



Measuring Performance in a Modern Police Organization

Malcolm K. Sparrow

Executive Session on Policing and Public Safety

This is one in a series of papers that will be published as a result of the Executive Session on Policing and Public Safety.

Harvard's Executive Sessions are a convening of individuals of independent standing who take joint responsibility for rethinking and improving society's responses to an issue. Members are selected based on their experiences, their reputation for thoughtfulness and their potential for helping to disseminate the work of the Session.

In the early 1980s, an Executive Session on Policing helped resolve many law enforcement issues of the day. It produced a number of papers and concepts that revolutionized policing. Thirty years later, law enforcement has changed and NIJ and the Harvard Kennedy School are again collaborating to help resolve law enforcement issues of the day.

Learn more about the Executive Session on Policing and Public Safety at:

www.NIJ.gov, keywords "Executive Session Policing"

www.hks.harvard.edu, keywords "Executive Session Policing"

Introduction

Perhaps everything the modern police executive needs to know about performance measurement has already been written. But much of the best work on the subject is both voluminous and now more than a decade old, so there is no guarantee that today's police executives have read it. Indeed, it appears that many police organizations have not yet taken some of its most important lessons to heart.

I hope, in this paper, to offer police executives some broad frameworks for recognizing the value of police work, to point out some common mistakes regarding performance measurement, and to draw police executives' attention to key pieces of literature that they might not have explored and may find useful. I also hope to bring to the police profession some of the general lessons learned in other security and regulatory professions about the special challenges of performance measurement in a risk-control or harm-reduction setting.

A research project entitled "Measuring What Matters," funded jointly by the National Institute of Justice (NIJ) and the Office of Community Oriented Policing Services (COPS Office), led

to the publication in July 1999 of a substantial collection of essays on the subject of measuring performance.¹ The 15 essays that make up that collection are fascinating, not least for the divergence of opinion they reveal among the experts of the day. The sharpest disagreements pit the champions of the New York Police Department's (NYPD's) early CompStat model (with its rigorous and almost single-minded focus on reductions in reported crime as the "bottom line" of policing) against a broad range of scholars who mostly espoused more expansive conceptions of the policing mission and pressed the case for more inclusive and more nuanced approaches to performance measurement.

Three years later, in 2002, the Police Executive Research Forum (PERF) published another major report on performance measurement, *Recognizing Value in Policing: The Challenge of Recognizing Police Performance*, authored principally by Mark H. Moore.² PERF followed that up in 2003 with a condensed document, *The "Bottom Line" of Policing: What Citizens Should Value (and Measure!) in Police Performance*,³ authored by Moore and Anthony Braga.

Despite the richness of the frameworks presented in these and other materials, a significant proportion of today's police organizations seem to remain narrowly focused on the same categories of indicators that have dominated the field for decades:

- (a) Reductions in the number of serious crimes reported, most commonly presented as local comparisons against an immediately preceding time period.
- (b) Clearance rates.
- (c) Response times.
- (d) Measures of enforcement productivity (e.g., numbers of arrests, citations or stop-and-frisk searches).

A few departments now use citizen satisfaction surveys on a regular basis, but most do not. Clearance rates are generally difficult to measure in a standardized and objective fashion, so category (b) tends to receive less emphasis than the other three. Categories (c) and (d) — response times and enforcement productivity metrics — are useful in showing that police are getting to calls fast and working hard but reveal nothing about whether they are working intelligently, using appropriate methods or having a positive impact.

Therefore, category (a) — reductions in the number of serious crime reports — tends to dominate many departments' internal and external claims of success, being the closest thing available to a genuine crime-control outcome measure. These measures have retained their prominence despite everything the field is supposed to have learned in the last 20 years about the limitations of reported crime statistics. Those limitations (which this paper will explore in greater detail later) include the following:

- (1) The focus is narrow because crime control is just one of several components of the police mission.
- (2) The focus on serious crimes is narrower still, as community concerns often revolve around other problems and patterns of behavior.
- (3) Relentless pressure to lower the numbers, without equivalent pressure to preserve the integrity of the recording and reporting systems, invites manipulation of crime statistics — suppression of reports and misclassification of crimes — and other forms of corruption.
- (4) Focusing on reported crime overlooks unreported crimes. Overall levels of victimization are generally two to three times higher than reported crime rates.⁴ Particularly low reporting rates apply to household thefts, rape, other sexual assaults, crimes against youths ages 12 to 17, violent crimes committed at schools, and crimes committed by someone the victim knows well.⁵
- (5) Pressure to reduce the numbers is counterproductive when dealing with invisible crimes (classically unreported or underreported crimes, such as crimes within the family, white collar crimes, consensual crimes such as drug dealing or bribery, and crimes involving intimidation). Successful campaigns against these types of crime often involve deliberate attempts to expose the problem by first driving reporting rates up, not down.⁶
- (6) A focus on crime rate reductions does not consider the costs or side effects of the strategies used to achieve them.
- (7) Emphasizing comparisons with prior time periods affords a short-term and very local perspective. It may give a department the chance to boast, even while its crime rates remain abysmal compared with other jurisdictions. Conversely, best performers (with low crime rates overall) might look bad when random fluctuations on a quarterly or annual basis raise their numbers. Genuine longer term trends may be masked by temporary changes, such as those caused by weather patterns or special events. More important than local short-term fluctuations are sustained longer term trends and comparisons with crime rates in similar communities. Pressure to beat one's own performance, year after year, can produce bizarre and perverse incentives.
- (8) Even if crime levels were once out of control, the reductions achievable will inevitably run out eventually, when rates plateau at more acceptable levels. At this point, the department's normal crime-control success story — assuming that reductions in reported crime rates had been its heart and soul — evaporates. Some executives fail to recognize the point at which legitimate reductions have been exhausted. Continuing to demand reductions at that point is like failing to set the torque control on a power screwdriver: first you drive the screw, which is useful work; but then you rip everything to shreds and even undo the value of your initial tightening. The same performance focus that initially produced legitimate gains becomes a destructive force if pressed too hard or for too long.

(9) A number is just a number, and reliance on it reduces all the complexity of real life to a zero or a one. One special crime, or one particular crime unsolved, may have a disproportionate impact on a community's sense of safety and security. Aggregate numbers fail to capture the significance of special cases.

Reported crime rates will always belong among the suite of indicators relevant for managing a complex police department, as will response times, clearance rates, enforcement productivity, community satisfaction and indicators of morale. But what will happen if police executives stress one or another of these to the virtual exclusion of all else? What will happen if relentless pressure is applied to lower the reported crime rate, but no counterbalancing controls are imposed on methods, the use of force, or the integrity of the recording and reporting systems? From the public's perspective, the resulting organizational behaviors can be ineffective, inappropriate and even disastrous.

If we acknowledged the limitations of reported crime rates and managed to lessen our dependence on them, then how would we recognize true success in crime control? And how might we better capture and describe it?

I believe the answer is the same across the full range of government's risk-control responsibilities, whether the harms to be controlled are criminal victimization, pollution, corruption, fraud, tax evasion, terrorism or other potential and actual harms. The definition of success in risk control or harm reduction is to

spot emerging problems early and suppress them before they do much harm.⁷ This is a very different idea from "allow problems to grow so hopelessly out of control that we can then get serious, all of a sudden, and produce substantial reductions year after year after year."

What do citizens expect of government agencies entrusted with crime control, risk control, or other harm-reduction duties? The public does not expect that governments will be able to prevent all crimes or contain all harms. But they do expect government agencies to provide the best protection possible, and at a reasonable price, by being:

- (a) Vigilant, so they can spot emerging threats early, pick up on precursors and warning signs, use their imaginations to work out what could happen, use their intelligence systems to discover what others are planning, and do all this before much harm is done.
- (b) Nimble, flexible enough to organize themselves quickly and appropriately around each emerging crime pattern rather than being locked into routines and processes designed for traditional issues.
- (c) Skillful, masters of the entire intervention toolkit, experienced (as craftsmen) in picking the best tools for each task, and adept at inventing new approaches when existing methods turn out to be irrelevant or insufficient to suppress an emerging threat.⁸

Real success in crime control — spotting emerging crime problems early and suppressing

them before they do much harm — would not produce substantial year-to-year reductions in crime figures because genuine and substantial reductions are available only when crime problems have first grown out of control. Neither would best practices produce enormous numbers of arrests, coercive interventions or any other specific activity because skill demands economy in the use of force and financial resources and rests on artful and well-tailored responses rather than extensive and costly campaigns.

Ironically, therefore, the two classes of metrics that still seem to wield the most influence in many departments — crime reduction and enforcement productivity — would utterly fail to reflect the very best performance in crime control.

Furthermore, we must take seriously the fact that other important duties of the police will never be captured through crime statistics or in measures of enforcement output. As NYPD Assistant Commissioner Ronald J. Wilhelm wrote in a November 2013 internal NYPD strategy document:

[W]e cannot continue to evaluate personnel on the simple measure of whether crime is up or down relative to a prior period. Most importantly, CompStat has ignored measurement of other core functions. Chiefly, we fail to measure what may be our highest priority: public satisfaction. We also fail to measure quality of life, integrity, community relations, administrative efficiency, and

employee satisfaction, to name just a few other important areas.⁹

Who Is Flying This Airplane, and What Kind of Training Have They Had?

At the most recent meeting of the Executive Session, we asked the police chiefs present, “Do you think your police department is more or less complicated than a Boeing 737?” (see photograph of Boeing 737 cockpit). They all concluded fairly quickly that they considered their departments more complicated and put forward various reasons.

First, their departments were made up mostly of people, whom they regarded as more complex and difficult to manage than the electrical, mechanical, hydraulic and software systems that make up a modern commercial jetliner.

Second, they felt their departments’ missions were multiple and ambiguous, rather than single and clear. Picking Denver as a prototypical flight destination, they wondered aloud, “What’s the equivalent of Denver for my police department?” Given a destination, flight paths can be mapped out in advance and scheduled within a minute, even across the globe. Unless something strange or unusual happens along the way, the airline pilot (and most likely an autopilot) follows the plan. For police agencies, “strange and unusual” is normal. Unexpected events happen all the time, often shifting a department’s priorities and course. As a routine matter, different constituencies have different priorities, obliging police executives to juggle conflicting and sometimes irreconcilable demands.



Photograph of Boeing 737 Cockpit
Source: Photograph by Christiaan van Heijst, www.jpvanheijst.com, used with permission.

Assuming that for these or other reasons, the answer is “more complicated,” then we might want to know how the training and practices of police executives compare with those of commercial pilots when it comes to using information in managing their enterprise.

