



## ARCHIVED - Archiving Content

### Archived Content

Information identified as archived is provided for reference, research or recordkeeping purposes. It is not subject to the Government of Canada Web Standards and has not been altered or updated since it was archived. Please contact us to request a format other than those available.

## ARCHIVÉE - Contenu archivé

### Contenu archivé

L'information dont il est indiqué qu'elle est archivée est fournie à des fins de référence, de recherche ou de tenue de documents. Elle n'est pas assujettie aux normes Web du gouvernement du Canada et elle n'a pas été modifiée ou mise à jour depuis son archivage. Pour obtenir cette information dans un autre format, veuillez communiquer avec nous.

This document is archival in nature and is intended for those who wish to consult archival documents made available from the collection of Public Safety Canada.

Some of these documents are available in only one official language. Translation, to be provided by Public Safety Canada, is available upon request.

Le présent document a une valeur archivistique et fait partie des documents d'archives rendus disponibles par Sécurité publique Canada à ceux qui souhaitent consulter ces documents issus de sa collection.

Certains de ces documents ne sont disponibles que dans une langue officielle. Sécurité publique Canada fournira une traduction sur demande.

# IDEAS IN AMERICAN POLICING



Number 6  
August 2004

## *Police Departments as Learning Laboratories*

*By Edward R. Maguire*

Like all organizations, police agencies collect, store, process, analyze, interpret, and react to information (Choo 1995). But they often fail to collect or assemble information useful for assessing their performance. Even when such information is available, they rarely analyze it in meaningful ways, learn from it, and implement changes based on what they have learned. In the absence of systematic information sources about their performance, police agencies, like many other public agencies, are often unable to state with any degree of precision how their performance has changed over time or how it compares with that of their peer agencies, particularly those situated in similar community

contexts. Lacking both the right kinds of information and the skills necessary to transform that information into meaningful and usable measures, police agencies tend to suffer from what management guru Peter Senge (1990) calls an “organizational learning disability.” While the extent of the disability certainly varies across agencies, it is to some extent systemic, affecting the entire policing industry. In this paper, I will suggest a systematic framework within which police agencies can overcome these difficulties and make greater use of information and measurement to enhance their capacity for systematic, organizational learning. As they begin to operate like learning laboratories, police organizations

will move closer to becoming “intelligent organizations,” capable of continuous reflection, adaptation, and renewal.

*Ideas in American Policing* presents commentary and insight from leading criminologists on issues of interest to scholars, practitioners, and policymakers. The papers published in this series are from the Police Foundation lecture series of the same name. Points of view in this document are those of the author and do not necessarily represent the official position of the Police Foundation.

© 2004 Police Foundation and Edward R. Maguire. All rights reserved.

**Edward R. Maguire** is an Associate Professor in the Administration of Justice program at George Mason University.

## Information

A substantial proportion of the information that an organization needs for scanning and intelligence analysis already exists within the organization.

Unfortunately, the information is scattered in bits and pieces, and the people who have the information are often unaware of its value to the organization and the need to share it with others (Choo 1995, 195).

Picture the administration of a police department, or any other kind of organization, as a giant information processing system, much like a human brain.

Enormous volumes of information flow into the brain from numerous sources. Some of the sources are internal, such as employee rumors or agency records, and some are external, such as citizen complaints or lawsuits. Some are local, such as media coverage, and some are distant, such as reports from professional accreditation teams. Some are formal and structured, such as official agency statistics,

and some are informal and unstructured, such as input received at community meetings. While all of these sources provide the organization with feedback about its performance, the information does not come packaged in a neat format that provides the organization with a clear direction. The sheer volume of information alone exceeds the cognitive capacity of any one individual to process it carefully, and even probably exceeds the organization's collective capacity for systematic, careful processing. Absent a state-of-the-art information management system, much of it is unlikely to be archived or retained in a format that enables it to be searched or retrieved. Much of it is vague, unclear, conflicting, biased, scattered, or otherwise difficult to interpret. The electronic age has enabled police organizations to store and retrieve more information than ever but very little of it, at least in its raw, unprocessed form, is useful for

gauging the organization's performance or charting its strategic future. It is a classic case of information overload.

To develop the foundation for systematic organizational learning, police agencies need to develop two new capacities. First, they must enhance their ability to process, interpret, make sense of, and learn from information already available within the agency. One important way to do this is to begin transforming raw data into meaningful and easy-to-understand measures on a variety of performance dimensions. In other words, a thoughtful and technically sound measurement strategy will enable the organization to extract more meaning from available data. To some extent, this is the idea behind data mining, an innovation that is growing popular in the private sector. Second, police agencies must collect new information that enables them to discover vital patterns and trends both inside the organization and in its external environment. Later I will provide examples of measures computed from a variety of sources that reveal useful information about how police organizations change over time and how they compare to their peers.

