of a given school on a scale determined entirely by scores on a single set of tests. The bottom line, according to the same website: "The API is based on test scores and is calculated in a way that encourages schools to raise the test scores of the lowest-scoring students." However, nothing about the scores or their display (a list of scores in similar schools, districts, or counties) can provide any information to decision makers on what elements of school or student behavior need to be addressed to improve those scores.

Washington state has attempted to expand the usefulness of an API-type index through adoption in 2009 of the state legislature-mandated Washington Achievement Index. It is currently used to select schools for the Washington Achievement Award. The index measures how all schools and districts are performing in five key areas (reading, writing, math, science, and graduation rate). Each of the key areas are evaluated across 4 indicators (achievement by non-low income students, achievement of low income students, achievement vs peers, and improvement from the previous year). The five outcomes and four indicators result in a five by four matrix with 20 measures. Each cell of the matrix contains a score, and the index is the average of the ratings across the 20 outcomes. By using the average, schools without data for some indicators are still included in the system and a separate system is not needed for different types of schools. The Washington state index retains the benefit of a single index number (the average of scores, ranging from 1–7) and the additional virtue of including more than a single outcome measure. However, it is, again, a lagging indicator—telling us something about outcomes but providing nothing that can help to identify which aspects of student or school performance need to be addressed. For more details about the Washington Achievement Index, please visit their website (http://www.sumner.wednet.edu/studentfamilyservices/academics/pages/achievementindex.html).

The final example of a potential magic matrix was published by the Center for American Progress (Boser, 2011). Called the Return on Educational Investment, the method calculates how much learning a district produces for every dollar spent, after controlling for factors such as cost of living and students in poverty. A vast majority of public school districts in the United States were evaluated using this method, and the interactive display mechanism provides viewers with the ability to see a district's basic ROI (return on investment) compared to other districts in the state. The color-coded evaluation matrix is used on both a state map and a matrix plotting the state achievement index against adjusted per-pupil spending (to see this interactive map, please go to http://www.americanprogress.org/issues/education/news/2011/01/19/8877/interactive-map-return-on-educational-investment/). This makes it possible to easily identify particularly effective or ineffective districts.

The education ROI provides a metric that has an appeal beyond either of the state-specific metrics discussed earlier, both because it takes into account factors that aren't directly addressed elsewhere (like cost of living and level of poverty) and uses graphic displays that illuminate relationships and encourage further investigation. The biggest disadvantage with this particular measure is that it is designed for whole districts rather than individual schools. Focus on the district level is largely due to the fact that most school districts do not calculate per-pupil spending for individual schools, clinging to the comfortable fiction that all schools in a district have essentially the same resources (see Keyworth, Detrich, and States, this volume, for an analysis of the inequities across schools within a district). However, using methods pioneered by researchers at the Center on Reinventing Public Education at the University of Washington, it should be possible to develop such per-school numbers, thus producing an ROI for individual schools (Roza & Miles, 2002; Roza & Hill, 2004).

While being more intuitively and visually appealing than the other education metrics discussed, the ROI shares a major drawback. The measure can identify districts and potentially schools that are at great risk of failing their students, but it can provide little to help decision makers at the state, district, or school level to get beyond a score that might be considered a final grade. It shows where action is needed; it doesn't provide the information necessary to diagnose the underlying weaknesses and to intervene and put the schools back on track.

IF NOT A MAGIC METRIC, THEN WHAT?

Clearly, a metric that assigns a final grade or acts as a lagging indicator can provide valuable triage information, identifying which schools are in the greatest distress and which are progressing well on their own. However, data for the next steps—intervention and remediation—must be immediately available. What would be helpful at this stage are indicators that identify schools or districts failing to meet the educational needs of their students and also provide decision makers with the information they need to address the issues that led to the failing grade. Unfortunately, these next steps have often involved the collection and presentation of masses of scattershot and unfocused data on district websites and district-produced school report cards. Numbers collected and presented in this way provide little guidance about what the information means and what to do with it. The result, as one study of district data needs and uses (Roza, 2004) puts it, is that "Most urban cities lack the strategic information to successfully identify and implement a district reform strategy."

