INTRODUCTION

Based on the real-life experiences of its creators, David Simon and Ed Burns, the HBO series *The Wire* is regarded by some as among the greatest cultural documents of our age. And with good reason. Focused on a single American city, Baltimore, the series drills down into a few major institutions—the police, the school system, municipal politics, the press—and provides an X-ray–like image of their workings and dysfunctions. The series has attracted an international audience because its themes of organizational dysfunction resonate broadly across Western societies.

One of the recurrent themes of *The Wire* is the salience of metrics: of measured performance as the hallmark of "accountability." Police commanders are obsessed with hitting the numbers—for example, cases solved, drug arrests, crime rates—and they do so by a variety of means that sacrifice effectiveness to meeting statistical targets. Politicians demand numbers that attest to police success in controlling crime. So the police units do their best to avoid having murders attributed to their district: when it turns out that a drug gang has been disposing of bodies in abandoned houses, the homicide sergeant discourages their discovery, since that would diminish the "clearance rate," the metric of the percentage of crimes solved. Much of the plot revolves around dedicated detectives seeking to develop a complex criminal case against a major drug lord. But since building that case will take months if not years, they are discouraged from doing so by the higher-ups, who want the cops to rack up favorable metrics by arresting lots of low-level drug dealers, despite the fact that those arrested will be replaced almost instantly. The mayor's office demands that the rate of major crimes decline by 5 percent before the end of the year, a target that can be reached only by overlooking actual crimes or downgrading their serious-

ness. In each case, they are engaged in "juking the stats"—improving their metrics either by distorting actual results, or by diverting their time and effort from crime prevention to less productive work.

Another plot line involves an ex-cop who teaches in a middle school in a neighborhood plagued by poverty, drug abuse, and family fragmentation. Students in the school perform poorly, and the school is in danger of being closed if the test scores of its students do not improve. So, in the six weeks before the standardized English reading and writing tests are to be administered, the teachers are instructed by their principal to focus all of class time on practicing for the tests, ignoring other subjects entirely (a strategy euphemistically referred to as "curriculum alignment"). "Teaching to the test," like juking the stats, is a way in which institutions are perverted, as effort is diverted from the institution's true purpose (education) to meeting the metric targets on which its survival has come to depend.

The distortive effects of performance metrics are felt at least as much across the Atlantic, in Great Britain.[1] There, another television series penned by a former real-life practitioner captures the same phenomenon. The series, *Bodies*, written by Jed Mercurio, a former hospital physician, takes place in the obstetrics and gynecology ward of a metropolitan hospital. In the first episode, a newly arrived senior surgeon performs an operation on a patient with complex comorbidities, after which she dies. His rival then provides him with this advice: "The superior surgeon uses his superior judgment to steer clear of any situation that might test his superior ability." That is, he avoids difficult cases as a way of maintaining his success rate. A classic strategy of "creaming," that is, avoiding risky instances that might have a negative impact on one's

measured performance. The cost of this tactic is that patients at greater risk for a failed surgery are left to an almost certain death without surgery.

Bodies is a medical drama, but the phenomena it depicts exist in the real world. Numerous studies have shown that when surgeons, for example, are rated or remunerated according to their success rates, some respond by refusing to operate on patients with more complex or critical conditions. Excluding the more difficult cases—those that involve the likelihood of poorer outcomes—improves the surgeons' success rates, and hence their metrics, their reputation, and their remuneration. That of course comes at the expense of the excluded patients, who pay with their lives. But those deaths do not show up in the metrics.

As we'll see, gaming the metrics occurs in every realm: in policing; in primary, secondary, and higher education; in medicine; in nonprofit organizations; and, of course, in business. And gaming is only one class of problems that inevitably arise when using performance metrics as the basis of reward or sanction. There are things that can be measured. There are things that are worth measuring. But what can be measured is not always what is worth measuring; what gets measured may have no relationship to what we really want to know. The costs of measuring may be greater than the benefits. The things that get measured may draw effort away from the things we really care about. And measurement may provide us with distorted knowledge—knowledge that seems solid but is actually deceptive.

We live in the age of measured accountability, of reward for measured performance, and belief in the virtues of publicizing those metrics through "transparency." But the identification

of accountability with metrics and with transparency is deceptive. Accountability ought to mean being held responsible for one's actions. But by a sort of linguistic sleight of hand, accountability has come to mean demonstrating success through standardized measurement, as if only that which can be counted really counts. Another assumption that is often taken for granted is that "accountability" demands that measurement of performance be made public, that is, "transparent."

The metric fixation is the seemingly irresistible pressure to measure performance, to publicize it, and to reward it, often in the face of evidence that this just doesn't work very well.