Nobel laureate Herbert Simon pointed out more than four decades ago the idea that managers and administrators have cognitive limitations in how much information they can

*The electronic age has enabled police organizations to store and retrieve more information than ever . . .*

process (Simon 1957). Faced with these limitations, they are often described as having to make decisions with “bounded rationality.” If this is true, as organization theorists believe, then reducing the complexity of the information available to administrators should improve decision making. Measurement is one important way to summarize large volumes of information into simple parcels that can be cognitively processed more easily than raw data.

## Measurement

I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind, it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be (Sir William Thomson, Lord Kelvin 1889, 73-74).

Imagine a world without measurement. Rather than using a blood pressure cuff, your physician places his fingers on your arm and concludes that your blood pressure is normal. Your child comes home from school without a quarterly report card but instead with a note from the teacher that says she did “pretty well.” With no speedometer or gas gauge in your automobile, you don’t know when you are speeding or running out of fuel. When deciding whether to

purchase a stock for your portfolio, your broker relies on his feelings and gut impressions rather than an in-depth analysis of the stock’s past performance. In many walks of life, measurement is not only accepted, it is expected. It provides us with feedback, helping us to make decisions, helping us to learn.

In each of the examples above, it would strike us as odd to rely on intuition, anecdotes, feelings, or gut-level impressions as the principal basis for forming a judgment or making a decision. We would find it odd because in each example there is an existing and accepted unit of measurement available to us: blood pressure is measured using millimeters of mercury (mmHG), a common measure of the pressure of gases; the A-F grading system is commonly used to measure student performance in individual courses and translates into an overall grade-point average that ranges from 0 to 4.0; speed is measured in miles per hour; fuel gauges measure the fraction of fuel remaining in the tank; and stock performance is measured using a variety of indices, including the price-to-earnings ratio.

It seems that we become uncomfortable with the idea of measuring a phenomenon when there is no easily apparent or ready-made unit of measurement available for it. But all units of measurement are man-made. For example, although the use of a

thermometer is now commonplace, temperature does not come equipped with a natural unit of measurement. As one scientist has observed, “temperature is a difficult concept to make quantitative” (Lindley 2004, 98). The history of the thermometer’s development stretched over several centuries. During that time, numerous scales were proposed for measuring temperature before the three primary scales used today—Celsius, Fahrenheit, and Kelvin—became widely adopted.

In addition, the more abstract the phenomenon, the more uncomfortable we are with the prospect of measuring it. Yet psychologists have been measuring abstract phenomena for more than a century. Psychometrics, the science of measuring psychological characteristics, dates back to at least 1879 (Galton 1879; Ludlow 1998). Perhaps the best known examples of psychometric measuring instruments are intelligence tests and the various tests of scholastic aptitude that students take to gain admission to colleges and universities, but there are many others, including personality tests and vocational tests. The *Mental Measurements Yearbook*, for example, contains reviews of more than 2,000 commercially available psychometric tests in the English language alone (Plake, Impara, Spies, and Pale 2003). Virtually every characteristic measured in psychometrics is abstract and

*... [S]ometimes we choose not to measure important phenomena, particularly when there is no natural or existing unit of measurement or when [it] is abstract or intangible.*

intangible, from depression and intelligence to job satisfaction and organizational commitment.

Although psychometrics has made the greatest number of technical contributions to social science measurement, it is not the only social science field to focus on measurement. A team of researchers working on the Project on Human Development in Chicago has developed a set of statistical methods called “ecometrics” (Raudenbush and Sampson 1999). Just as psychometrics measures individual-level phenomena, ecometrics focuses on measuring the characteristics of aggregates like neighborhoods or communities. Similar methods have been developed for studying organizations, where measures are often computed for individuals and then aggregated to form organization-level scores (Camp, Saylor, and Harer 1997; Camp, Saylor, and Wright 1999).

Sociometry, a longstanding field that focuses on measuring interpersonal relationships in groups, is yet another branch of social science measurement (Moreno 1960). Collectively, the various measurement subfields in the social sciences have been developing for more than a century now, and important new scientific developments are still emerging today.

Measurement is valuable. It helps us make judgments, make decisions, and learn. Yet sometimes we choose not to measure important phenomena, particularly when there is no natural or existing unit of measurement or when the phenomenon is abstract or intangible. Both conditions—the lack of a uniform unit of measurement, and a concept that is abstract and intangible—apply to the measurement of police performance.

Yet the measurement sciences

teach us that these problems can be overcome. They are unlikely to be overcome by those without expertise in the technical aspects of measurement. There is a lengthy tradition of research demonstrating that measuring—especially in aggregates like neighborhoods, communities, or organizations—without paying adequate attention to technical details can produce seriously misleading results (Raudenbush and Sampson 1999; Camp, Saylor, and Wright 1999). Consider the basic ecometric equation used to measure the characteristics of Chicago neighborhoods. Filled with Greek letters and multiple subscripts, it is sufficient to scare away all but the most competent statisticians (Raudenbush and Sampson 1999). Any serious attempt to measure abstract concepts, including virtually all efforts to measure the performance of police organizations, should pay careful attention to the technical aspects of measurement.