This paper describes a set of indicators that can serve as a guide through the data wilderness and can help school officials and community members make sense of the mountains of data. It is not so much a cookbook as it is a blueprint to action, and it begins with a more refined definition of "indicator." Innes (1990; see also Norris, Atkisson, et al., 1997; Innes & Booher, 2000) has described an indicator as "simply a set of rules for gathering and organizing data so they can be assigned meaning." They 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. So 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 hospital, or a school system—indicators can provide a general sense of how well the system is functioning. The trick is to find educational indicators that have meaning, are easy to read, have been validated by research as related to student learning, and can be presented comprehensibly in a graph or chart or in a page or two of text rather than in a volume.

In developing a set of indicators of school quality and improvement, Celio and Harvey (2005) adopted several basic principles that are explained below:

Indicators should be neither top-down (developed by experts or those in leadership) nor built from the grassroots but rather evidence-based.

There is a long history of conflict between advocates of the two traditional approaches to indicator development. Community organizations and city governments have conducted listening sessions and focus groups to identify indicators of effectiveness in different areas. On the other hand, government agencies have developed metrics that don't always make much sense to those who have to live with them (e.g., adequate yearly progress.) This question is actually moot; what is important is whether the indicators are linked by reliable evidence to the quality being sought. In other words, does the indicator measure what it is intended to measure.

As far as indicators are concerned, experience has shown that less may be more, but one is not enough.

The single metric (magic or otherwise) cannot possibly provide enough information beyond that needed to identify the schools most in need of attention. A single metric, no matter how attractive, makes it difficult to understand its implications or motivate to action. On the other hand, schools and school systems are now awash in data. A school or district report card made up of dozens of data items for dozens of subgroups of students confuses rather than enlightens; the human mind has a limited capacity to absorb unlimited data if the data are not organized in a meaningful way.

Parsimony and power must be respected.

The number of indicators can spiral out of control when developers try to cover all bases and please all stakeholders. Success rests in parsimoniously selecting a limited number of indicators and judging their power to communicate useful information plainly and succinctly.

Current status data are necessary but not sufficient.

Knowing the graduation rate of a particular school can tell you how a single group of students at that school fared, but not much else. The data are out of context. There are, of course, year-to-year fluctuations and these can confuse matters, but trend data are crucial for understanding the overall trajectory of the particular institution. Is the trend generally positive? Negative? Stable?

Proxies for key elements such as adequacy of funding or teacher effectiveness are inevitable.

It would be wonderful if school and community leadership could have immediate access to information about school-level factors that have been found to affect school quality (e.g., school culture, the effectiveness of the teachers, and the connectedness of students to the school). However, such information is either not readily or not consistently available in most districts. If an indicator has been shown to be connected to school quality but is not readily available, then it might be necessary to identify proxies for that indicator. Areas for which there are no universally accepted indicators cannot be excused from assessment and reporting for that reason.

Presentation cannot be an afterthought.

Every school board member, principal, and teacher is familiar with the reams of computer output supposedly designed to convey critical information on how students and schools are doing.

Often these reports are an excellent illustration of the truth of an observation made generations ago: "Getting information from a table is like extracting sunbeams from a cucumber" (Farquhar & Farquhar, 1891, cited in Wainer, 2004).

The heightened emphasis on school accountability, along with the realization that matrices and tables seldom have the desired impact has caused software developers to flood the market with tools for the conversion of input and output data to dashboards, sample report cards, and other display mechanisms. Many of these displays are very colorful but are not easily readable or understandable. The plethora of what Edward Tufte (1983) called "chartjunk" has inspired volumes to educate the reader in effective graphic techniques.[1] These books concentrate on presentation of content, making it clear that the way information is presented is critical to its usefulness.