The pilot of a Boeing 737 has access to at least 50 types of information on a continuing basis. Not all of them require constant monitoring, as some of the instruments in the cockpit beep or squeak or flash when they need attention. At least 10 to 12 types of information are monitored constantly. What do we expect of pilots? That they know, through their training, how to combine different types of information and interpret them in

context, so they can quickly recognize important conditions of the plane and of the environment and know how they should respond.

A simple question like, “Am I in danger of stalling?” (i.e., flying too slowly to retain control of the aircraft) requires at least seven types of information to resolve: altitude, air temperature, windspeed, engine power, flap deployment, weight and weight distribution, together with knowledge of the technical parameters that determine the edge of the flight envelope. Some of these factors relate to the plane, and some relate to external conditions. All these indicators must be combined to identify a potential stall.

Thanks to the availability of simulators, commercial airline pilots are now trained to recognize and deal with an amazing array of possible scenarios, many of which they will never encounter in real life.¹⁰ They learn how various scenarios would manifest themselves through a variety of indicators, so they can recognize them quickly, diagnose them accurately and respond appropriately.

With so many forms of information available to the pilot, we might wonder how many of the dials in the cockpit are designated key performance indicators or, for that matter, how many of them are regarded as performance indicators at all. The answer is none. No single dial in the cockpit tells the pilot how well he or she is doing or is used to judge performance. Airline pilots appreciate and use an extensive array of different types of information, habitually use them in combination and in relation to one another, but are very slow to label any of them performance indicators.

By contrast, many public sector agencies (and this problem is by no means confined to police departments) pay close attention to a comparatively small number of indicators and seem all too keen to select one or two of them (usually no more than five) as key performance indicators. Designating a measure as a performance indicator usually means we determine in advance whether we expect it to move up or down, and we may even set a particular level as a target. Public sector executives then often seem surprised when the narrow range of information monitored produces partial or inadequate interpretations

of what is really happening and when the narrow performance focus drives behaviors that turn out to be perverse and contrary to the public interest.

Executive Session participants were quick to point out the lack of “simulator training” currently available for senior police managers. Simulations in the form of practical exercises are frequently used in entry-level training for recruits and in training courses for specialty functions (e.g., hostage negotiations, SWAT). For executives, simulator-style training is occasionally available in crisis leadership courses, where trainees are invited to take their turn at the helm in a crisis response exercise. But absent a crisis, most executive teams operate without any special training to help them interpret the myriad signals available or recognize important conditions quickly and pick the best response to different scenarios.

In the absence of such training, many executive teams muddle through, having learned most of what they know through their own experience on the way up through the managerial ranks rather than through formal training. As one chief noted, the closest equivalent to executive-level simulator training is when one department has the opportunity to learn from the misery of another. A collegial network of police executives, ready to share both their successes and failures, is a valuable asset to the profession.

Any “numbers game” is, of course, a simplification. Carl Klockars expresses this beautifully (and with a nod to Arthur “Dooley” Wilson) in *Measuring What Matters*:

SCENARIOS FOR POLICE EXECUTIVES

Imagining a simulator for police executives might be interesting. Here are three scenarios:

(1) You notice the following:

- (a) The conviction rate for cases prosecuted is unusually low and falling.
- (b) The proportion of complaints against officers that are internally generated is rising.
- (c) The number of internal requests for transfers is unusually high, as is the use of sick days.
- (d) Based on community surveys, public confidence in the agency is rising, but the response rate on surveys is falling.
- (e) The reported crime rate is dropping fast.

What is happening? What else would you want to know? What might you do?

(2) You notice the following:

- (a) Drug-related arrests are up 30 percent over last year.
- (b) The department has just (proudly) announced “record drug seizures” for the last quarter.
- (c) The proportion of drug-related prosecutions that result in jail terms has fallen to record low levels.
- (d) Street-level drug prices are at record lows.

What is happening? What else would you want to know? What might you do?

(3) You notice the following:

- (a) The number of domestic violence calls received by the department has risen sharply over the last 6 months.
- (b) The number of homicides determined to be domestic violence-related has dropped by one-third in the same period.
- (c) The proportion of domestic violence arrests that result in charges has just dropped below 40 percent for the first time.
- (d) The proportion of domestic violence cases prosecuted that result in convictions has also fallen below 70 percent for the first time.

What is happening? What else would you want to know? What might you do?

These three scenarios were presented to the Executive Session in the context of our discussions on

performance measurement. Given the time limitations, we did not have the chance to explore them at length, and none of the police leaders present ventured an opinion as to what he or she thought might be happening in each case. But they did agree that they would quite likely reach very different conclusions from one another, given their varied experiences and biases.

These scenarios are relatively simple ones involving only four or five indicators. Even so, using these indicators in combination and treating them as important management information might lead one to very different conclusions than treating one or another of them as a performance metric and specifying whether it should move up or down or reach some target level.

Scenario 1 might reveal an agency in crisis with plummeting morale. The community may be giving up on the police, no longer bothering to report crimes. Only “friends of the department” bother to respond to surveys; everyone else has given up on them. Corruption problems may be surfacing, evidenced by a higher rate of internally generated complaints and loss of cases in court, perhaps due to tainted evidence or community distrust.

Scenario 2 might suggest a failing drug-control strategy, with numerous arrests exhausting the capacity of the justice system but no positive or structural impact on the drug problem.

Scenario 3 might reveal the early stages of an energetic and successful campaign against domestic violence, which is helping to expose the problem more effectively while using arrests and prosecutions aggressively as deterrence tactics (even if prosecutions are not taken all the way through the system).

Notice that the two scenarios suggesting various forms of failure and trouble (1 and 2) show a reduction in reported crime rates (traditionally regarded as a sign of success). One of them (scenario 2) shows high levels of enforcement productivity.

The only story here that suggests an effective crime-control strategy (scenario 3) shows an increase in reported crime rates and a decrease in the conviction rate (traditionally signals of poor performance).

Even these relatively straightforward examples show how too close a focus on one or two classes of metric could blind an organization to meaning and context and possibly mislead everyone about what is really happening.

In every instance of measurement, the conversion of a thing, event, or occasion to a number requires ignoring or discarding all other meaning that thing, event, or occasion might have. The easy way to appreciate this very hard point in all its paradox and irony is to remember this: a kiss is just a kiss, a sigh is just a sigh, and a crime is just a crime, as time goes by. (Which, of course, anyone who has kissed, sighed, or committed, investigated, or been the victim of a crime knows is not true.)¹¹

Filmmakers and writers of police television dramas seem to have picked up on the collision that happens, somewhere in the middle ranks of a police department, between high-level general strategies (often numbers driven) and the rich texture of real life. It does not really matter whether one is watching Detective McNulty in “The Wire,” Detective Chief Superintendent Foyle in “Foyle’s War,” or any other fictional midrank police officer.

At some point or another, in most episodes, an assistant commissioner or some other senior officer is going to show up unexpectedly at the local police station. Never is his visit good news! The senior officers are invariably portrayed as interfering nitwits, ignorant and dismissive of what is happening at the street level. Often, they appear with their own new pet initiatives, requiring this many cases to be made or that many arrests of a particular type. Their agendas seem motivated by political, personal or

career-enhancing imperatives or are blatantly corrupt. The senior officers never listen and never help, but walk in with their general solutions that are not well thought out and demand a campaign of arrests or prosecutions that they claim will serve as a visible public demonstration of how seriously the department takes such-and-such an issue. Such visits interrupt valuable work and create imperatives for frontline officers that seem unhelpful, absurd or not in the public interest. Morale suffers, and the star of the show is left to fend off the ridiculous pressures from above so that his or her subordinates can carry on with “real police work,” knowing full well it will never be recognized by the department.

Of course, incompetent and corrupt senior management always makes for good theater. But the creators of police dramas do seem to have picked up on a specific organizational dynamic relevant to policing. The episode is about a specific scenario, which has all the quirky peculiarities of real life. That is what makes each episode unique, interesting and worth watching. The senior officers are unflatteringly portrayed as besotted with a general strategy and a narrow performance focus, neither of which does justice to the complex and varied nature of police work.

If a department appreciates that complexity; does well on vigilance, nimbleness, and skill; and therefore excels at spotting emerging problems early and suppressing them before they do much harm, what will success look like? How will such a department demonstrate its crime-control expertise?

The answer is that its performance account will consist largely of problem-specific project accounts describing emerging crime patterns and what happened to each one. Each project account will describe how the department spotted the problem in the first place, how it analyzed and subsequently understood the problem, what the department and its partners did about the problem, and what happened as a result. Some in policing call that the scanning, analysis, response and assessment (SARA) model. It is a straightforward, problem-oriented account that has little to do with aggregate numbers of any particular kind.

Broader Frameworks for Monitoring and Measurement

If the analogy of the cockpit turns out to be useful, then the following sections may help senior managers develop a broader sense of the banks of instruments they might want to have available in their cockpits and a clearer sense of how to use them for different purposes. To understand the diversity of indicators that could be relevant and useful in a police department's cockpit, we should remember (1) the breadth of the policing mission, (2) the multiple dimensions of performance that are therefore relevant, (3) the different managerial purposes that measurement can serve, and (4) the different types of work that occur within an agency, each of which naturally generates different types of information.

Much literature is available on all of these subjects, but for the sake of efficiency here I will refer readers to one principal author for each:

Herman Goldstein, for breadth of mission;¹² Mark Moore, for multiple dimensions of performance;¹³ Robert Behn, for different managerial purposes;¹⁴ and my own work on operational risk control for different types of work.¹⁵

Broadening the Frame: The Mission of Policing

In his 1977 book, *Policing a Free Society*, Herman Goldstein summarized the functions of the police:¹⁶

- (1) To prevent and control conduct widely recognized as threatening to life and property (serious crime).
- (2) To aid individuals who are in danger of physical harm, such as the victims of criminal attacks.
- (3) To protect constitutional guarantees such as the right of free speech and assembly.
- (4) To facilitate the movement of people and vehicles.
- (5) To assist those who cannot care for themselves: the intoxicated, the addicted, the mentally ill, the physically disabled, the old and the young.
- (6) To resolve conflict, whether between individuals, groups of individuals, or individuals and their government.
- (7) To identify problems that have the potential to become more serious for the individual citizen, the police or the government.
- (8) To create and maintain a feeling of security in the community.

This template, now 37 years old, seems a solid place to start and perfectly usable today as a frame for describing the multiple components of the police mission. Each of the eight areas Goldstein mentions might constitute one chapter in a police department's annual report. Perhaps, at the outset, the first chapter (crime control) might be more thickly populated than some of the others, but with some effort a better balance might be achieved.