Although measurement methods require technical sophistication, the resulting measures are, or should be, simple and easily interpretable. Often this means transforming data into standardized scales with known properties. Consider standardized aptitude tests like the Scholastic Aptitude Test (SAT). Students taking the SAT do not receive a lengthy report detailing their performance item by item. Instead, they receive a simple, easy-to-read report that

summarizes their verbal and math scores on a known scale ranging from 200-800, together with percentiles informing them how their scores compared with others who took the test. Underlying the scoring of the SAT are a battery of complex statistical methods but these are all transparent to those who take the test. Similarly, researchers at the Federal Bureau of Prisons use advanced statistical models to compute performance measures for prison facilities. They recognize that the method they use “to construct such measures is technically complex but, from the point of view of data consumers, it is not necessary to understand the technical details to effectively utilize the information” (Camp, Saylor, and Wright 1999). Measurement is difficult but interpreting measures should not be. Measurement plays a vital role in police departments as learning laboratories.

Even if police agencies were to begin collecting new information useful for gauging their performance, and to develop the internal capacity for expert measurement so that new and existing information could be transformed into meaningful performance measures, many would still be lacking a crucial ingredient for organizational learning: a comparison group that enables them to interpret their scores. The reason the SAT and other standardized tests are such valuable measures is that they

enable comparison. Most measures in policing do not permit such comparisons. For instance, suppose a police agency implements a new citizen survey designed to measure the friendliness and responsiveness of officers. If the agency’s research team conducts the survey once and computes friendliness and responsiveness measures from the survey data, how will they know if the agency has performed well? Without a source of comparison, they won’t know. Their plight reminds me of an unsolvable math problem. “I have \$2.00 to purchase some widgets. How many can I buy?” They are missing crucial information. Absent the capacity to compare, the measures we have computed from our raw information are much less meaningful than they could be. That is where comparative performance measurement comes in.

### Comparative Performance Measurement

Organizational leaders often have difficulty assessing the performance of their own organizations. One reason is that information may not be available to help them compare how well their organizations are doing relative to others providing similar services (Gormley and Weimer 1999, 30).

Performance measurement is everywhere, from schools to corporate boardrooms, from hospitals to local governments. It

is now common to measure the performance of employees, of sub-units and project teams, and of organizations as a whole. Here the focus is on *comparative* performance measures that can be used to compare police organizations over place or time. I sketch a framework through which systematic measurement of the performance of police organizations can be used as a renewable resource for continuous improvement in the policing industry (Maguire 2004). If implemented correctly, this performance measurement framework can serve as a springboard through which police organizations can become like *learning laboratories*, reaching into their internal and external environments, taking measurements, making diagnoses, implementing change, and regularly repeating the process.

I have already written a detailed introduction to

*Measurement plays a vital role in police departments as learning laboratories.*

*Police agencies provide a variety of public services . . . and this variety makes it difficult to measure their performance.*

comparative performance measurement in policing for the Commission on Accreditation for Law Enforcement Agencies (CALEA) (Maguire 2003, 2004). Here I will provide only the basic framework and begin by emphasizing the core idea that police performance is multidimensional. Police agencies provide a variety of public services to their communities, and this variety makes it difficult to measure their performance. One agency might do a terrific job at dealing with mentally ill suspects, but it may not be adequately prepared to respond to natural disasters or other large-scale critical incidents. Another may routinely invest in state-of-the-art information technologies, but its investigations may produce lower clearance or conviction rates compared with similar agencies. The idea that police agencies might be very successful in some ways but less successful in others is not unique to the police. Even in corporations, where the

principal measure of performance is the famed “bottom line,” other dimensions of performance matter. Recent corporate scandals have demonstrated the perils of focusing on profit while ignoring other dimensions of performance, such as maintaining fair and accurate accounting and employment practices. The idea that performance is multidimensional is central to the development of good comparative performance measures.

If performance is multidimensional, what are the dimensions that ought to be measured in policing? Answering

this conceptual question is the crucial first step in performance measurement. While scholars can facilitate or contribute to discussions and debates about the dimensions of police performance, answering this question ought to be the work of the police in consultation with their communities. My purpose here is not to suggest the dimensions on which police performance should be measured, but I will provide two examples from other scholars. The first, which appears in a recent book by Mark Moore, contains seven dimensions of police performance that he believes ought to be measured (Moore 2002).