THE BIRTH OF AN INDICATOR SYSTEM

Outside medicine, few fields are subject to such intense public analysis as education. Given the sheer volume of data about schools and the hundreds of articles published each year detailing evidence of school effectiveness or ineffectiveness, 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 (Marzano, 2000) In fact, the proposed indicator system grew from an extensive study of the literature on school effectiveness and reform. It also included research into the ways in which school district leaders use data in making decisions. It was informed by analyses of how leaders in other areas of community life try to understand how the public institutions for which they are responsible measure up against public institutions elsewhere. Finally, it was designed to assist leaders of individual schools, school districts, and state education systems improve school management. The foundation of the indicators is work completed over a 6-year period at the University of Washington's Center on Reinventing Public Education, much of it supported by

[1] Edward R. Tufte coined this term in *The Visual Display of Quantitative Information* and expanded on the concept in later books. A good example of a workbook designed to correct for the most egregious designs is Stephen Few's I*nformation Dashboard Design: The Effective Visual Communication of Data*.

the Wallace Foundation. This work focused on the achievement gap, national and statewide dropout statistics, national studies of school superintendents and principals, and extensive work on school and school district reform, including school finance.

Based on the work described above, seven indicators were selected, for which both status and change measures were identified.

1. Student achievement (scores on standards-based math and reading tests);
2. Elimination of the achievement gap (status and change in reading and math achievement for 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/district to attract students where there is opportunity for choice by parents/students);
4. Student engagement with school (proxy measure of school engagement, including attendance, tardiness, and involvement in school activities);
5. Student retention and completion (retention of students during the school year and completion of the requirements appropriate at each school level: elementary, middle, and high);
6. Teacher attraction and retention (proxy measure of teacher attraction using applications per job opening and non-retirement turnover); and
7. Funding equity and efficiency (proxy measure using amount of funding per student expected by policy and amount actually received; return on investment using calculated per-student funding).

A sample of the display mechanism designed for this indicator system is presented on the next page which is not unlike the displays used in consumer product review publications. Each school level in a district (in the example, all schools in the fictitious Rebel Valley School District) is rated on status and change in the seven areas. The status indicators are a snapshot of how the schools perform right now in comparison to a comparison group, in this case, other middle schools in the district.