Goldstein's framework pushes police to view their role more broadly than as part of the criminal justice system. The community policing movement and the evolution of CompStat systems into broader CitiStat systems¹⁷ have pushed police departments farther in recognizing their role as a part of municipal government. As Moore and Poethig point out in their *Measuring What Matters* essay:

If we conceive of the police as nothing more than "the first step in the criminal justice system," then we might easily miss the contributions that they make "outside the box" of crime control, law enforcement, and arresting people. On the other hand, if we conceive of the police as an agency of municipal government that shares with other agencies the broad responsibility for strengthening the quality of urban life, then we are in a better position to notice that the police contribute much more to those goals than is captured by the simple idea of reducing crime. We also notice that the police have capabilities that go

far beyond their ability to make arrests and that these capabilities are valuable to the enterprise of city government. In short, the police are a more valuable asset when viewed from the vantage point of trying to strengthen urban life than they are when viewed from the narrower perspective of reducing crime through making arrests.¹⁸

An obvious example of this broader role, as Moore and Poethig point out, is emergency response:

[P]artly because the police department is the only agency that works 24 hours a day, 7 days a week, and makes house calls, police will continue to be the "first responders" to a wide variety of emergencies. These emergencies can be medical (although ambulance services increasingly take care of these) or they can be social, such as deranged people threatening themselves or others, homeless children found wandering the streets with no parents to care for them, or drunks at risk of freezing to death after falling asleep on a park bench.¹⁹

The terrorist attacks of September 11, 2001, brought another kind of multiagency network to the fore. Police departments, especially those in major cities, were now required (and with no diminution of their other responsibilities) to participate in a range of counterterrorism, domestic security and intelligence collaborations at regional, national and international levels. Reporting systems that focused too heavily on

local crime statistics would likely overlook their contributions on this front. Of course, much of the work conducted in the security domain cannot be discussed in public as readily as routine crime-control activities. Nevertheless, it would be a shame not to include this aspect of the modern police mission in a department's performance account, even if parts of the story could be presented only to those stakeholders authorized to hear it.

Broadening the Frame: Dimensions of Performance

Working from Goldstein's (and other commentators') broader sense of the policing mission, Mark Moore develops a framework for holding police departments accountable. In *Recognizing Value in Policing: The Challenge of Measuring Police Performance*, he and his coauthors explore the range of data types and methods of observation that could provide a basis for assessing police performance in each of the following seven dimensions:²⁰

- (1) Reducing criminal victimization.
- (2) Calling offenders to account.
- (3) Reducing fear and enhancing personal security.
- (4) Guaranteeing safety in public spaces (including traffic safety).
- (5) Using financial resources fairly, efficiently and effectively.

(6) Using force and authority fairly, efficiently and effectively.

(7) Satisfying customer demands/achieving legitimacy with those policed.

Moore notes that reported crime statistics only partially capture levels of criminal victimization and that criminal victimization is only one of the seven dimensions of performance that matter to the public. In particular, he takes issue with the notion that crime statistics should be treated as the "bottom line" of policing. William J. Bratton had stated this view in an essay he contributed to NIJ's *Measuring What Matters*: "Crime statistics have become the department's bottom line, the best indicator of how the police are doing, precinct by precinct and citywide."²¹

Moore says the analogy with a commercial firm's bottom line is flawed because the crime statistics themselves do not capture the costs of the methods used to achieve them. Reductions in crime levels ought, he says, to be treated as part of gross revenues (not net revenues), and the use by police of financial resources and especially the use of force or authority should be regarded as costs and therefore counted against the revenues.²² The "bottom line" is therefore served if police manage to bring down crime rates and are economical with their uses of force and money.

[C]rime reduction is closer to the idea of the *gross revenues* a private company earns by producing and selling particular products and services, not its *profit*. [...]

Managers ought to be interested in trying to widen the difference between the valuable results the police produce (reduced crime), and the costs incurred in producing those results.²³

In the early days of CompStat in New York City, with widespread use of aggressive street order maintenance tactics, it did not appear that huge numbers of arrests were regarded as a cost. In fact, the number of arrests seems to have been a source of pride for the department. George Kelling, in the essay he contributed to *Measuring What Matters*, commented:

Even by more widely touted measurements, New York police do relatively well; so many people have been arrested that neither jails nor prisons can hold them. If the number of cells was expanded, few doubt that New York City police could fill almost any added capacity as well.²⁴

Carl Klockars, in his essay for *Measuring What Matters*, stressed the importance of the skill of minimizing the use of force.²⁵ He points out how much officers vary in this regard and in their predisposition to resort to violent or nonviolent means to exercise control in their encounters with civilians:

In any police agency there are officers who are well known for their ability to walk into an out-of-control situation and stabilize it peacefully. (There are others, of course, who can turn any situation

into a riot.) The skill of such officers is [in] knowing how to work in ways that minimize the use of force.²⁶

Stuart Scheingold, in his essay for *Measuring What Matters*, warned that "... the kinds of police practices associated with zero-tolerance and hyper law enforcement seem likely to increase the mistrust of the police that robs crime control of its consensus building capacity."²⁷ Scheingold also recalls the warnings about the effects of zero tolerance and hyper law enforcement that Wesley Skogan provided in his 1990 book, *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*:

Even if they are conducted in a strictly legal fashion, aggressive tactics such as saturating areas with police, stopping cars frequently, conducting extensive field interrogations and searches, and bursting into apartments suspected of harboring gambling or drugs can undermine police-community relations in black and Hispanic neighborhoods.²⁸

The underlying tensions between police and predominantly minority communities may amplify these costs, according to Skogan:

[R]esidents of poor and minority neighborhoods with serious disorder problems often have antagonistic relations with the police. They regard the police as another of their problems, frequently perceiving them to be arrogant, brutal, racist, and corrupt.²⁹

Moore, along with these and many other commentators, has urged police to recognize coercive power as a precious commodity to be used sparingly and invites them to search for artful interventions that use authority efficiently rather than defaulting to massive enforcement campaigns of one kind or another.³⁰ The point is not to rule out enforcement campaigns where these are required. Rather, the point is that a full accounting of police performance demands serious attention to the costs as well as the gross revenues of such campaigns.

Moore also proposes an expansive view of client satisfaction that includes an assessment of the experience of those arrested or cited. Not that we would expect them to be “pleased” that they were caught! But we would hope to hear that their rights were respected, that excessive force was not used against them, and that they were treated with dignity and courtesy even while being brought to book.³¹

Broadening the Frame: Purposes of Measuring Performance

Robert Behn’s 2003 article, “Why Measure Performance? Different Purposes Require Different Measures,”³² has recently been celebrated as one of the most influential articles in the history of the journal in which it was published, *Public Administration Review*. I often recommend this as foundational reading for practitioners wanting to rethink their department’s use of metrics and indicators.

In this article, Behn lays out eight different managerial purposes that may be served by

Table 1. Eight Purposes That Public Managers Have for Measuring Performance

Purpose: The public manager’s question that the performance measure can help answer.	
Evaluate	How well is my public agency performing?
Control	How can I ensure that my subordinates are doing the right thing?
Budget	On what programs, people or projects should my agency spend the public’s money?
Motivate	How can I motivate line staff, middle managers, nonprofit and for-profit collaborators, stakeholders, and citizens to do the things necessary to improve performance?
Promote	How can I convince political superiors, legislators, stakeholders, journalists and citizens that my agency is doing a good job?
Celebrate	What accomplishments are worthy of the important organizational ritual of celebrating success?
Learn	Why is what working or not working?
Improve	What exactly should who do differently to improve performance?

Source: Robert D. Behn, “Why Measure Performance? Different Purposes Require Different Measures,” *Public Administration Review* 63 (5) (Sept./Oct. 2003): 586-606, table 1.

various types of performance measurement and monitoring. He summarizes these in a table,³³ reproduced here as table 1.

Behn also explores the different types of indicators that will serve these eight managerial purposes.³⁴

For example, evaluation of agency performance at the aggregate level (purpose 1) will involve attempts to link agency outputs with societal outcomes and will need to consider the effects of exogenous factors.

Motivating teams internally (purpose 4) may use real-time, functionally specific activity and productivity metrics disaggregated to the team level. These might be compared with production

targets or even pitted against one another in an internal competition.

Organizational learning (purpose 7) may rest more on disaggregated data that can reveal anomalies and deviations from norms, revealing problems to be understood and resolved.

Improvement of departmental processes (purpose 8) will require linking internal process design changes with their effects on agency outputs, efficiency and outcomes.

Broadening the Frame: Recognizing Different Types of Work

The fourth way to broaden the frame is to recognize the characteristics of different classes of work that coexist within a complex organization.

First, and most familiar, is functional work, which groups workers with other similarly skilled workers. Investigators work together within a detective bureau, auditors in the audit bureau, educators in the training department, and lawyers in the general counsel's office. Functional professionalism requires staying current, skilled and state-of-the-art.

For any specialist functional unit, the most readily available metrics cover quantity and quality. Quantity is easiest to measure: how many cases opened and closed, how many audits completed, how many hours of training delivered. But each function should also have and emphasize its own quality metrics lest concerns over quantity drown out concerns over quality.

Ideas about quality may include indications of significance or relevance (of a case or an audit) as well as assessments of whether the activities are conducted more or less professionally. A functional unit, if it divides its output quantity by its inputs, may also furnish indications of productivity or efficiency (e.g., the number of cases closed per investigator, the number of audits completed per auditor or the number of training hours delivered per instructor).

Second, and also familiar, is process-based work. By this I mean any kind of transactional work that is repeated hundreds or thousands of times. Obvious examples include tax agencies processing tax returns, environmental agencies issuing or renewing environmental permits for industry, and a licensing authority processing applications, as well as police departments responding to emergency calls or to complaints from the public about officer conduct. Given the repetitive nature and volume of such work, organizations invest in process design (involving procedures, protocols, automation and flow patterns) to handle the loads in expeditious, smooth and predictable ways.

Five categories of management information are associated with every process:

- (1) *Volume*: Measures of transaction volume (and trends in volume over time). This is important management information, as it affects resource allocation decisions even though the incoming volume may not be under the organization's control.

- (2) *Timeliness*: How long it takes to process transactions. Many processes have mandatory timelines. Measures used may include average processing times, worst-case times, or percentage of the transaction load that is handled within mandated or specified time periods.
- (3) *Accuracy*: The proportion of the determinations or judgments made that turn out to be “correct.” Metrics on accuracy usually come from case reviews or audits of samples conducted after the fact. Getting a decision right early is preferable to getting it right eventually (after review or on appeal), as it is quicker, cheaper and generally more satisfactory for all the stakeholders.
- (4) *Cost-efficiency*: Indicators of the cost of running the process divided by the transaction volume. (Processing costs per transaction can be driven down through process improvement, process redesign and automation.)
- (5) *Client satisfaction*: Almost every process transaction involves a client (a caller, a complainant, an applicant). Clients can be sampled and surveyed retrospectively to determine the nature of their experience with the process. Even if the client did not get the decision he or she wanted, he or she can still legitimately comment on whether the process seemed fair and efficient and whether he or she was treated properly.