The second was outlined in a previous *Ideas in American Policing* monograph by Stephen Mastrofski (1999) where he described six dimensions of police service quality that ought to be measured. Notice that all six of Mastrofski’s dimensions could be classified under Moore’s seventh dimension. This illustrates how quickly the process of determining dimensions of police

**Figure 1. Seven Dimensions from Moore’s Recognizing Value in Policing**

- Reduce criminal victimization
- Call offenders to account
- Reduce fear and enhance personal security
- Guarantee safety in public spaces
- Use financial resources fairly, efficiently, and effectively
- Use force and authority fairly, efficiently, and effectively
- Satisfy customer demands/achieve legitimacy with those policed

*Figure 2. Six Dimensions from Mastrofski's Policing for People*

- Attentiveness
- Reliability
- Responsiveness
- Competence
- Manners
- Fairness

performance can become overly complex. I will return to the topic of information complexity in performance measurement shortly.

These are just two possible schemes for thinking about the dimensions of police performance worth measuring. While both of these make sense, my experience suggests that there are at least as many potential performance measurement frameworks as there are people to imagine them.

Once the dimensions of policing have been selected, we must determine how to measure them. Traditional performance measures in policing are often derived from administrative data maintained by the police department. Although these data can be very useful, they should not be the only source used in a comprehensive performance measurement system. Several alternatives exist, including:

- *General community surveys.* These surveys are based on random samples of the community, including people who may or may not have had contact with the police. Questions tend to focus on a

handful of domains, including perceived crime and disorder, fear, victimization history, and overall satisfaction with police. General community surveys are commonplace and valuable but they also have some limitations.

- *Citizen contact surveys.* These surveys are administered to those who have had recent contact with the police. Contact surveys can be a crucial source of management intelligence about police behavior and the perceptions of clients, including both voluntary clients (e.g., those who call the police) and involuntary clients (e.g., arrestees). For instance, when Toronto police surveyed rape victims, they found that victims rated the performance of investigators as quite high but gave substantially lower ratings to the patrol officers (Rape Victims Rate Police Performance 1998). Surveys of different contact types, such as arrestees, crime victims, or drivers cited by the police, could be very useful—if administered in multiple

cities—for learning whether a police agency is perceived as more or less fair or effective than others. Contact surveys could also be administered over time within a single agency.

- *Employee surveys.* Employee surveys can also serve as versatile sources of management intelligence. They are used commonly to explore issues such as morale, job satisfaction, organizational social climate, and environment of integrity (e.g., Camp, Saylor, and Miles 1997). In one ambitious study, researchers relied on survey responses from more than 3,000 individual police officers

*Traditional performance measures in policing are often derived from administrative data maintained by the police department.*



to form an aggregate measure of the “environment of integrity” in thirty police agencies. The results showed that police agencies vary widely in their overall environments of integrity (Klockars, Ivkovich, Harver, and Haberfeld 2000).

- *Direct observation.* Direct observation by trained observers is another potentially valuable method for collecting performance measures. While systematic observation of police on patrol by researchers is common within police agencies, it has not often been used to compare agencies. However, there are many examples of using direct observation for comparison purposes. For example, the Early Childhood Environment Rating Scale uses trained observers to rate the quality of child care facilities based on direct observation of the space and furnishings, the interaction between children and teachers, and several other dimensions of performance (Harms, Clifford, and Cryer 1998). In New

York, systematic observation has been used to measure the “smoothness” of 670 miles of roads in fifty-nine districts (Fund for the City of New York 2001). In Chicago, it has been used to record the volume of physical and social disorder in neighborhoods (Sampson and Raudenbush 1999). While direct observation can produce meaningful performance measures, it is personnel intensive and therefore very expensive.

- *Independent testing or simulation studies.* Rather than observing performance in completely natural settings, independent tests create artificial opportunities to measure performance. For instance, the Insurance Institute for Highway Safety uses crash tests to rank vehicle safety. Many firms hire people to pose as customers, known as secret shoppers or mystery shoppers, who visit their facilities to perform checks on quality of service, cashier

accuracy, ethical standards, and many other issues. Internal affairs units in large police agencies have conducted various kinds of integrity tests for many years. ABC News (2001) conducted independent integrity tests of police in New York and Los Angeles by turning over forty wallets or purses to police officers chosen at random. All of the wallets and purses were turned in by the officers with contents intact. While controversial, testing and simulation offer some interesting promise for collecting performance data that are truly independent of the police.

## Some Important Decisions

Once the dimensions of performance have been selected, it is important to resolve some important issues about how to use the data. I will not outline these issues in detail here, but will mention three of them briefly: minimizing information complexity, accounting for differences in the relative importance of the dimensions, and ensuring fair comparisons.