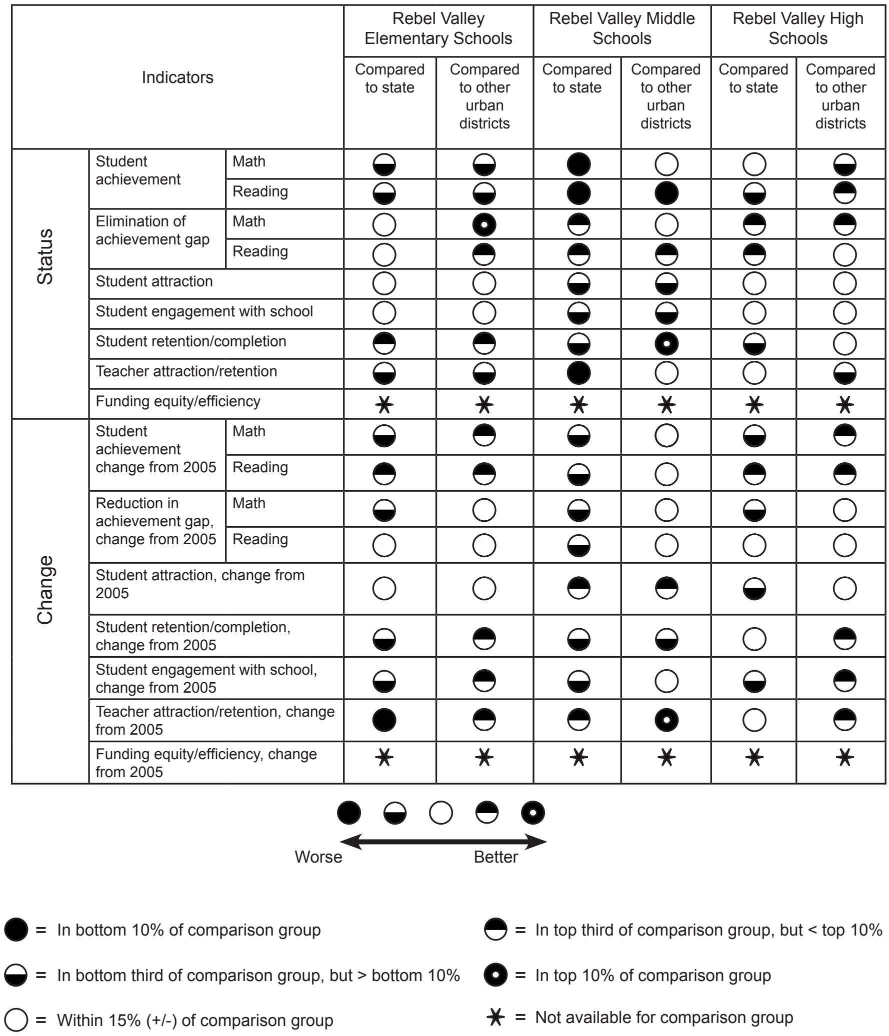

Figure 1. Sample indicator system.

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

Student 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 like widgets on a production line. Whether a bane or a boon, test scores are essential to any indicator system (Wainer, 1992; Wainer & Brown, 2004).[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 (*New York Times*, 2002) made an impassioned plea for using scale scores in reporting criterion-referenced test performance, noting that the cut-point used to determine the standard is simply a predetermined point on the scale score distribution, not a magic number. Thus, moving the cut-point in one direction or another could make a radical difference in the percent meeting standards (Shaw, 2004).[3]

A scale score is neither the raw score a student earns (i.e., the number of correct answers) or a percentage of correct answers. It is a number on a scale that is derived from the raw score but takes into account differences on the forms of the test students take. A well-known example of scale scores is the SAT. For both verbal and math portions of the test, the scale runs from 200 to 800, and the two scale scores are added together to get a total SAT score. Using scale scores and research on what different scale scores mean in terms of acquisition of required knowledge and skill, most states set two or more cutoff points along their scales, with the most important division being between students who are considered 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

---

[2] The indicator presented in this report uses scale scores unadjusted for the racial or economic composition of the student body or the geographic location of the school. Wainer has made a strong argument that using unadjusted scores creates a situation called the Simpson Paradox, in which average scores for subgroups may actually be higher than the average for the group as a whole because the subgroups are of different sizes. 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.

[3] In fact, Washington state changed the cutoff points for fourth- and sixth-grade Washington Assessment of Student Learning (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.

labeled “meets standard” and “exceeds standard.”

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 line. Using scale scores permits educational leaders to understand where their students stand as they monitor their efforts to improve achievement or close the achievement gap. “Percent meeting standard” provides no such guidance. Scale scores also permit leaders to detect change over the entire range of scores. 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 indicator system 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, physical sciences, social sciences, and so forth. The correlation among the scores is very high and statistically significant.[4] It is conceivable that one of the two scores might be used by itself. However, reading and math scores together are the scores most generally accepted as meaningful.

Status*:* The circles on the display in the achievement row represent scale score data from schools based on what is known as relative distribution and density analysis. Hancock and Morris (1999) explicated in detail this method of analysis and presentation, which was developed specifically to study and report on the achievement gap between groups in society with special attention to income variations. They wanted to provide a full picture of the distribution of different measures rather than simple 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 incomes of male and female employees), it does not rely on the assumed distribution of scores as represented by the standard bell-shaped curve.

Change: Rate of change analysis provides information on how scores have changed over time, in our example, over 5 years. Changes from year to year are likely to be highly unstable but potentially indicative of progress toward

[4] For example, correlations among the scores on the WASL were very high for reading-math (around 0.76), but much more modest for listening-writing (around 0.40). Correlations were very similar across the grades tested (4th, 7th, and 10th) and across years of testing (1998–2003).

academic achievement across the spectrum of students (Kane & Staiger, 2001).[5] 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. In this example, 2005 is the base year. Each subsequent year shows the percentage of change from 2005.

Elimination of the 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 5 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 a sizzling hot issue. Educators need to see what is happening with respect to the achievement gap, both at the district level and within individual schools.

The achievement gap presented in this chapter defines the gap in terms of racial and ethnic groups. Those descriptors were the only student-level data available in the fictitious school district being analyzed. However, some districts also collect information on free or reduced-price lunch status and family composition (e.g., single parent household) that could be used to analyze student achievement. The same approach could be taken, independent of the descriptors used.