The third type of work, risk-based work, differs from functional work and process work. In the vocabulary of the police profession, the most

familiar phrase for this would be problem-oriented work. It is not about one particular method (functional), and it is not about quality management in the context of an established process; rather, it is about a specific risk concentration, issue or problem. For risk-based work, professionalism begins with spotting the problems early and discerning their character and dynamics, long before anyone decides which tactics might be relevant or whether any of the organization’s routine processes touch the problem at all.

The following extract from *The Character of Harms* describes some of the frustrations that arise when agencies depend too heavily on functional and process-based metrics and fail to recognize problem-oriented or risk-based work as different:

Even the combination of [the functional and process-based] performance stories falls short with respect to demonstrating effective harm-control. It might show the agency worked hard to apply the functional tools it has, and that it operated its established processes with alacrity and precision. But it leaves open the question of whether tool selection is effective, and whether the processes established touch the pressing issues of the day. The audience can see agencies working hard, which they like. But they would like to be convinced they are also working effectively on the problems they choose to address, and that they are identifying and selecting the most

important issues to address. Neither the function-based performance story (e.g., “we completed this many high-quality investigations”) nor the process-based version (“we cleared our backlogs and streamlined our system”) provide such assurance.³⁵

The risk-based or problem-oriented performance story is quite different from the other two. It consists of a collection of project-based accounts, each comprising five elements:³⁶

- (1) A description of the data or intelligence that alerted the department to the problem and showed it to be of sufficient scope and significance to warrant special attention.
- (2) A description of the metrics selected (which are project-specific) to be used in determining whether the condition “got better”; in other words, outcome or impact metrics through which project success or failure might be judged.
- (3) A description of the action plan adopted and maybe a second or third one used if and when it was determined that the first one failed.
- (4) A description of what happened to the project metrics over time and how the project team interpreted those changes.
- (5) A decision about project closure, hopefully with the problem sufficiently abated that resources could be withdrawn so the department could move on to the next pressing issue. Continuing attention to the

problem would enter a longer term (and less resource-intensive) monitoring and maintenance phase.

If a department succeeds in its crime-control mission, the heart of its success story will be the collection of project-based accounts, each describing an emerging crime pattern spotted early and dealt with skillfully. The more vigilant the department becomes in spotting emerging problems early, the less available significant crime reductions will be. Aggregate levels of crime may remain low but relatively steady, even as the department works hard to spot new threats before they have a chance to grow out of control.

Criminologists and other evaluators of police performance — who tend to use changes in the aggregate reported crime rate as the outcome variable in their analysis — may not recognize best practice, as the crime reductions visible at the aggregate level may not be statistically significant or may not be present at all.³⁷

However, the risk-based performance account — which describes how the department kept the aggregate rates low — is always available. That account relies much less on aggregate crime numbers and instead describes the collection of problem-specific interventions. Relevant metrics are tailored to each project and derive from disaggregated data filtered to fit the problem being addressed.

On Setting Goals Using Performance Metrics

With the distinctions between functional, process-based and risk-based work clarified, it might be worth observing which types of metrics associated with each type of work could or should be used for goal-setting — that is, as performance metrics.

Some departments set targets for functional outputs, including enforcement activities such as arrests, stops, searches and traffic citations. This only makes sense in particular circumstances and should never be the default position or become normal practice. If you want quality work from a carpenter, it makes no sense to demand that he or she drill a certain number of holes or hammer a quota of nails. The essence of craftsmanship involves mastery of all the tools and the ability to select among them based on a clear understanding of the specific task in hand. Functional quotas make little sense in this context.

The use of quotas for enforcement activity is always contentious, subject to officer whistleblowing and media scrutiny.³⁸ The danger is that setting enforcement goals may bias the judgment of police officers in potentially dangerous ways and may contribute to aggressive or oppressive police conduct. Some jurisdictions have banned the use of quotas for traffic tickets.³⁹ Mandating a certain level of enforcement activity drives up the costs of policing for society and therefore reduces the “bottom line.” There may be smarter ways, more economical with respect

to the use of force or authority, to procure the compliance or behavioral changes sought.

Some departments deny setting quotas but admit to using enforcement productivity as a measure of patrol officers’ effort — in other words, as a performance metric. In terms of the effect on police operations, the two ideas are virtually indistinguishable. Measures of overall productivity spanning the full range of patrol tasks (including but not limited to enforcement) can reveal over time who is working hard and who is not. That, of course, is important management information.

The danger of setting goals specifically for enforcement actions is that it may incline officers to stretch the limits of legality and fairness for the sake of another arrest or ticket. It is liable to produce quantity without regard to quality, relevance or side effects. It undermines the importance of discretion and judgment exercised at the frontline.

Despite these general cautions, the use of some kind of enforcement campaign might still be appropriate in specific circumstances, but the need would arise only when a specific type of enforcement focus has been selected as the solution to address a specific crime problem. All of the following conditions would need to be satisfied:

- (1) A specific pattern of crime is being addressed.
- (2) The problem has been properly analyzed.

- (3) All the intervention options have been given full and open-minded consideration.
- (4) An action plan has been chosen that requires a period of intensive enforcement attention to specific violations (which we might call an enforcement campaign).
- (5) Instructions to frontline officers make it clear that, despite the context, every enforcement action must, given the circumstances of the case, be legally justifiable and appropriate.
- (6) During the implementation of the plan, management constantly and carefully monitors the legality, reasonableness, relevance, impact and side effects of the enforcement activities so that operations can wind down or change course as soon as is appropriate.

In this context, the use of functional focus fits Behn's managerial purposes of control (to make sure employees are carrying out the plan) and motivation (to increase intensity of effort). But the plan being carried out is one thoughtfully designed in response to a specific crime problem. The campaign is appropriate, therefore, because it constitutes an intervention for a specific problem. This differs from setting enforcement quotas just because "that's what we do to show we're serious" or because particular executives have an ideological preference for certain types of activity. A wise craftsman does not "believe in" any one tool any more than any other. Enforcement campaigns should be seldom used and their occasional use should be thoughtful, carefully monitored and temporary.

If unrelenting pressure is applied to police officers to meet activity quotas of any kind (enforcement-related or not), they will surely find ways to produce quantity, even at the expense of quality, relevance, appropriateness or their own better judgment. If the goals set are unreasonable or not achievable through legitimate means, then illegitimate means may well be employed, producing behaviors not in the public interest.

As Andrew P. Scipione, Commissioner of Police in New South Wales, Australia, wrote in 2012:

There is a delicate balance. A preoccupation with numbers is unhealthy if it distracts from the primary need to apprehend the most serious criminals and care for the most traumatized victims; it is unhealthier still if it causes police on the streets to set aside sound judgment and the public good in the pursuit of arrest quotas, lest they attract management criticism or compromise their chances of promotion.⁴⁰

What about setting process-based goals? For core processes, it rarely makes sense to set goals for the incoming volume of transactions, as that is not usually under the organization's control. It is appropriate to set goals for timeliness in processing, accuracy in determinations and client satisfaction. All three of these are candidates for close-to-real-time monitoring and target-setting. But managers should be careful to adjust the balance between them, as too great an emphasis

on timeliness can end up damaging accuracy and client satisfaction.

Cost-efficiency is more often improved by periodically rethinking or redesigning the underlying processes (using technology and business process improvement methods) than by exerting constant pressure on operational staff.

What about setting goals for risk-based work? Executives should demand vigilance from their crime analysis and intelligence staff so that their department spots emerging problems earlier rather than later. Executives should demand nimbleness and fluidity from those responsible for allocating resources within the system so that the right set of players can be formed into the right kind of team, at the right level, to tackle each unique pattern of crime or harm that appears. In addition, executives can emphasize their expectations for creativity and skill from those to whom problem-solving projects are entrusted.

Those involved in risk-based (problem-oriented) work should retain a strong methodological focus on achieving relevant outcomes rather than churning out enormous volumes of familiar outputs. It is tempting, once a plausible plan has been selected that involves activities, to allow the monitoring systems to focus on the production of those activities rather than achievement of the desired outcomes. What starts as thoughtful risk-based work can morph over time into functional quotas — not least because functional quotas are much easier to monitor.

A Closer Look at Crime Reduction as the Bottom Line of Policing

Holding these broader frameworks in mind, let us return for a moment to the dangers of relying too heavily on reported crime rates as the bottom line for policing. More needs to be said about the ways in which a relentless focus on reductions in reported crime rates (a) produces too narrow a focus and (b) invites manipulation of statistics and other forms of corruption.

The Dangers of Narrowness

The focus is narrow because crime control is just one of several components of the police mission. The focus on serious crimes is narrower still, as community concerns often revolve around other problems. The focus on aggregate numbers is misleading, as some crimes have a disproportionate impact on community life or levels of fear.

Focusing on reported crime overlooks unreported crimes, which are generally more numerous than reported crimes. Particular types of crime — sexual assaults, domestic violence and other crimes within the family, white-collar crimes, crimes involving intimidation or where victims have a reason not to want to go to the authorities — are all notoriously underreported. Klockars⁴¹ points out that even homicide rates may be unreliable to a degree. Apparent suicides, deaths reported as accidental, and unresolved cases of missing persons may all act as masks for some murders. “Particularly vulnerable to having their murders misclassified this way are transients,

street people, illegal aliens, and others who, if missed at all, are not missed for long."⁴²

Consensual crimes (such as bribery or drug dealing), with no immediate victim to complain, go essentially unreported and therefore appear in police statistics only if they are uncovered through police operations.

Most commentators recognize the difficulty and expense of carrying out victimization surveys in order to get better estimates. But most find no available alternative and recommend police departments consider the option seriously. As Moore and colleagues put it:

Determining the level of *unreported crime* is important not only to get a more accurate measure of the real rate of criminal victimization, but also to determine how much confidence citizens have in asking the police for help.

The only way to measure the underlying rate of victimization is to conduct a general survey of citizens asking about their victimization, and their reasons for failing to report crime to the police.