### *Minimizing information complexity*

Imagine for a moment that a group of agencies has settled on Moore’s seven dimensions of performance and they select five measures of each dimension. They will have a total of

*. . . [T]esting and simulation offer . . . interesting promise for collecting performance data that are truly independent of the police.*

*... [T]he way data are presented can affect our ability to process information.*

thirty-five separate measures. Although they will have reduced somewhat the information complexity that was discussed earlier, they are still faced with a volume of information that is pushing the limits of their cognitive processing abilities. The idea that human beings have limits in the amount of information they can process is a fundamental principle of cognitive psychology (Miller 1956). At the same time, recent developments in the study of information processing demonstrate that the way data are presented can affect our ability to process information. Research shows that presenting data efficiently using diagrams, graphics, and other visualization techniques can enhance cognitive processing capacity (Card, MacKinley, and Schneiderman 1999; Larkin and Simon 1987; Tufte 1997; Wainer 1997). In sum, the evidence suggests that the total number of performance

measures should be small enough not to overwhelm our cognitive processing limits. At the same time, we can increase the number of performance measures that can be processed if we pay careful attention to the way we present those measures, using simple, intuitive visual presentations rather than complex tables containing dozens or hundreds of numbers. The idea of enhancing cognition through graphical summaries of data has already been proposed in a number of substantive areas, including medical charts (Powsner and Tufte 1994), psychiatric data (Powsner and Tufte 1997), and student test scores (Wainer 2002).

Another option for reducing information complexity is to form composite scores that combine the items within each dimension into a single score. When a student takes the SAT, the Graduate Record Examination (GRE), or other similar standardized tests, the overall scores represent composites of the individual test questions. These composite scores are standardized to fall within a certain range, such as 200-800 for the SAT. Nobody is very interested in their performance on individual test questions, only the overall score within each dimension, such as math and verbal. Measurement experts can use similar methods to create standardized, composite scores of police performance within each dimension. It will take time for

the science to mature, but the development of standardized, composite scores of police performance is a realistic possibility.

### *Accounting for differences in the relative importance of the dimensions*

Some dimensions of performance are more important than others. A common complaint about some performance measurement systems is that they treat each measure equally. This is acceptable as long as the different domains of performance are equivalent but if some are much more important than others, it is misleading. Sometimes it is useful to assign greater weight to certain measures when computing composite performance scores. There are a variety of methods for computing such weights but they require technical expertise. The more challenging question is how to assign the weights in a manner that is not totally arbitrary. How do we quantify differences in importance between multiple goals? If the differences are minor, they may be worth overlooking. If there are major differences in importance, such as the friendliness of the hospital staff versus its mortality rate, then these differences need to be accounted for within the performance measurement framework (Gormley and Weimer 1999). Is a low crime rate more important than treating citizens

fairly? If so, is it twice as important? Three times as important? One method for determining the relative weight of different dimensions of performance is to conduct focus groups or surveys of citizens. Thinking about the relative importance of different dimensions of police performance can also be a useful exercise that will help a police agency clarify its goals.

### *Ensuring fair comparisons*

For comparative performance measurement to work, substantial investment needs to be made in keeping comparisons fair. Police agencies are located in very different environments and these differences need to be taken into account when comparing agencies. This concern with making fair comparisons between organizations is not unique to the policing industry. Just as police departments located in the poorest, most disorganized communities might be expected to have the highest crime rates, hospitals admitting the most at-risk patients might be expected to have the highest death rates, and prisons admitting the worst offenders might be expected to have the highest recidivism rates. However, we can control for these variations in the inputs of organizations when measuring their performance.

There are two primary methods for controlling for variations in inputs: stratification (forming peer groups) and

calculating risk-adjusted performance measures. Stratification is the easiest. It works by establishing peer groups of similar agencies. All of the agencies within a peer group will be compared with the other peer agencies. One way to reduce embarrassment or fear would be to set a threshold for a minimum peer group size, such as ten agencies. All members in the peer group would know the names of the other participating agencies. However, while each agency would receive its rankings on various measures relative to the other peer agencies as a group, it would not have access to the other agencies' individual rankings. Thus, for example, only the agency in last place would know that it came in last place.

While stratification is the easiest method, sometimes it can also be tricky. Some agencies and some jurisdictions are so unique that it might be difficult to classify them within a peer group. For instance, I recently spoke about peer groups with the chief of the Nishnawbe-Aski Police Service, a Canadian tribal police agency. With only ninety-three officers and a jurisdiction the size of France, it has some detachments that can be reached only by air. Finding a suitable peer group for this agency is likely to be difficult.<sup>1</sup>

---

<sup>1</sup> I am grateful to Chief Wesley Luloff of the Nishnawbe-Aski Police Service in Ontario, Canada, for this observation.

Furthermore, sometimes the unique factors that might make it challenging to find a suitable peer group may be subtle or difficult to anticipate. For instance, some “edge cities” (Garreau 1991) may not have a large population but their proximity to large urban areas means they face issues that render them unique compared with other similarly sized communities.<sup>2</sup>

The more complicated method is calculating risk-adjusted measures that use statistical controls to adjust the measures for variations in inputs. This process will require technical expertise and a substantial investment in testing and calibration to ensure that the risk-adjustment procedures are scientifically defensible. Furthermore, since risk adjusted crime rates are based on an implicit assumption that demographic and structural characteristics—such as poverty, race, and region—influence crime, the risk-adjustment procedures might inspire controversy.<sup>3</sup> The scientific foundation for this process in policing is not yet well established. In a previous *Ideas in American Policing* monograph,

---

<sup>2</sup> I am grateful to Chief John Douglas of the Overland Park Police Department in Kansas for this observation.