Used properly, measurement, as we'll see, can be a good thing. So can transparency. But they can also distort, divert, displace, distract, and discourage. While we are bound to live in an age of measurement, we live in an age of mismeasurement, over-measurement, misleading measurement, and counterproductive measurement. This book is not about the evils of measuring. It is about the unintended negative consequences of trying to substitute standardized measures of performance for personal judgment based on experience. The problem is not measurement, but excessive measurement and inappropriate measurement—not metrics, but metric fixation.

We are often told that gathering metrics of measured performance and then making them available to the public is a way to improve the functioning of our institutions. Nowhere have the virtues of accountability, performance metrics, and transparency been more touted than in the field of medicine. And understandably so, for nowhere are the stakes higher. The health sector not only makes up over 17 percent of the U.S. economy, but lives are also on the line. Surely, the logic goes, measures of performance can help save dollars and save lives.

Gathering standardized information about the success rates of surgeons, or the survival rate of patients admitted to particular hospitals, is supposed to be helpful. For if doctors or hospitals are remunerated by government agencies or private insurers based on their success rates in keeping patients alive, then such measurements should create incentives for better care. And if the success rates of doctors and hospitals are publicized, the resulting transparency will allow the public to choose among doctors and among hospitals. All in all, metrics, accountability, and transparency will provide the cure for what ails the medical professions. What could go wrong?

A good deal, as we have already seen. When their scores are used as a basis of reward and punishment, surgeons, as do others under such scrutiny, engage in creaming, that is, they avoid the riskier cases. When hospitals are penalized based on the percentage of patients who fail to survive for thirty days beyond surgery, patients are sometimes kept alive for thirty-one days, so that their mortality is not reflected in the hospital's metrics.[2] In England, in an attempt to reduce wait times in emergency wards, the Department of Health adopted a policy that penalized hospitals with wait times longer than four hours. The program succeeded—at least on the surface. In fact, some hospitals responded by keeping incoming patients in queues of ambulances, beyond the doors of the hospital, until the staff was confident that the patient could be seen within the allotted four hours of being admitted.[3]

We'll explore these issues in the realm of medicine in greater depth. But what is striking is that the problems that arise in healthcare arise in many other institutions—in K-12 and college education; in policing and other public services; in business and finance; and in charitable organizations. Those who work in any of these fields will have some sense

of such problems in their institutions. And social scientists have examined and anatomized them in one or another of these realms. What has gone largely unnoticed is the recurrence of the same unintended negative consequences of performance metrics, accountability, and transparency across a wide range of institutions.[4]

As with many insights, once you've become aware of metric fixation, you are likely to find it almost everywhere—and not just in television dramas.

The catchwords of metric fixation are all around us. Google's Ngram—which instantly searches through thousands of scanned books and other publications—provides a rough but telling portrait of changes in our culture and society. Set the parameters by years, type in a term or phrase, and up pops a graph showing the incidence of the words from 1800 to the present. Type in "accountability" and you will see a line that begins to curve upward around 1965, with an increasingly rising slope after 1985. So too with "metrics," which begins its steep increase around 1985. "Benchmarks" follows the same pattern, as does "performance indicators."

This book argues that while they are a potentially valuable tool, the virtues of accountability metrics have been oversold, and their costs are often underappreciated. It offers an etiology and diagnosis, but also a prognosis for how metric fixation can be avoided, and its pains alleviated.

The most characteristic feature of metric fixation is the aspiration to replace judgment based on experience with standardized measurement. For judgment is understood as personal, subjective, and self-interested. Metrics, by contrast, are supposed to provide information that is hard and objective. The strategy is to improve institutional efficiency by offering re-

wards to those whose metrics are highest, or whose benchmarks or targets have been reached, and to penalize those who fall behind. Policies based on these assumptions have been on the march for several decades, and as the ever-rising slope of the Ngram graphs indicate, their assumed truth goes marching on.

To be sure, there are many situations where decision-making based on standardized measurement is superior to judgment based upon personal experience and expertise. Decisions based on big data are useful when the experience of any single practitioner is likely to be too limited to develop an intuitive feel for or reliable measure of efficacy. When a physician confronts the symptoms of a rare disorder, for example, she is better advised to rely on standardized criteria based on the aggregation of many cases. Checklists—standardized procedures for how to proceed under routine conditions—have been shown to be valuable in fields as varied as airlines and medicine.[5] And, as recounted in the book *Moneyball*, statistical analysis can sometimes discover that clearly measureable but neglected characteristics are more significant than is recognized by intuitive understanding based on accumulated experience.[6]

Used judiciously, then, measurement of the previously unmeasured can provide real benefits. The attempt to measure performance—while pocked with pitfalls, as we will see—is intrinsically desirable. If what is *actually* measured is a reasonable proxy for what is *intended* to be measured, and if it is combined with judgment, then measurement can help practitioners to assess their own performance, both for individuals and for organizations. But problems arise when such measures become the criteria used to reward and punish—when metrics become the basis of pay-for-performance or ratings.