... [I]f one wants to get close to the real level of victimization, and to learn about the extent to which the police have earned citizens' confidence in responding to criminal offenses, then there is little choice but to complement information on reported crime with

information gained from general surveys of local populations.⁴³

Moore and his coauthors also note the efficiencies available once a system for conducting victimization surveys has been set up:

[W]e can use that system to *answer many other important questions about policing*. Specifically, we can learn a great deal about citizens' fears and their self-defense efforts, as well as their criminal victimization. We can learn about their general attitudes toward the police and how those attitudes are formed.⁴⁴

Moore and Braga⁴⁵ also point out the value of using other data sources as a way of cross-checking or validating trends suggested by police statistics. Particularly useful would be public health data from hospital emergency rooms that might reveal "the physical attacks that happen behind closed doors, or which are otherwise not reported to the police."⁴⁶ Langton and colleagues estimate that, nationwide, 31 percent of victimizations from 2006 to 2010 involving a weapon and injury to the victim went unreported to police.⁴⁷ Even when assaults involved the use of a firearm, roughly 4 in 10 cases went unreported to police.⁴⁸ Many such cases are visible to public health systems.

As noted earlier, pressure to reduce the numbers is counterproductive when dealing with the whole class of invisible crimes (classically unreported or underreported crimes). Successful campaigns against these types of crime often

involve deliberate attempts to expose the problem by first driving reporting rates up, not down.⁴⁹

Corruption in Reporting Crime Statistics

Relentless pressure to reduce the number of crimes reported, without equivalent pressure to preserve the integrity of the recording and reporting systems, invites manipulation of statistics. The most obvious forms of manipulation involve suppression of reports (failing to take reports of crime from victims) and misclassification of crimes to lower categories in order to make the serious crime statistics look better.

This ought not to surprise anyone. It is an obvious danger naturally produced by too narrow a focus on one numerical metric. John A. Eterno and Eli B. Silverman, in their 2012 book, *The Crime Numbers Game: Management by Manipulation*,⁵⁰ have compiled an extensive history of corruption scandals involving manipulation of crime statistics that spans decades and many countries. England and Wales, France, and the Australian states of New South Wales and Victoria have all experienced major scandals.⁵¹ In the 1980s, the Chicago Police Department suffered a major scandal, accused of “killing crime” on a massive scale by refusing to write up official reports of offenses reported to them.⁵² Detectives were caught “unfounding” (determining a case was unverifiable) complaints of rape, robbery and assault without investigation.⁵³ Other major U.S. cities have had similar scandals, including Baltimore, Washington, D.C., New York City, Atlanta, Boca Raton, Fla., and Chicago.⁵⁴

Cheating on crime statistics, as a formidable temptation for police departments, goes with the territory. It remains a current problem, with a continuing stream of allegations and inquiries in various jurisdictions. In the United Kingdom, Chief Inspector of Constabulary Tom Winsor testified to a Commons Committee in December 2013 that British police forces were undoubtedly manipulating crime statistics, and the question was only “where, how much, how severe?”⁵⁵ According to the *Guardian* newspaper, he later hinted that he thought the cheating might be on an “industrial” scale.⁵⁶

Two investigative reports published by *Chicago Magazine* in April and May 2014 raised new questions about manipulation of crime statistics by the Chicago Police Department, including the downgrading of homicide cases.⁵⁷ In Scotland, the U.K. Statistics Authority rejected the government’s claims to have achieved record low crime rates in August 2014 amid claims from retired police officers that crime figures had been manipulated “to present a more rosy picture of law and order to the public.”⁵⁸

Eterno and Silverman also show, through their survey of relevant public management literature, that such dangers are well understood in other fields and by now ought to be broadly recognized throughout the public sector.⁵⁹

In 1976, social psychologist Donald Campbell wrote:

The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort or corrupt the social processes it is intended to monitor.⁶⁰

Addressing the phenomenon of crime statistics in particular, Campbell notes:

Crime rates are in general very corruptible indicators. For many crimes, changes in rates are a reflection of changes in the activity of the police rather than changes in the number of criminal acts [references deleted]. It seems to be well documented that a well-publicized, deliberate effort at social change — Nixon’s crackdown on crime — had as its main effect the corruption of crime-rate indicators ... achieved through under-recording and by downgrading the crimes to less serious classifications.⁶¹

Diane Vaughan, who has studied many forms of organizational misconduct, says that trouble arises when the social context puts greater emphasis on achieving the ends than on restricting the means:

When the achievement of the desired goals received strong cultural emphasis, while much less emphasis is placed on the norms regulating the means, these norms tend to lose their power to regulate behavior.⁶²

What I have elsewhere termed “performance-enhancing risk-taking” turns out to be one of the most common sources of public-sector corruption. Officials take actions that are illegal, unethical or excessively risky, and they do so not because they are inherently bad people or motivated by personal gain but because they want to help their agency do better or look better and they are often placed under intense pressure from a narrow set of quantitative performance metrics.⁶³ As a consequence of the pressures, an internal culture emerges that celebrates those who “know what’s required and are prepared to step up” — whether that means torturing terror suspects to extract information, “testilying” to procure convictions, or cooking the books to achieve the crime reductions that executives demand.

Few in the police profession will need to be told how one might “cook the books.” Robert Zink, the Recording Secretary of NYPD’s Patrolmen’s Benevolent Association, put it plainly in 2004:

[S]o how do you fake a crime decrease? It’s pretty simple. Don’t file reports, misclassify crimes from felonies to misdemeanors, under-value the property lost to crime so it’s not a felony, and report a series of crimes as a single event. A particularly insidious way to fudge the numbers is to make it difficult or impossible for people to report crimes — in other words, make the victims feel like criminals so they walk away just to spare themselves further pain and suffering.⁶⁴

Integrity Issues Related to CompStat and Crime Reporting at the NYPD

There is good reason to pay special attention to the NYPD's early experience with CompStat. A great many other departments have copied it and now run similar systems or variants of it. The use of CompStat-style systems has therefore substantially affected how many police departments view performance measurement, both in the ways they describe their performance at the agency level and in the ways they hold subunits accountable. So the effects of such systems — both positive and negative — need to be broadly understood.

The original form of CompStat as implemented by the NYPD in 1994 had a set of very particular features.

The system focused almost exclusively on reported rates for Part I Index crimes.⁶⁵ The principal data source was “reported crime,” and the performance focus was unambiguously and always to “drive the numbers down.” Crime analysis was principally based on place and time so that hot spots could be identified and dealt with. Accountability for performance was organized geographically and by precinct and was therefore intensely focused on precinct commanders. The precinct commanders were required to stand and deliver their results, or explain the lack of them, at the “podium” in the CompStat meetings — a famously stressful experience.

The managerial style employed seemed combative and adversarial. Garry McCarthy,

who experienced the system in New York and replicated it in Chicago, recalled the New York experience of CompStat meetings in an interview with *Chicago Magazine*: “When I was a commander in New York, it was full contact,” he said in 2012. “And if you weren’t careful, you could lose an eye.”⁶⁶

The early days of CompStat also focused on a particular class of offenders and offenses and promoted a somewhat standardized response. As Bratton wrote in 1999, “Intensive quality-of-life enforcement has become the order of the day in the NYPD.”⁶⁷ The targets were disorganized street criminals, as Bratton noted: “In fact, the criminal element responsible for most street crime is nothing but a bunch of disorganized individuals, many of whom are not very good at what they do. The police have all the advantages in training, equipment, organization, and strategy.”⁶⁸ The primary intervention strategy solution was aggressive street-order maintenance: “We can turn the tables on the criminal element. Instead of reacting to them, we can create a sense of police presence and police effectiveness that makes criminals react to us.”⁶⁹

The CompStat system was also used to drive enforcement quotas throughout the city for arrests, citations and stops/searches as well as to press senior management’s demands for reductions in reported crime rates.

There is no doubt that the system was effective in driving performance of the type chosen. As Bratton wrote: “Goals, it turns out, are an extremely important part of lifting a

low-performing organization to higher levels of accomplishment and revitalizing an organizational culture.”⁷⁰

In 2002, Moore described the early implementation of CompStat in the NYPD thus:

It was close to a strict liability system focused on outcomes. As a result, it is not surprising that the system ran a strong current through the department. Given that the system was consciously constructed to be behaviorally powerful, it becomes particularly important to know what it was motivating managers to do.⁷¹

Eterno and Silverman surveyed retired NYPD managers at or above the rank of captain to try to understand the effects of the CompStat system on the nature and extent of crime statistics manipulation. One of their survey findings was that respondents reported a diminished sense of pressure for integrity in reporting, even as CompStat imposed enormous pressure for crime reductions.⁷²

Public concerns about what CompStat was motivating managers to do were heightened when the *Village Voice* published transcripts of Officer Adrian Schoolcraft’s secretly recorded audiotapes of roll calls and other police meetings. Journalist Graham Rayman authored five articles, published in *The Village Voice* from May 4 to August 25, 2010, exploring the manipulation of crime statistics and the use of arrest and stop-and-frisk quotas in Brooklyn’s Bedford-Stuyvesant’s 81st Precinct,

as well as a 2012 follow-up article confirming the results of the investigation into Schoolcraft’s allegations.⁷³

In addition, 2010 saw the first wave of internal investigations within the NYPD for crime statistics manipulation. A department spokesman told the *New York Times* in February 2010 that Commissioner Raymond Kelly had disciplined four precinct commanders and seven other senior managers for downgrading crime reports.⁷⁴ In October 2010, the department disciplined the former commander of the 81st Precinct and four others for downgrading or refusing to take crime reports.⁷⁵

In January 2011, Commissioner Kelly appointed a panel of three former federal prosecutors to examine NYPD’s crime statistics recording and reporting practices, giving them 3 to 6 months to report.⁷⁶ Sadly, one of the three died while the panel was conducting its investigation, but the remaining two published the final report on April 8, 2013.⁷⁷

The Crime Reporting Review (CRR) Committee acknowledged the danger of senior managers exerting pressure on subordinates to manipulate crime statistics and noted that there had been substantiated reports of manipulation in the past.⁷⁸ They reported that the NYPD’s own investigation of the 81st Precinct had “largely substantiated” Officer Schoolcraft’s allegations regarding precinct supervisors exerting significant pressure on patrol officers not only to meet quotas on activities such as issuing traffic

tickets but also to manipulate or not take crime reports.⁷⁹

The CRR Committee had not been asked to determine the extent to which the NYPD had manipulated or was manipulating crime statistics. Rather, their task was to examine the control systems in place to see if they were robust enough to guarantee integrity in the face of the obvious and substantial pressures of the CompStat environment.