<sup>3</sup> I am grateful to the members of the Performance Measurement Subcommittee of the Commission on Accreditation for Law Enforcement Agencies for this observation.

Lawrence Sherman (1998) proposed treating risk-adjusted measures of crime as performance measures for police agencies. At the time, he presented a hypothetical example of risk adjusted homicide rates. Since then, a team of researchers based at Georgia State University, Carnegie-Mellon University, the University of Missouri-St. Louis, and the National Consortium on Violence Research has begun to develop such measures.<sup>4</sup> According to their preliminary estimates, just five demographic and socioeconomic factors are responsible for more than two-thirds of the variation in homicide rates across cities. This initial finding suggests that holding police accountable for crime rates without risk adjustment means holding them responsible for social and economic factors largely beyond their control.

### Why This Approach Might Work

There are many ways to change organizations. The vision I have just outlined is only one potential method for improving police organizations. All organizations are capable of self-learning, adaptation, adjustment, experimentation, and innovation. To do so, organizations need information

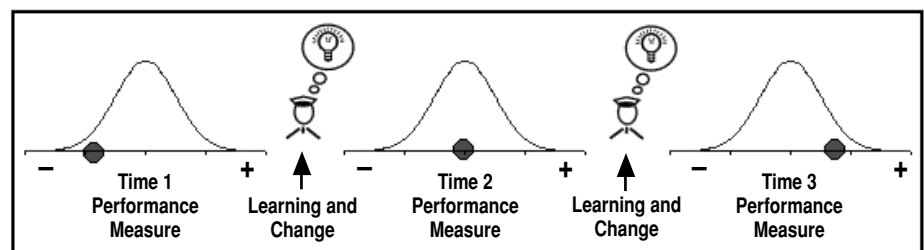
and feedback. Comparative performance measures will provide police organizations with crucial information: how they are doing relative to other police agencies on a variety of performance dimensions and how they are improving relative to their own previous levels of performance. A variety of evidence from research in the information sciences suggests that organizations which engage in greater levels of “environmental scanning” have greater levels of performance (Choo 1995). So far, this research has taken place in private-sector firms and nonprofit agencies. The performance measurement scheme outlined here represents just one form of environmental scanning. There is also evidence that ranking organizations can stimulate those who are performing poorly to implement changes designed to improve their performance (Gormley and Weimer 1999).

The theory here is quite simple: comparison motivates change. Comparative performance measures are likely to induce change among chronically low performers who,

absent the measures I have proposed, probably do not even know that their performance is weak. Picture a process whereby the agencies learning that their performance is low relative to their peers initiate changes designed to improve their performance. If their performance improves, others are now the lowest performers. As this process of learning and change continues to take place, the net result is continuous improvement in the policing industry. Figure 3 depicts this continuous improvement process. Each performance measurement cycle generates learning and change within the organization, resulting in improved performance.

Many police agencies do not know how their performance rates relative to their peers. There do exist some extreme indicators of poor performance that might be sufficient to induce change in police organizations. Lawsuits, negative publicity, court orders, and consent decrees are all engines that drive reform, but these catalysts only activate when performance has degenerated to an extreme. What police administrators need are simple,

Figure 3. An Organizational Learning Cycle



<sup>4</sup> [Http://www.cjgsu.net/initiatives/HomRates-PR-2003-08-03.htm](http://www.cjgsu.net/initiatives/HomRates-PR-2003-08-03.htm)

*As the science of performance measurement develops, it is likely that some agencies will respond by criticizing the validity of the measures.*

clear, easy-to-understand indicators that measure performance regularly across the full range of the performance spectrum, from negative, to neutral, to positive.

### **Why This Approach Might Not Work**

There are many reasons to believe that systematic comparative performance measurement might never be implemented and might not improve organizational learning even if it is implemented. Exploring the full range of potential pitfalls would require a whole paper in itself. Here I will briefly explore just a handful of them.

- Many agencies may be unwilling to participate, especially those with the worst performance.
- There are plenty of examples of performance measures generating perverse behaviors

within organizations, such as falsifying or manipulating data or allowing unmeasured aspects of performance to suffer due to a myopic focus on those that are measured. For example, security guards at one nuclear power plant surprised everyone by repelling a series of surprise simulated terrorist attacks. A later investigation revealed that they had been tipped off about the timing of the simulations (U.S. Department of Energy 2004).

- As the science of performance measurement develops, it is likely that some agencies will respond by criticizing the validity of the measures. Again, there are plenty of examples of this from other policy domains, especially in measuring the performance of hospitals (Iezzoni 2003). Although there is some evidence to suggest that environmental scanning and comparative performance measures can induce change in other

industries, there is no evidence that they will produce such effects in policing.