Status: The indicator system uses an analytic, graphic method based on the cumulative distribution of individual student scores—a method of presentation often used in such areas as medical research, marketing, and insurance. 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 (Olson, 2002).

Change: The gap status that can be displayed in the graph recommended by Olson (2012) shows a particular moment in time. This functions to inform educators how a specific school is performing relative to a comparison school. There are three possible outcomes from these data: The school of interest can

[5] In *Improving School Accountability Measures*, Kane and Staiger emphasized the imprecision of school-level test score means. They estimated 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, I estimate that the confidence interval for the average fifth-grade reading score in a school with 60 students per grade level would extend from roughly the 25th to the 75th percentile!

narrow the achievement gap with the comparison school; there is no difference between the two schools; or the school of interest is outperforming the comparison school.

Student attraction

Makers of toothpaste, producers of TV shows, and designers of teen clothing all conduct extensive research into what makes a product attractive to potential consumers. 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 knowing what aspects of a school are attractive to its target market is to look at families' choices when choices are available. Many public school districts now offer some level of choice for parents, ranging from magnet or alternative schools that are open (usually by lottery) to all students in the district to permitting parents to rank their school choices from among all schools within the system. Where some level of choice is available, the indicator system uses an indicator called "student attraction," which differs in definition according to district policies.

Status and change: In a school district where parents and students can designate any school in the system as first, second, or third choice and where the choices are relatively equal (e.g., every middle school is equally attractive), then the percentage of students should be comparable to the capacity of the school. Therefore, if a school has the capacity to enroll 20% of the sixth graders in the district, and all schools have the same ability to attract students, the district could expect that about 20% of the incoming sixth graders would choose that school.

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. Districts without an established policy of school choice can learn a great deal by comparing the number of children living within a school's catchment area with the number of students actually enrolled 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 appropriate schools for their children.

Student engagement with school

One of the most potent behavioral predictors of failure in school and subsequent dropping out is attendance patterns (Celio, 1989; National Research Council, 2001; Hale, 1998;). In a summary of national research on the issue, DeKalb (1999) reported the following:

> 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 (Garry, 1996). 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."

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 (Bryk & Thum, 1989; Fine, 1986; Finn & Voelkl, 1993; Johnson, Crosnoe, & Elder, 2001). An international study conducted by Programme for International Student Assessment (PISA) in 2000 was designed to assess the sense of belonging and participation of students in school.[6] In a report of the study, Willms (2003) noted that the term "engagement" as used in international research refers to "the extent to which students identify with and value schooling outcomes, and participate in academic and non-academic school activities."

Status and change: As with most assessments of engagement, the primary source of data for the PISA study was self-reported 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 possess such data, especially data that make it possible to identify trends over several years. The indicator system, 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

[6] The two measures used to assess the sense of belonging were based on responses to six items describing the students' personal feelings about belonging, acceptance by peers, and support from teachers, along with frequency of absence, class skipping, and late arrival at school during the 2 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 extracurricular activities, but because of the need for cross-cultural applicability, these potentially more sensitive and meaningful measures weren't used in the PISA study. Instead, absenteeism was used as the most important aspect of participation.

level), percentage tardiness, and percentage of students belonging to school-sponsored activities.

Student retention and completion

The retention and completion indicator assesses "leaks" from the system at each school level—elementary, middle, and high. Here, the indicator system shows the proportion of students who enter the system at the beginning of a cycle—first grade, sixth grade, and ninth grade—and who are still in school at the end of the cycle. Although the "completion rate" commonly reported now is just for high school students, the retention and completion indicator can provide valuable information to school leaders at every grade span. If a particular cohort of students begins sixth grade together and only 70% of that number are present at the end of eighth grade, this may carry a warning to district leaders: For some reason, parents are abandoning the school and departing students are not being replaced, 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 a district to track each student through the system, letting them know whether an individual who started at a particular middle school remained there through 3 years. However, only 31 of 50 states (and the District of Columbia) 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 districtwide monitoring (Celio & Harvey, 2005). Thus, the ideal (i.e., 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. Considerable controversy has surrounded these reports, however; the way the data are collected and computed makes all the difference in their usefulness as indicators of school health.