First, the committee examined the process-level controls and protocols governing the taking and classifying of crime reports. They determined that these were insufficient to counteract the danger of intentional manipulation.⁸⁰

They then examined the audit systems that the NYPD used to review crime reports and crime classification decisions after the fact. They found, not surprisingly, that “the risk of suppression (as measured by its effect on reported crime) may be more substantial than the risk of downgrading.”⁸¹

With downgrading, there is at least a trail of records that can be reviewed, from the initial “scratch report” to the final system entry, and a set of policy criteria are available to test the determinations made. The system of audits conducted by the Quality Assurance Division (QAD) that the department used to test crime classifications uncovered patterns of downgrading with respect to robberies, burglaries and larcenies. QAD audits identified patterns of larcenies downgraded to lost property when complainants did not see their property

being stolen, robberies being downgraded to larcenies, burglaries being downgraded to larcenies by omitting a complaint of an unlawful entry, and attempted burglaries evidenced by broken locks or attempted breakdown of a door being classified as criminal mischief.⁸²

With crime suppression — failing to take a complaint, discouraging or intimidating a complainant, or otherwise making it awkward or unpleasant to report crime — there is often no written record to review, except in the case of 911 “radio runs” when the central call system retains a note of the original call.

The most plausible way to detect patterns of suppression is to audit the radio runs and revisit the callers, comparing what happened within the department with what should have happened as a result of the call. The NYPD implemented such audits (called SPRINT audits) in 2010. The CRR Committee noted that before 2010 no mechanisms were in place to determine the level of crime suppression, and hence there was no way of estimating the effect that any patterns of crime suppression may have had on NYPD’s recorded crime rates.⁸³ The committee also noted that implementation of the SPRINT audits had resulted in a relatively high rate of disciplinary actions, with investigations of 71 commands in 2011 and 57 commands in 2012, and administrative sanctions being sought against 173 officers as a result of those investigations.⁸⁴ The committee recommended a substantial increase in the breadth, depth and volume of SPRINT audits in the future as, at the time of the committee’s report (2013), these had been applied

only to limited categories of calls (robbery and burglary).⁸⁵

Wesley Skogan points out that this type of audit, even if used extensively, can still only provide a partial view of crime suppression practices.⁸⁶ Sampling the transactions that pass through a centralized 911 system misses all the other attempts by citizens to report crimes through the panoply of alternate channels now available — beepers, cell phones, voice mail and face-to-face reporting.⁸⁷

Almost nobody denies that the relentless pressure to reduce reported (or, to be more precise, recorded) crime rates produced patterns of misclassification and crime report suppression. The question of the extent to which manipulation of crime statistics became institutionalized at NYPD, or whether the cases of manipulation that have come to light were isolated, remains unresolved. This paper cannot and does not attempt to resolve that question. Nobody doubts that crime has fallen dramatically in New York City in the last 20 years and that the city is now much safer than it was before. The CRR Committee was keen to make that clear, noting in its conclusion that declines in overall crime rates in NYC over the previous 10 or 20 years were indeed “historic.”⁸⁸

The purpose here is not to undermine the achievements of the NYPD. CompStat was a major innovation when it was introduced and no doubt contributed substantially to the crime rate reductions achieved, even though criminologists still disagree about how much.⁸⁹

There is evidence that the NYPD has recognized some of the dangers of its original CompStat model. The NYPD has disciplined a substantial number of officers over issues of crime suppression and misclassification and increased the number and types of audits designed to help guarantee integrity in reporting. Precinct-level officers interviewed by the CRR Committee during the review period, 2011-2013, indicated that “the culture surrounding complaint-reporting had changed [improved] from ‘what it had been.’”⁹⁰ The NYPD has also backed off its use of enforcement quotas to some degree. These are signs that the NYPD is rebalancing the pressures acting on its officers and paying more attention to the means as well as to the ends of reducing reported crime rates.

Looking forward, and with a focus on best practice, I stress that intense pressure for a reduction in crime numbers is liable — for many reasons well understood throughout private industry and in public management — to produce corruption, manipulation of statistics, and other patterns of behavior not in the public interest, unless an equivalent and counterbalancing level of attention is paid to the integrity of recording and reporting systems, the governance of means and careful monitoring of side effects.

Departments that do great work, but leave themselves open to integrity risks, may face public cynicism about their self-reported results and have an ugly cloud over their successes.

Implications for CompStat-Like Systems

There has been much debate as to whether the early focus of CompStat and the methods that the system drove were ever right, even for New York City. The 1999 collection of essays, *Measuring What Matters*,⁹¹ reflects much of that debate. But this narrow form is certainly not appropriate now and certainly is not appropriate in general or in other jurisdictions.

In the past 20 years, the police profession has had the opportunity to learn a great deal about the strengths and weaknesses of the original CompStat model and to figure out what types of modifications to the original model can improve it or better adapt it for use in other jurisdictions. Many variations can now be found. Some of them still reflect particular and narrow aspects of the original form. Other versions are much broader, more mature, and seem both more versatile and, in some ways, more humane. Which version a department uses is likely to have a significant effect on its approach to performance measurement and reporting.

Table 2 shows six dimensions in which CompStat-like systems for crime analysis and accountability might be either narrower or broader. These six dimensions provide an opportunity for police executives to examine their own system and determine where, along each spectrum, their current CompStat implementation lies. Any dimension in which a system sits close to the narrow end represents an opportunity for improvement and might enable a department to replace its current model with a more versatile and less restrictive version.

Table 2. CompStat Implementation Options

	Narrow forms	Broader (mature) forms
Data sources	Reported crime rates	Multiple sources, including victimization surveys
Forms of analysis	Geographic (by precinct and cluster) and temporal	Versatile, considering full range of relevant dimensions
Performance focus	Drive the numbers down	Emphasize increased reporting to expose and deal with hidden problems
Locus of responsibility	Precinct commanders	Tailored to each problem
Managerial style	Adversarial	Cooperative/coaching
Preferred tactics	Directed patrol, street order maintenance	Full range of interventions

Many of the issues discussed so far in this paper would naturally push toward the broader end of each spectrum. Recognizing Goldstein's more inclusive view of the police mission, Moore's multiple dimensions of performance, and the need for organizational nimbleness and skill in addressing many different types of problems that police face should all help to shake police departments out of too narrow a focus.

Data sources: As this paper has already made clear, Part I index crimes remain important, but community concerns frequently center on other issues. Many crimes are not reported, and therefore police would need to use a broader range of data sources — including public health information and victimization surveys — even to be able to see the full range of problems that matter.

Forms of analysis: Crime analysis should no longer revolve solely or mainly around hot spot analysis. Goldstein repeatedly pressed police to recognize that crime problems were often not concentrated geographically but were concentrated in other dimensions (e.g., repeat offenders who roamed the city, repeat victims, methods of commission, patterns of behavior and types of victims).⁹² Therefore, adding forms of analysis that focus on dimensions other than time and place is important for broadening the range of problems that crime analysis can reveal.

Performance focus: The performance focus should be carefully chosen, depending on the character of the problem being addressed. “Driving numbers down” is an appropriate focus only for crimes where discovery rates are high and the frequency of the crime is currently far greater than normal levels (i.e., there is slack to be taken up). It is not appropriate for the many types of invisible problems that need to be better exposed before they can be dealt with. And “reducing the numbers” will not be possible in perpetuity as crime rates will inevitably level off once they have reached reasonable levels. For jurisdictions that are relatively safe to begin with, reductions may not be possible, even in the short term.

Locus of responsibility: It makes sense to make precinct commanders unambiguously responsible for problems that are tightly concentrated within one precinct. But many issues are citywide or fit awkwardly into the precinct organization. A mature problem-oriented organization will use more fluid systems that allow for the formation of problem-solving

teams at many different levels of an organization to match the breadth and distribution of various problems.

Managerial style: In terms of managerial style, pressure to perform is one thing. But a modern police department has no place for tyrannical management, deliberate humiliation of officers in front of their peers, or attempts to catch them out with analytic findings not previously shared. Mature forms of CompStat should embody congenial and cooperative managerial relationships, even as they remain ruthlessly analytical and outcome oriented. Adversarial managerial styles exercised at high levels within a department tend to trickle all the way down, resulting in intolerable pressure on frontline officers and, ultimately, inappropriate forms of police action on the streets.

Preferred tactics: Aggressive zero-tolerance style policing is relevant only to specific classes of street crime and, as many commentators have observed, can destroy community relationships and cooperation. Persistent use of aggressive policing tactics, particularly in disadvantaged and minority neighborhoods, may be a recipe for anti-police riots in the end, given some appropriate spark. A mature CompStat system should bring no underlying preference for any particular set of tactics. Teams working on problems should be required and expected to consider the full range of interventions available to them and to invent new methods where necessary.

Table 2 offers a simple tool for diagnosing whether and in which particular ways an existing CompStat implementation might still be narrow or immature, highlighting opportunities to develop more sophisticated ideas about performance and more nuanced organizational responses to a broader range of issues across the six dimensions presented.

Conclusion

The principal purpose of this paper has been to highlight some of the narrower traditions into which police organizations fall when describing their value and reporting their performance. A modern police organization needs a broader view of its mission (per Goldstein),⁹³ a broader view of the dimensions of performance (per Moore),⁹⁴ and a clear understanding of the metrics that

Recommended Further Readings

- In *Measuring What Matters*, pages 43 to 47, Wesley Skogan provides a practical discussion of methods for assessing levels of disorder in neighborhoods.⁹⁵
- In *Measuring What Matters*, pages 47 to 50, Skogan also offers advice on measuring the fear of crime and its impact on community behavior and the use of public spaces.⁹⁶
- In *Measuring What Matters*, pages 58 to 59, Darrel Stephens offers ideas about measuring disorder and describes work by volunteer groups in St. Petersburg to survey residents and record physical conditions in neighborhoods.⁹⁷
- In *Measuring What Matters*, pages 202 to 204, Carl Klockars provides examples of simple-to-administer crime victim surveys, asking for feedback on the nature and quality of services delivered to them by police departments.⁹⁸
- In *Measuring What Matters*, pages 208 to 212, Klockars also stresses the importance of assessing levels of integrity within a police department (as a matter of organizational culture, rather than trying to measure actual levels of corruption) and describes some practical ways of doing what many might claim cannot be done.⁹⁹
- In *The “Bottom Line” of Policing*, pages 30 to 75,¹⁰⁰ in a section entitled “Measuring Performance on the Seven Dimensions,” Mark Moore and Anthony Braga provide a rich collection of practical ways to gather data, use existing data, and generate indicators relevant to each of the seven dimensions of performance. These are summarized in note form in table 3, pages 80 to 82,¹⁰¹ with suggestions for immediate investments prioritized in table 4, pages 83 to 85.¹⁰²
- Chapter 8 of *The Character of Harms* explores the special challenges that go with the class of invisible risks (those with low discovery rates). Much of that discussion revolves around issues of measurement, in particular the methods required to resolve the inherent ambiguity of readily available metrics.¹⁰³
- The 2012 Bureau of Justice Statistics Special Report, *Victimizations Not Reported to the Police, 2006-2010*, by Lynn Langton and colleagues, provides an interesting and data-rich analysis of the frequency with which various forms of victimization are not reported to the police.¹⁰⁴
- Chapter 12 of *The Character of Harms* provides a more rigorous analysis of performance-enhancing risks, describing the special threats to integrity that commonly arise when too much emphasis is placed on too narrow a set of quantitative metrics.¹⁰⁵

go with different types of work. Overall, police departments need more complex cockpits. Police executives need a more sophisticated understanding of how to use different types of information to understand the condition of their organizations and what is happening in the communities they serve.