Schemes of measured performance are deceptively attractive because they often "prove" themselves by spotting the most egregious cases of error or neglect, but are then applied to all cases. Tools appropriate for discovering real misconduct become tools for measuring all performance. The initial findings of performance measurement may lead poor performers to improve, or to drop out of the market. But in many cases, the extension of standardized measurement may be of diminishing utility, or even counterproductive—sliding from sensible solutions to metric madness. Above all, measurement may become counterproductive when it tries to measure the unmeasurable and quantify the unquantifiable.

Concrete interests of power, money, and status are at stake. Metric fixation leads to a diversion of resources away from frontline producers toward managers, administrators, and those who gather and manipulate data.

When metrics are used by managers as a tool to control professionals, it often creates a tension between the managers who seek to measure and reward performance, and the ethos of the professionals (doctors, nurses, policemen, teachers, professors, etc.). The professional ethos is based on mastery of a body of specialized knowledge acquired through an extended process of education and training; autonomy and control over work; an identification with one's professional group and a sense of responsibility toward colleagues; a high valuation of intrinsic rewards; and a commitment to the interests of clients above considerations of cost.[7]

That tension is sometimes necessary and desirable, for the professional ethos tends to discount issues of cost and opportunity cost. That is, the professional is inclined to see only the advantages of providing more of his or her services, without much attention to the limits of resources, or their alter-

nate uses. Professionals don't like to think about costs. Metrics folks do. When the two groups work together, the result can be greater satisfaction for both. When they are pitted against one another, the result is conflict and declining morale.

While there are vested interests at stake that sometimes lead from reasonable metrics to metric madness, the cause lies as much in the uncritical adoption of metric ideology. Like every culture, the culture of metric accountability has its own unquestioned sacred terms and its characteristic blind spots.[8] Yet today it is so dominant that its flaws tend to go unnoticed.

You might wonder how a historian came to write a book about the tyranny of metrics. It happened as I came to recognize that troubling developments in my own professional experience were reflections of much larger patterns in our society. Microlevel discontents led to macrolevel analysis, as I came to understand that cultural patterns that were damaging my narrow professional turf were warping many contemporary institutions.

I was drawn into the subject through my experience as the chair of my department at a private university. There are many facets to such a job: mentoring faculty members to help them develop as scholars and teachers; hiring new faculty; trying to ensure that necessary courses get taught; maintaining relations with deans and others in the university administration. Those responsibilities were on top of my roles as a faculty member: teaching, researching, and keeping up with my professional fields. With all those roles, I was quite satisfied. Time devoted to thinking about and working with faculty members contributed to making them better teachers and scholars. I was proud of the range and quality of the courses that we were teaching,

and relations with other departments were fine. Teaching, researching, and writing were demanding, but satisfying.

Then, things began to change. Like all colleges and universities, our institution gets evaluated every decade by an accrediting body, the Middle States Commission on Higher Education. It issued a report that included demands for more metrics on which to base future "assessment"—a buzzword in higher education that usually means more measurement of performance. Soon, I found my time increasingly devoted to answering queries for more and more statistical information about the activities of the department, which diverted my time from tasks such as research, teaching, and mentoring faculty. There were new scales for evaluating the achievements of our graduating majors—scales that added no useful insights to our previous measuring instrument, namely grades. I worked out a way of doing this speedily, without taking up much time of the faculty, simply by translating the grades the faculty had awarded into the four-category scale created for purposes of assessment. Over time, gathering and processing the information, in turn, required the university to hire ever more data specialists. (It has since gone so far as to appoint a vice-president for assessment.) Some of their reports were genuinely useful: for example, in producing spreadsheets that showed the average grade awarded in each course. But much of the information was of no real use, and indeed, was read by no one. Yet once the culture of performance documentation caught on, department chairs found themselves in a sort of data arms race. I led the department through a required year-long departmental self-assessment—a useful exercise, as it turned out. But before sending it up the bureaucratic chain, I was urged to add more statistical appendices—because if I didn't, the report would look less rigorous than that of other

departments. One fellow chair—a solid senior scholar—devoted most of one summer to compiling a binder full of data, complete with colored charts, to try to convince the dean of the need to fill a faculty slot in his department.