As a measure of a school's ability to retain its students, the completion rate recommended by Greene and Winters (2002) of the Manhattan Institute has some real advantages over the traditional dropout rate. Greene demonstrated that the completion rate is straightforward and stark. The Manhattan Institute approach is used in the indicator system. The retention-completion rate measures school completion in the aggregate, comparing 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 such as 9th grade. 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 about individual students.

It is unlikely that the retention and 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 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 schools. Changes across the district may require a more general response.

Status: In most schools, we might expect the number of students in a grade to be approximately the same each year. Barring an obvious situation 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.

Change: To see how a school's ability to retain its students changes over time, it is necessary to look at cohorts of students. The National Governors Association has recommended a cohort analysis approach to graduation rates. Most states have indicated their commitment to implementing this method. If the recommendation is implemented, many districts will have data for use in a retention and completion indicator system.

Teacher attraction and retention

Teacher effectiveness is not an element listed in the indicator system because currently no way of assessing it independent of student performance exists. There is a circular quality to many analyses of teaching 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, or having a major in the subjects taught—has been shown to be reliably related to significantly greater student achievement.

Systems that rate teacher effectiveness have traditionally looked retrospectively at teachers to see how their students did over time. Such approaches are of little help when examining an entire district to determine which schools require assistance; a concentration of less effective teachers may account for the poor performance of students, but so also may changes in demographic makeup of the school, innovative curricular approaches, or new leadership. At the moment, there is no direct way to measure the effectiveness of all teachers, or specific teachers, within a school or district. Until a measure of teacher effectiveness can be developed and added to the indicator system, teacher attraction and retention are suggested as proxies.

Although not an ideal indicator of the important role teachers play in school effectiveness, the measurement of teacher attraction and retention is useful. Just as it would be useful to know what parents are looking for in a school before they enroll their children, it would be helpful to know how potential employees view a 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.

When 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, teachers' perceptions of a school can play an important role in determining whether that school will attract the desired number and quality of teachers.

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 (2004) reported that calling around to a few schools in Los Angeles revealed that some schools received, at best, 1 to 3 applications per position while others had up to 130 applicants. Similar studies in the Seattle Public School District found much the same pattern: Some schools had a handful of applicants per opening while others received dozens (Roza, 2004). It should not be a surprise that schools with few applicants wind up doing the best they can when hiring while schools with many applicants can be choosier, selecting teachers with the qualifications and qualities they seek.

Attraction and retention are not the same thing. A school might 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 could develop because potential applicants have not yet heard that the school is not a good place to work. There could also be more positive explanations, for example, 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, this indicator, like others in the indicator system, cannot provide a diagnosis or prescription, but it can act as an early warning system, alerting leaders to look closely at what is happening in a school. Also, as with other indicators, a particular rating on the teacher attraction and retention indicator can call for additional attention from the school superintendent and school board.

Funding equity/efficiency

Although the achievement gap between groups of students has received a lot of public and professional attention recently, another type of gap has elicited little comment: a funding gap between school districts and even schools within a district. A gap between districts cannot be solved by districts, but a disparity in funding among schools in the same district can be.

The Center on Reinventing Public Education (CRPE) has conducted detailed studies of budgeting practices in more than a half dozen major school districts over the past 10 years (Roza & Miles, 2002; Roza & Hill, 2004). This research revealed substantial differences in the actual (as opposed to the budgeted) funding levels of schools within each of these districts. Surprisingly, these differences are often largely invisible not just to the public eye but to the eyes of district leaders.

The differences fly beneath the radar of both district leaders and the general public because schools are "resourced" rather than funded. That is to say, they are provided with a certain number of teachers depending on enrollment and not on funds to pay for teachers, much less the categorical funding that is intended for particular groups of students such as special education, English language learners, and low socio-economic status students. The budgeting process in many districts makes it difficult to determine exactly how much funding is going to each school.