Another purpose was to point police executives toward existing resources on this subject that they might find useful. For those interested in developing a more comprehensive suite of instruments for their “cockpit” and a clearer sense of how to use them all, I would particularly recommend the selections referenced in the “Recommended Further Readings” as relevant, practical and accessible.

In discussing this subject with police executives in the classroom, two questions invariably come up. First, someone will object, “So what’s new? Haven’t you just invented the balanced scorecard for public agencies?”

Not exactly. The cockpit I imagine is not really a “balanced scorecard”; it is hardly a “scorecard” at all. It is a richer information environment with much more sophisticated users. Police executives need a more comprehensive suite of managerial information but — like the airline pilot — should be slow to label any specific dial a performance indicator. As soon as you designate an indicator as a performance indicator, you need to think carefully about the currents that will run through the organization and make sure that you anticipate and can control perverse behaviors that may arise.

The second question that students raise is this: “So, to get the public and the politicians to pay attention to this broader, more complex performance story, should we withhold the traditional statistics (reported Part I crime figures) so we are not held hostage by them?”

My normal advice is basically this: “No, it’s a democracy, and transparency is the default setting. You cannot and would not want to hide that information. It remains important managerial information. Your job is not to withhold the traditional performance account but to dethrone it. You have to provide a richer story that better reflects the breadth of your mission and the contributions your agency makes. You have to provide that story even if the press, the politicians and the public do not seem to be asking for it just yet. Educate them about what matters, by giving it to them whether they ask for it or not. In the end, you can reshape their expectations.”

Endnotes

1. *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, edited by Robert H. Langworthy. Research Report. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, July 1999. NCJ 170610, <https://www.ncjrs.gov/pdffiles1/nij/170610.pdf>.
2. Moore, Mark H., with David Thacher, Andrea Dodge and Tobias Moore, *Recognizing Value in Policing: The Challenge of Measuring Police Performance*. Washington, D.C.: Police Executive Research Forum, 2002.

3. Moore, Mark H., and Anthony Braga, *The "Bottom Line" of Policing: What Citizens Should Value (and Measure!) in Police Performance*. Washington, D.C.: Police Executive Research Forum, 2003. http://www.policeforum.org/assets/docs/Free_Online_Documents/Police_Evaluation/the%20bottom%20line%20of%20policing%202003.pdf.

4. Skogan, Wesley G., "Measuring What Matters: Crime, Disorder, and Fear," in *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, edited by Robert H. Langworthy. Research Report. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, July 1999. NCJ 170610, <https://www.ncjrs.gov/pdffiles1/nij/170610.pdf>, pp. 36-53, at 38.

5. For a detailed study of the frequency with which various forms of victimization are not reported to the police, see Langton, Lynn, Marcus Berzofsky, Christopher Krebs and Hope Smiley-McDonald, *Victimizations Not Reported to the Police, 2006-2010*. Special Report. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics, August 2012. NCJ 238536, <http://www.bjs.gov/content/pub/pdf/vnrrp0610.pdf>.

6. For a detailed discussion of invisible harms and the special challenges involved in controlling them, see Chapter 8, "Invisible Harms," in Sparrow, Malcolm K., *The Character of Harms: Operational Challenges in Control*. Cambridge, England: Cambridge University Press, 2008.

7. Sparrow, *The Character of Harms*, pp. 143-144.

8. Adapted from Sparrow, Malcolm K., "Foreword," in *Financial Supervision in the 21st Century*, edited by A. Joanne Kellerman, Jakob de Haan, and Femke de Vries. New York, New York: Springer-Verlag, 2013. This framework was first described in Sparrow, Malcolm K., "The Sabotage of Harms: An Emerging Art Form for Public Managers," ESADE Institute of Public Governance & Management, E-newsletter, March 2012 (published in English, Spanish and Catalan).

9. The report, which is not public, was provided to the author by Assistant Commissioner Wilhelmy.

10. The advent of the six-axis or "full-flight" simulators for pilot training in the 1990s is regarded as one of the most significant recent contributors to enhanced safety in commercial aviation.

11. Klockars, Carl B., "Some Really Cheap Ways of Measuring What Really Matters," in *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, edited by Robert H. Langworthy. Research Report. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, July 1999. NCJ 170610, <https://www.ncjrs.gov/pdffiles1/nij/170610.pdf>, pp. 195-214, at 197.

12. See, e.g., Goldstein, Herman, *Policing a Free Society*. Cambridge, Massachusetts: Ballinger, 1977.

13. See, e.g., Moore and Braga, *The "Bottom Line" of Policing*; Moore et al., *Recognizing Value in Policing*.

14. See, e.g., Behn, Robert D., "Why Measure Performance? Different Purposes Require Different Measures," *Public Administration Review* 63 (5) (Sept./Oct. 2003): 586-606.
15. See, e.g., Sparrow, *The Character of Harms*.
16. Goldstein, *Policing a Free Society*, p. 35.
17. Behn, Robert D., "On the Core Drivers of: CitiStat." *Bob Behn's Public Management Report* 2 (3) (November 2004).
18. Moore, Mark H., and Margaret Poethig, "The Police as an Agency of Municipal Government: Implications for Measuring Police Effectiveness," in *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, edited by Robert H. Langworthy. Research Report. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, July 1999. NCJ 170610, <https://www.ncjrs.gov/pdffiles1/nij/170610.pdf>, pp. 151-167, at 153.
19. *Ibid.*, p. 157.
20. Moore et al., *Recognizing Value in Policing*, p. 76, table 2.
21. Bratton, William J., "Great Expectations: How Higher Expectations for Police Departments Can Lead to a Decrease in Crime," in *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, edited by Robert H. Langworthy. Research Report. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, July 1999. NCJ 170610, <https://www.ncjrs.gov/pdffiles1/nij/170610.pdf>, pp. 11-26, at 15.
22. Moore et al., *Recognizing Value in Policing*, pp. 24-25.
23. *Ibid.*, p. 25.
24. Kelling, George, "Measuring What Matters: A New Way of Thinking About Crime and Public Order," in *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, edited by Robert H. Langworthy. Research Report. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, July 1999. NCJ 170610, <https://www.ncjrs.gov/pdffiles1/nij/170610.pdf>, pp. 27-35, at 27.
25. Klockars, "Some Really Cheap Ways of Measuring What Really Matters," pp. 205-208.
26. *Ibid.*, p. 206.
27. Scheingold, Stuart A. "Constituent Expectations of the Police and Police Expectations of Constituents," in *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, edited by Robert H. Langworthy. Research Report. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, July 1999. NCJ 170610, <https://www.ncjrs.gov/pdffiles1/nij/170610.pdf>, pp. 183-192, at 186.
28. Skogan, Wesley G., *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*. New York, New York: The Free Press, 1990, p. 166.
29. *Ibid.*, p. 172.

30. Moore et al., *Recognizing Value in Policing*.

31. *Ibid.*, p. 79.

32. Behn, "Why Measure Performance?"

33. *Ibid.*, p. 588, table 1.

34. For a summary of this analysis, see *ibid.*, p. 593, table 2.

35. Sparrow, *The Character of Harms*, pp. 123-124. See Chapter 6, "Puzzles of Measurement," for a discussion of the unique characteristics of a risk-control performance account.

36. The formal structure and essential rigors of the problem-oriented approach are described in detail in *ibid.*, chapter 7. The associated performance measurement challenges are explored in chapters 5 and 6.

37. For a detailed discussion of the potential tensions between problem-oriented policing and "evidence-based" program evaluation techniques, see Sparrow, Malcolm K., *Governing Science*. New Perspectives in Policing Bulletin. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, 2011.

38. See, e.g., Hipolit, Melissa, "Former Police Officer Exposes Chesterfield's Ticket Quota Goals," CBS 6 News at Noon, July 14, 2014. <http://wtvr.com/2014/07/14/chesterfield-quota-investigation>.

39. "Whew! No More Traffic Ticket Quotas in Illinois," *CNBC News, Law Section*, June 16, 2014. <http://www.cnbc.com/id/101762636#>.

40. Scipione, Andrew P., "Foreword," in Eterno, John A., and Eli B. Silverman, *The Crime Numbers Game: Management by Manipulation*. Boca Raton, Florida: CRC Press, 2012, p. xviii.

41. Klockars, "Some Really Cheap Ways of Measuring What Really Matters," p. 196.

42. *Ibid.*

43. Moore et al., *Recognizing Value in Policing*, pp. 42-43.

44. *Ibid.*, p. 38.

45. Moore and Braga, *The "Bottom Line" of Policing*.

46. *Ibid.*, p. 37.

47. Langton et al., *Victimizations Not Reported to the Police, 2006-2010*, p. 5.

48. *Ibid.*

49. See note 6.

50. Eterno and Silverman, *The Crime Numbers Game*.

51. *Ibid.*, pp. 85-104.

52. Skogan, Wesley G., *Police and Community in Chicago: A Tale of Three Cities*. New York, New York: Oxford University Press, 2006. Chapter 3 describes crime-recording practices used by Chicago police during the 1980s.