- The new science of performance measurement could easily become tainted or dominated by political considerations.
- Finally, the vision outlined here assumes there is a cadre of people with expertise in transforming raw police data into useful information. This assumption is not true right now, but universities could easily create degree or certificate programs to develop people with these skills once the demand for such expertise began to exceed the supply.

### **Implementing the Vision**

This vision cannot be implemented overnight. It will take time and it will require investment and support from several important sources.

First, and most obviously, it will require the energetic participation of police leaders. The training I have provided on performance measurement to hundreds of police administrators and accreditation managers at law enforcement conferences in the United States and Canada over the past two years convinces me there is a growing demand for education and training on these ideas.

Second, the development of the science of performance measurement will require

financial and logistical support from government. The National Institute of Justice in the U.S., the Home Office in the U.K., and other similar government agencies will need to underwrite the investment in developing the scientific methods on which performance measures are based. There needs to be careful attention paid to testing, calibration, and refinement of the measures. Much like the Educational Testing Service's efforts to continue building and improving standardized tests, this will be an ongoing process of incremental development.

Third, the systematic adoption of comparative performance measurement is likely to create a new occupational niche for people with combined expertise in survey research, data analysis, measurement, and information visualization. The vision outlined here would require the development of a new civilian position in those police agencies

large enough to support it. The qualified individual would be someone with expertise in performance measurement who can ensure that the appropriate data are collected, analyzed, summarized, and reported in appropriate ways. Perhaps these individuals will also have expertise in auditing the kinds of data used to create such measures. Just as the development of crime mapping led to the development of crime analyst positions in agencies around the nation, embracing performance measures will mean hiring those with the expertise to implement them.

Fourth, universities would need to create programs that teach performance measuring skills and certify these experts.

Finally, since many smaller agencies would be unable to hire their own performance measurement gurus, government would need to provide technical assistance to some agencies. Perhaps this could be organized along the framework of regional

community policing institutes, offering services to those agencies within each region who are too small or otherwise unable to measure their own performance.

## Conclusion

The ideas outlined here are ambitious. They are not cheap, not easy, and not fast but they do hold promise for improving policing. The first step is to begin conducting pilot studies. The Commission on Accreditation for Law Enforcement Agencies has initiated just such a project with about a dozen agencies. Their goal is to determine if it is possible to collect uniform measures on several dimensions of police performance. The process will not be easy, but in the spirit of safe and just communities, I believe it is worth the effort.

## References

- ABC News. 2001. Testing police honesty: A Primetime investigation with lost wallets. Primetime, Thursday, May 17. ABC News: New York.
- Camp, Scott D., William G. Saylor, and Miles D. Harer. 1997. Aggregating individual-level evaluations of the organizational social climate: A multilevel investigation of the work environment at the Federal Bureau of Prisons. *Justice Quarterly*, 14(4), 739–761.

*... [T]he systematic adoption of comparative performance measurement is likely to create a new occupational niche ...*

- Camp, Scott D., William G. Saylor, and Kevin N. Wright. 1999. Creating performance measures from survey data: A practical discussion. *Corrections Management Quarterly*, 3(1): 71–80.
- Card, Stuart K., Jock D. MacKinley, and Ben Schneiderman. 1999. *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann Publishers, Inc.
- Choo, Chun Wei. 1995. *Information management for the intelligent organization: The art of scanning the environment*. Medford, NJ: American Society for Information Science and Technology.
- Fund for the City of New York. 2001. How smooth are New York City's streets? Center on Municipal Government Performance.
- Gallagher, Catherine, Edward R. Maguire, Stephen D. Mastroski, and Michael D. Reisig. 2001. The public image of the police: Final report to the International Association of Chiefs of Police. Manassas, VA: George Mason University, Administration of Justice Program.
- Galton, Francis. 1879. Psychometric experiments. *Brain: A Journal of Neurology*. 2: 149-162.
- Gardner, Howard. 1983. *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Garreau, Joel. 1991. *Edge city: Life on the new frontier*. New York: Doubleday.
- Gormley, William T., Jr., and David L. Weimer. 1999. *Organizational report cards*. Cambridge, MA: Harvard University Press.
- Harms, T., R.M. Clifford, and D. Cryer. 1998. *Early childhood environment rating scale*. New York: Teacher College Press.
- Iezzoni, Lisa I. 2003. *Risk adjustment for measuring healthcare outcomes*. Third Edition. Chicago: Health Administration Press.
- Klockars, Carl B., Sanja Kutnjak Ivkovich, William E. Harver, and Maria R. Haberfeld. 2000. The measurement of police integrity. Research in Brief. Washington, DC: National Institute of Justice.
- Larkin, Jill, and Herbert A. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, Vol. 11, No. 1, pp. 65-99.
- Lindley, David. 2004. *Degrees Kelvin: A tale of genius, invention, and tragedy*. Washington, DC: National Academies Press.
- Ludlow, Larry H. 1998. Galton: The first psychometrician? *Popular Measurement*, 1(1): 13–14.
- Maguire, Edward R. 2003. Measuring the performance of law enforcement agencies, Part 1. *CALEA Update*, 83. <http://www.calea.org/newweb/newsletter/No83/measurement>.
- Maguire, Edward R. 2004. Measuring the performance of law enforcement agencies, Part 2. *CALEA Update*, 84. <http://www.calea.org/newweb/newsletter/No84/maguirepart2.htm>
- Mastroski, Stephen D. 1999. *Policing for people*. Ideas in American Policing. Washington, DC: Police Foundation, (March).
- Miller, George A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81–97.
- Moore, Mark H. 2002. *Recognizing value in policing: The challenge of measuring police performance*. Washington, DC: Police Executive Research Forum.
- Moreno, Jacob Levy. 1960. *The sociometry reader*. Glencoe, IL: The Free Press.
- Plake, Barbara, James C. Impara, Robert A. Spies, and Barbara S. Pale, eds. 2003. *The fifteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- Powsner, Seth M. and Edward R. Tufte. 1994. Graphical summary of patient status. *The Lancet*, 344: 386–389.
- Powsner, Seth M. and Edward R. Tufte. 1997. Summarizing