When CRPE researchers examined school funding, 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 were intended.[7] The result in all the districts studied was that the most needy schools tended to receive lower per-student funding than schools with fewer needs. In effect, schools with the most challenges were subsidizing schools with the fewest.

Status and change: The CRPE research developed two measures of funding equity that are used in the indicator system. The first uses actual versus budgeted teacher salaries and the second uses a weighted index of resource allocation to compare expected funding with actual funding for 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 (the district's average teacher salary multiplied by the number of teachers assigned to the school) with how it spends the money (the

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

real salaries of teachers in the school). Some of the teachers in the school may make the minimum salary while others may be paid at the top of the scale. This measure throws light on a finance fiction—namely that budgeting (or "resourcing") schools on the basis of average teacher salaries represents what is spent on teacher salaries in individual schools.

The CRPE researchers found that every district queried about the effects of average versus actual teacher salaries was convinced that the average teacher salary within the schools would closely match the average district salary (Roza & Hill, 2004). That was not the case in any single district studied. In fact, the disparity among schools within a district in real teacher salaries amounted to a gain of as much as $1 million in some schools. This could only be made up with corresponding losses in other district schools.

This disparity might not be of great significance if all teachers possessed equal experience and ability. In practice, what happens is that very needy schools tend to be staffed largely with new and inexperienced teachers, at the bottom of the salary scale. Once those teachers 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 (2004) concluded that "there is good reason to believe that schools with higher average salaries have more capable teachers."

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 (Miles, Ware, & Roza, 2003; Miller, Roza, & Swartz, 2004; Miles & Roza, 2006). If support were attached to students rather than buildings, a school serving a large number of low-income, educationally vulnerable students should receive more funding than one serving children of upper-income, professional families. The disadvantaged students not only need more assistance, but districts can draw on specific sources of state and federal funds (e.g., Title I) to help them.

As with other indicators, the funding equity indicator cannot tell how a particular situation came about or how to address it, but it can provide educational leaders with a tool that lets them see a meaningful summary of complex data. What they do with these insights is up to them.

CONCLUSIONS AND RECOMMENDATIONS

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

The indicator system consists of both status indicators and trend indicators in seven areas. Each indicator tells part of the story, but even taken together they cannot possibly tell the whole story. However, the 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 wakeup call (to shock, enlighten, and jump-start) or a guide to the goal or standard to be attained, or both.

Most school districts are already collecting the data underlying these indicators. Much of that information is also available to members of the public, who are likely to find it even more difficult to comprehend than teachers and principals. The nation's educational data problem today is not that not enough data exist. Quite the contrary. The problem is that educators and parents are awash in data they find hard to understand. The indicator system described 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.

REFERENCES

Boser, U. (2011). *Return on educational investment: A district-by-district evaluation of U.S. educational productivity*. Center for American Progress.

Bryk, A. S., & Thum, Y. M. (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal*, *26*(3), 353–383.

Celio, M. B. (1989). *Falling through the cracks: A study of dropouts in the Seattle Public Schools.* A Report for the Seattle–King County Economic Development Council.

Celio, M. B., & Harvey, J. (2005). *Buried treasure: Developing a management guide from mountains of educational data*. Seattle, WA: Center on Reinventing Public Education. http://www.crpe.org/publications/buried-treasure-developing-management-guide-mountains-school-data

Council of Chief State School Officers. (2006) *Implementing the NGA graduation rate compact: State-level issues.* http://www.dataqualitycampaign.org/files/Publications-Implementing_the_NGA_Graduation_Rate_Compact_State-level_Issues.pdf

DeKalb, J. (1999). Student truancy. *ERIC Digest*, #125. (ED429334.)

Few, S. (2006). *Information dashboard design: The effective visual communication of data*. Sebastopol, CA: O'Reilly Media.

Fine, M. (1986). Why urban adolescents drop into and out of public school. *Teachers College Record*, *87,* 393–409.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117–142.

Finn, J. D., & Voelkl, K. E. (1993). School characteristics related to student engagement. *Journal of Negro Education*, *62*(3), 249–268.