53. Skogan, "Measuring What Matters: Crime, Disorder, and Fear," p. 38.

54. Eterno and Silverman, *The Crime Numbers Game*, pp. 11-12. See also the discussion of alleged manipulation of crime statistics by the NYPD, p. 23, under "Integrity Issues Related to CompStat and Crime Reporting at the NYPD."
55. Rawlinson, Kevin, "Police Crime Figures Being Manipulated, Admits Chief Inspector," *The Guardian*, Dec. 18, 2013. <http://www.theguardian.com/uk-news/2013/dec/18/police-crime-figures-manipulation-chief-inspector>.
56. *Ibid.*
57. Bernstein, David, and Noah Isackson, "The Truth About Chicago's Crime Rates," *Chicago Magazine*, Special Report, Part 1: Apr. 7, 2014. Part 2: May 19, 2014. <http://www.chicagomag.com/Chicago-Magazine/May-2014/Chicago-crime-rates>.
58. Hutcheon, Paul, "Independent Watchdog Refuses to Endorse Police Statistics That Claim Crime Is at a Near 40-Year Low," *Sunday Herald*, Aug. 17, 2014. <http://www.heraldscotland.com/news/home-news/independent-watchdog-refuses-to-endorse-police-statistics-that-claim-crime-is-at-a-ne.25062572>.
59. Eterno and Silverman, *The Crime Numbers Game*, pp. 9-12.
60. Campbell, Donald T., *Assessing the Impact of Planned Social Change*. Occasional Paper Series, #8. Hanover, New Hampshire: Dartmouth College, The Public Affairs Center, Dec. 1976, p. 49, cited in Eterno and Silverman, *The Crime Numbers Game*, p. 10.
61. Campbell, *Assessing the Impact of Planned Social Change*, pp. 50-51.
62. Vaughan, Diane, *Controlling Unlawful Organizational Behavior: Social Structure and Corporate Misconduct*. Chicago, Illinois: University of Chicago Press, 1983, p. 55, summarizing Merton, Robert K., "Social Structure and Anomie," in Robert K. Merton, *Social Theory and Social Structure*, enlarged ed., New York: New York: The Free Press, 1968, pp. 185-214.
63. For a detailed discussion of this phenomenon, see Chapter 12, "Performance-Enhancing Risks," in Sparrow, *The Character of Harms*.
64. Zink, Robert, "The Trouble with Compstat," *PBA Magazine*, Summer 2004, cited in Eterno and Silverman, *The Crime Numbers Game*, p. 27.
65. Part I Index crimes, as defined under the Uniform Crime Reporting system, are aggravated assault, forcible rape, murder, robbery, arson, burglary, larceny-theft and motor vehicle theft.
66. Bernstein and Isackson, "The Truth About Chicago's Crime Rates," Part 1.
67. Bratton, "Great Expectations," p. 15.
68. *Ibid.*, p. 17.
69. *Ibid.*, p. 17.
70. *Ibid.*, p. 11.
71. Moore et al., *Recognizing Value in Policing*, p. 150.

72. Eterno and Silverman, *The Crime Numbers Game*, pp. 28-34.

73. Rayman, Graham, "The NYPD Tapes: Inside Bed-Stuy's 81st Precinct," *The Village Voice*, May 4, 2010; Rayman, Graham, "The NYPD Tapes, Part 2: Ghost Town," *The Village Voice*, May 11, 2010; Rayman, Graham, "The NYPD Tapes, Part 3: A Detective Comes Forward About Downgraded Sexual Assaults," *The Village Voice*, June 8, 2010; Rayman, Graham, "The NYPD Tapes, Part 4: The Whistle Blower, Adrian Schoolcraft," *The Village Voice*, June 15, 2010; Rayman, Graham, "The NYPD Tapes, Part 5: The Corroboration," *The Village Voice*, Aug. 25, 2010; Rayman, Graham, "The NYPD Tapes Confirmed," *The Village Voice*, Mar. 7, 2012. The entire series of articles is available online at <http://www.villagevoice.com/specialReports/the-nypd-tapes-the-village-voices-series-on-adrian-schoolcraft-by-graham-rayman-4393217/>.

74. Rashbaum, William K., "Retired Officers Raise Questions on Crime Data," *The New York Times*, Feb. 6, 2010. http://www.nytimes.com/2010/02/07/nyregion/07crime.html?pagewanted=all&_r=0.

75. Baker, Al, and Ray Rivera, "5 Officers Face Charges in Fudging of Statistics," *The New York Times*, Oct. 15, 2010. <http://www.nytimes.com/2010/10/16/nyregion/16statistics.html>.

76. Baker, Al, and William K. Rashbaum, "New York City to Examine Reliability of Its Crime Reports," *The New York Times*, Jan. 5, 2011. <http://www.nytimes.com/2011/01/06/nyregion/06crime.html?pagewanted=all>.

77. Kelley, David N., and Sharon L. McCarthy, "The Report of the Crime Reporting Review Committee to Commissioner Raymond W. Kelly Concerning CompStat Auditing," Apr. 8, 2013. http://www.nyc.gov/html/nypd/downloads/pdf/public_information/crime_reporting_review_committee_final_report_2013.pdf.

78. *Ibid.*, p. 18.

79. *Ibid.*, p. 17, note 43.

80. *Ibid.*, p. 21.

81. *Ibid.*, p. 52.

82. *Ibid.*, pp. 41-42.

83. *Ibid.*, p. 52.

84. *Ibid.*, p. 53.

85. *Ibid.*, p. 58.

86. Skogan, "Measuring What Matters: Crime, Disorder, and Fear."

87. *Ibid.*, p. 38.

88. Kelley and McCarthy, "The Report of the Crime Reporting Review Committee to Commissioner Raymond W. Kelly Concerning CompStat Auditing," p. 54.

89. For a thorough analysis of this issue, see Chapter 4, "Distinguishing CompStat's Effects," in Behn, Robert D., *The PerformanceStat Potential: A Leadership Strategy for Producing Results*. Washington, D.C.: Brookings Institution

Press, 2014. Behn, who is a great admirer of *PerformanceStat* systems in general, surveys the literature on the subject, considers the 10 other explanations most commonly put forward as alternative explanations, and concludes that due to the limitations of the methods of social science, we will probably never know for sure what proportion of the crime reductions seen from 1994 onward was due to CompStat and what proportion is attributable to other factors.

90. Kelley and McCarthy, "The Report of the Crime Reporting Review Committee to Commissioner Raymond W. Kelly Concerning CompStat Auditing," p. 21.

91. See note 1.

92. See, e.g., Goldstein, Herman, "Improving Policing: A Problem-Oriented Approach," *Crime & Delinquency* 25 (April 1979): 236-258.

93. See, e.g., Goldstein, "Improving Policing: A Problem-Oriented Approach"; Goldstein, *Policing a Free Society*.

94. See, e.g., Moore and Braga, *The "Bottom Line" of Policing*; Moore et al., *Recognizing Value in Policing*.

95. Skogan, "Measuring What Matters: Crime, Disorder, and Fear," pp. 43-47.

96. *Ibid.*, pp. 47-50.

97. Stephens, Darrel W. "Measuring What Matters," in *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, edited by Robert H. Langworthy. Research Report. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, July 1999. NCJ 170610, <https://www.ncjrs.gov/pdffiles1/nij/170610.pdf>, pp. 55-64, at 58-59.

98. Klockars, "Some Really Cheap Ways of Measuring What Really Matters," pp. 202-204.

99. *Ibid.*, pp. 208-212.

100. Moore and Braga, *The "Bottom Line" of Policing*, pp. 30-75.

101. *Ibid.*, pp. 80-82, table 3.

102. *Ibid.*, pp. 83-85, table 4.

103. Sparrow, *The Character of Harms*, Chapter 8, "Invisible Risks."

104. Langton et al., *Victimizations Not Reported to the Police, 2006-2010*.

105. Sparrow, *The Character of Harms*, Chapter 12, "Performance Enhancing Risks."

Author Note

Malcolm K. Sparrow is professor of the practice of public management at the John F. Kennedy School of Government at Harvard University. This bulletin was written for the Executive Session on Policing and Public Safety at the school.

Findings and conclusions in this publication are those of the authors and do not necessarily reflect the official position or policies of the U.S. Department of Justice.



Official Business
Penalty for Private Use \$300



Members of the Executive Session on Policing and Public Safety

Commissioner Anthony Batts, Baltimore Police Department

Professor David Bayley, Distinguished Professor (Emeritus), School of Criminal Justice, State University of New York at Albany

Professor Anthony Braga, Senior Research Fellow, Program in Criminal Justice Policy and Management, John F. Kennedy School of Government, Harvard University; and Don M. Gottfredson Professor of Evidence-Based Criminology, School of Criminal Justice, Rutgers University

Chief Jane Castor, Tampa Police Department

Ms. Christine Cole (Facilitator), Executive Director, Program in Criminal Justice Policy and Management, John F. Kennedy School of Government, Harvard University

Commissioner Edward Davis, Boston Police Department (retired)

Chief Michael Davis, Director, Public Safety Division, Northeastern University

Mr. Ronald Davis, Director, Office of Community Oriented Policing Services, United States Department of Justice

Ms. Madeline deLone, Executive Director, The Innocence Project

Dr. Richard Dudley, Clinical and Forensic Psychiatrist

Chief Edward Flynn, Milwaukee Police Department

Colonel Rick Fuentes, Superintendent, New Jersey State Police

District Attorney George Gascón, San Francisco District Attorney's Office

Mr. Gil Kerlikowske, Director, Office of National Drug Control Policy

Professor John H. Laub, Distinguished University Professor, Department of Criminology and Criminal Justice, College of Behavioral and Social Sciences, University of Maryland, and former Director of the National Institute of Justice

Chief Susan Manheimer, San Mateo Police Department

Superintendent Garry McCarthy, Chicago Police Department

Professor Tracey Meares, Walton Hale Hamilton Professor of Law, Yale Law School

Dr. Bernard K. Melekian, Director, Office of Community Oriented Policing Services (retired), United States Department of Justice

Ms. Sue Rahr, Director, Washington State Criminal Justice Training Commission

Commissioner Charles Ramsey, Philadelphia Police Department

Professor Greg Ridgeway, Associate Professor of Criminology, University of Pennsylvania, and former Acting Director, National Institute of Justice

Professor David Sklansky, Yosef Osheawich Professor of Law, University of California, Berkeley, School of Law

Mr. Sean Smoot, Director and Chief Legal Counsel, Police Benevolent and Protective Association of Illinois

Professor Malcolm Sparrow, Professor of Practice of Public Management, John F. Kennedy School of Government, Harvard University

Mr. Darrel Stephens, Executive Director, Major Cities Chiefs Association

Mr. Christopher Stone, President, Open Society Foundations

Mr. Richard Van Houten, President, Fort Worth Police Officers Association

Lieutenant Paul M. Weber, Los Angeles Police Department

Professor David Weisburd, Walter E. Meyer Professor of Law and Criminal Justice, Faculty of Law, The Hebrew University; and Distinguished Professor, Department of Criminology, Law and Society, George Mason University

Dr. Chuck Wexler, Executive Director, Police Executive Research Forum

Learn more about the Executive Session at:

www.NIJ.gov, keywords "Executive Session Policing"
www.hks.harvard.edu, keywords "Executive Session Policing"