My experience was irritating, not shattering: a pin-prick not a blow. But it stimulated me to inquire more deeply into the forces leading to this wasteful diversion of time and effort. The Middle States Commission, from which the stimulus for more data originated, operates with a mandate from the U.S. Department of Education. That department, under the leadership of Margaret Spellings, had convened a Commission on the Future of Higher Education, which published its report in 2006 emphasizing the need for greater accountability and the gathering of more data, and directing the regional accrediting agencies to make "performance outcomes" the core of their assessment.[9] That mode of evaluation, in turn, filtered down to the Middle States Commission, and from there to the administration of my university, and eventually down to me. Spellings had been the director of the Domestic Policy Council under President George W. Bush at the time of the passage of the No Child Left Behind Act in 2001. At first, I had thought that legislation—which expanded the evaluation of teachers and schools based on the scores of their students on standardized tests—was a positive step. But in time I came to hear searing critiques of it by erstwhile supporters, such as the former assistant secretary of education, Diane Ravitch. And classroom teachers of my acquaintance told me that while they loved teaching, they found that the increasing regimentation of the curriculum, intended to maximize performance on the tests, was sucking away their enthusiasm.

Such accounts led me to investigate, using my own intellectual toolkit, the broader historical and cultural roots and

contemporary manifestations of the culture of measured and rewarded performance that is permeating ever more institutions. My professional interests had been on the borders between history, economics, sociology, and politics. I had long been interested in the history of what we have come to call "public policy," and had published a book on Adam Smith as a public policy analyst. I had also written about the history of conservative approaches to public policy, and some of the thinkers I had written about, such as Michael Oakeshott and Friedrich Hayek, turned out to provide critical insights into our contemporary apotheosis of measured performance. I had been interested in the history of capitalism, especially the ways in which intellectuals have thought about the social, moral, and political prerequisites and ramifications of business. A recurrent concern among modern Western intellectuals about whom I had written was the potentially pernicious spillover effects of concepts and predispositions from business and from the discipline of economics into other realms of life. And so, my personal experience of professional discontent proved serendipitous, stimulating me to investigations that drew upon a wide range of my interests. The spirits presiding over this book are those of Matthew Arnold, the great Victorian cultural critic, and of my teacher, Robert K. Merton, who schooled me to look out for the unanticipated and unintended consequences of social action—and for serendipity in scholarship.[10]

As I began to investigate these issues, a book by a sociologist at the Harvard Business School, Rakesh Khurana's *From Higher Aims to Hired Hands: The Social Transformation of American Business Schools and the Unfulfilled Promise of Management as a Profession*, opened my eyes to the intellectual history of business schools themselves, and the broader impact of what

gets taught in them. These insights led me to wider investigations of the changing culture and ideologies in the field of management, the sometimes dubious nature of which is nicely captured in the title of Adrian Wooldridge's book, *The Witch Doctors* (a second edition carries the more benign title, *Masters of Management*).

I proceeded to consult a wide range of scholarly literatures, in fields from economics and politics, to history, anthropology, psychology, sociology, public administration, and organizational behavior. I made extensive use of social scientific studies of the actual behavior of teachers, professors, doctors, and policemen in the real world.

In surveying the scholarship on the topic from a variety of fields, I was struck by the degree to which academic disciplines tend to be walled off from one another, and by the gap between academic research and real world practice. I found remarkable, for example, how much of recent economic literature on incentives and motivation was a formalization of what psychologists had already discovered. But much of what psychologists had discovered was long known by managers with judgment. Yet although there is a large body of scholarship in the fields of psychology and economics that call into question the premises and effectiveness of pay for measured performance, that literature seems to have done little to halt the spread of metric fixation.[11]

That is why I wrote this book. Little of what this book has to say is entirely new—it is based on synthesizing research and insights drawn from many other authors. Many of the dysfunctions connected with what I've termed "metric fixation" have been documented and analyzed by scholars writing about one or another domain: education, medicine, policing, profit-oriented enterprises, and nonprofits. A few students of

organizational behavior, writing in rather specialized venues, have analyzed some of the broader patterns of success and dysfunction. What no one has really done is put it all together and make it accessible to all of us who guide and work in these institutions, from politicians deciding on the fate of educational and medical systems, to members of boards of directors of corporations, to trustees of universities and nonprofit organizations, and down to the peons (such as department chairs). This book is for them. More broadly, it's for anyone who wants to understand one of the big reasons why so many contemporary organizations function less well than they ought to, diminishing productivity while frustrating those who work in them.

Though the thrust of the argument rubs against the received wisdom of many contemporary institutions, I've aimed not at novelty but at distilled wisdom. Readers eager to pigeonhole the argument into some existing ideological framework will be disappointed, as it draws not only from a variety of disciplines but from a variety of political orientations. I have drawn upon evidence and insight from wherever they were to be found. I hope that readers will approach the book with the same open mind.