- clinical psychiatric data. *Psychiatric Services* 48: 1458–1461.
- Rape victims rate police performance. 1998. *Toronto Star*, July 23: C3.
- Raudenbush, Stephen W. and Robert J. Sampson. 1999. Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology*, 29, 1–41.
- Sampson, Robert J., and Stephen W. Raudenbush. 1999. Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology*, 105, 603-651.
- Senge, Peter M. 1990. *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday/Currency.
- Sherman, Lawrence J. 1998. *Evidence-based policing*. Ideas in American Policing. Washington, DC: Police Foundation, (July).
- Simon, Herbert A. 1957. *Models of man: Social and rationale*. New York: John Wiley and Sons.
- Thomson, Sir William (Lord Kelvin). 1889. *Electrical units of measurement. Popular lectures and addresses*, Volume 1 of 3. London: Macmillan.
- Tufte, Edward R. 1997. *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- U.S. Department of Energy. 2004. Inspection report: Protective force performance test improprieties. Washington, DC: U.S. Department of Energy, Office of Inspector General.
- Wainer, Howard. 2002. Clear thinking made visible: Redesigning score reports for students. *Chance*, 15: 56-58.
- Wainer, Howard. 1997. *Visual revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. New York: Copernicus.



## ABOUT THE POLICE FOUNDATION

The Police Foundation is a national, independent, not-for-profit organization dedicated to supporting innovation and improvement in policing through its research, technical assistance, communication, and professional services programs. Established in 1970, the foundation has conducted seminal research in police behavior, policy, and procedure, and works to transfer to local agencies the best new information about practices for dealing effectively with a range of important police operational and administrative concerns. Motivating all of the foundation's efforts is the goal of efficient, humane policing that operates within the framework of democratic principles and the highest ideals of the nation.

### DIVISION OF RESEARCH, EVALUATION, & PROFESSIONAL SERVICES

Karen L. Amendola  
*Chief Operating Officer*

Robert C. Davis  
*Director of Research*

M. Joe Ryan  
*Director, Crime Mapping &  
Problem Analysis Laboratory*

Edwin E. Hamilton  
*Director of  
Professional Services*

Greg Jones  
*Research Associate*

Raymond Johnston, Sr.  
*Systems Engineer*

Vanessa Correa  
*Program Associate*

Eliab Tarkghen  
*Systems Engineer*

Vanessa Ruvalcaba  
*Research Assistant*

Tamika McDowell  
*Sr. Administrative Assistant*

Kevin Cozzolino  
*Graduate Research Associate*

### RESEARCH ADVISORY COMMITTEE

David Weisburd, *Chair*  
*Hebrew University and University of Maryland*

Edward R. Maguire  
*George Mason University*

David Klinger  
*University of Missouri-St. Louis*

Tracey L. Meares  
*University of Chicago Law School*

### BOARD OF DIRECTORS

William G. Milliken  
***Chairman***

Hubert Williams  
***President***

George H. Bohlinger III

David D. Cole

Wade Henderson

William H. Hudnut III

David B. Lewis

W. Walter Menninger

Laurie O. Robinson

Weldon J. Rougeau

Alfred A. Slocum

Andrew L. Sonner

Maria Vizcarrondo-DeSoto



1201 Connecticut Avenue, NW, Washington, DC 20036-2636  
(202) 833-1460 • Fax: (202) 659-9149 • E-mail: [pinfo@policefoundation.org](mailto:pinfo@policefoundation.org)  
[www.policefoundation.org](http://www.policefoundation.org)