Garry, E. (1996). *Truancy: First step to a lifetime of problems*. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention. (ED 408 666.)

GreatSchools. http://www.greatschools.org/students/local-facts-resources/31-the-api-10-things-parents-should-know.gs

Greene, J. P., & Winters, M. A. (2002, November). *Public school graduation rates in the United States* (Civic Report No.31). New York: The Manhattan Institute for Policy Research.

Hale, L. (1998). *High school dropouts cost everyone something!* Bethesda, MD: National Association of School Psychologists.

Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences*. New York: Springer-Verlag.

Holland, P. W. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics, 27(*1), 3–17.

Innes, J. E. (1990). *Knowledge and public policy: The search for meaningful indicators*. (2nd ed.). New Brunswick, NJ: Transaction Publishers.

Innes, J. E., & Booher, D. E. (2000). Indicators for sustainable communities: A strategy building on complexity theory and distributed intelligence. *Planning Theory and Practice*, *1*(2), 173–186.

Johnson, M. K., Crosnoe, R., & Elder, G. H. (2001). Students' attachment and academic engagement: The role of race and ethnicity. *Sociology of Education 74*, 318–340.

Kane, T. J., & Staiger, D. O. (2001). *Improving school accountability measures*. (Working Paper #8156.) Cambridge, MA: National Bureau of Economic Research.

Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us*. Aurora, CO: Mid-continent Research for Education and Learning.

Miles, K., & Roza, M. (2006). Understanding Student-Weighted Allocation as a Means to Greater School Resource Equity. *Peabody Journal Of Education, 81*(3), 39-62.

Miles, K. H., Ware, K., & Roza, M. (2003). Leveling the playing field: Creating funding equity through student-based budgeting. *Phi Delta Kappan*, *85*(2), 114–119. http://www.kappanmagazine.org/content/85/2/114.abstract

Miller, L., Roza, M., & Swartz, C. (2004). *A cost allocation model for shared district resources: A means for comparing spending across schools.* Working Paper. Seattle, WA: Center for Reinventing Public Education.

National Research Council. (2001). *Understanding dropouts: Statistics, strategies, and high-stakes testing.* A. Beatty, U. Neisser, W. T. Trent, & J. P. Heubert, J. P., Eds. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Norris, T., Atkisson, A., et al. (1997). *Community indicators handbook: Measuring progress toward healthy and sustainable communities*. Oakland, CA: Redefining Progress.

Olson, L. (2002). Testing experts develop new method of presenting achievement gap data. *Education Week, 21*(26), 11.

Rothstein, R. (2002, May 22). Lessons: Testing reaches a fork in the road. *New York Times*. http://www.nytimes.com/2002/05/22/nyregion/lessons-testing-reaches-a-fork-in-the-road.html

Roza, M. (2004). Rethinking data capacity. In P. T. Hill & H. James (Eds.), *Making school reform work: New partnerships for real change* (pp. 105–106). Washington, DC: Brookings Institution Press.

Roza, M., & Hill, P. (2004). How Within-District Spending Inequities Help Some Schools to Fail. *Brookings Papers On Education Policy*, 201-227.

Roza, M., & Miles, K.H. (2002). *A new look at inequities in school funding.* Seattle, WA: Center on Reinventing Public Education.

State of Washington, Office of Superintendent of Public Instruction. School Achievement Index. http://www.k12.wa.us/accountability/

Shaw, L. (2004, September 2). Improvement on WASL carries asterisk. *Seattle Times*.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Wainer, H. (2004). *Graphic discovery: A trout in the milk and other visual adventures*. Princeton, NJ: Princeton University Press.

Wainer, H. (1994). On the academic performance of New Jersey's public school children: Fourth and eighth grade mathematics in 1992. *Education Policy Analysis Archives*, *2*(10).

Wainer, H., & Brown, L. (2004). Two statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. *The American Statistician, 58,* 117–123. http://www.statlit.org/pdf/2004wainer_threeparadoxes.pdf

Willms, J. D. (2003). *Student engagement at school: A sense of belonging and participation*. Results from PISA 2000. Organisation for Economic Co-operation and Development.