

Chapter 18

Performance Measurement in Public Administration*

Success can be a great liar. —Friedrich Nietzsche

INTRODUCTION

“Performance” is one of those seductive terms whose meaning appears self-evident but is not. The introduction of performance indicators into public management has gathered steam recently, carrying both a potential for greater effectiveness and substantial risk, especially in developing countries. It is thus necessary to unbundle the concept of performance, and review the country- and sector-specific conditions that make for reform success, failure, or disaster. This chapter begins by stressing that performance is inherently relative and culture-specific—not as justification for inaction, but as prerequisite for moving to a more efficient set of performance rules. A recapitulation of performance indicators then suggests an “accountability tradeoff,” by which accountability can be either tight or broad but not both, and the resulting need to use a combination of indicators. Next, we list ten caveats and risks, from actual international experience. The major risks stem from proceeding as if performance indicators were easy to define, implement, and monitor; neglecting costs, including transaction costs; and disregarding due process in the name of results, which leads in time to bad process and bad results. The chapter goes on to outline steps for a realistic and sustainable strengthening of performance orientation in the public sector, and concludes with the customary synthesis of key points and directions for reform.

In recent years, several developed countries and some developing countries have made increasing use of performance concepts and results

* An earlier version of this chapter was published in the December 1999 issue of the *Asian Journal of Political Science*. Adapted and reprinted by permission.

indicators, both in their managerial practices and in the formulation and execution of public programs, assisted—and often pushed—by external aid agencies. Occasionally (e.g., in Singapore), the results have justified the substantial investment in time, effort, and stress; typically, they have not. The key determinant of success or failure—as argued throughout this book—is whether the changes were realistic, introduced gradually, and consistent with both the methodological complexity of the topic and the specific country realities (especially administrative capacity and the governance regime).

Several factors have led to the recent focus on performance. The main ones are the pervasive dissatisfaction with government employees' unresponsiveness to the public; the dynamics of Wagner's Law, which systematically increases the relative size of government, and hence puts pressure on the public finances;¹ and the New Public Management (NPM) paradigm. As previewed in chapter 1, the genesis of the NPM can be dated to the early 1980s (essentially starting from the Thatcher reforms in the United Kingdom), and its heyday was marked by the completion in the early 1990s of the public-sector revolution implemented in New Zealand.²

Clearly, to forget the real purpose of spending monies obtained from the people eventually generates a culture of means rather than ends, disregard for the public, and the legendary bureaucratic mentality that considers it a success to formulate tight and internally consistent controls—regardless of whether they are necessary or even helpful in executing the functions assigned to the government. Thus, a focus on policy and performance is highly appropriate—*provided* that it does not lead to forgetting the importance of integrity and of due process. (This is a real risk, which has already materialized in some countries, with serious costs in terms of corruption and government credibility).

But what *is* performance? The reforms that have introduced performance-based management mechanisms (usually assisted by aid in developing countries) have only rarely been mindful of the country context and circumstances (including administrative capacity). In some countries, however, they have been introduced mechanically, without any regard for the need to adapt to local circumstances or even correctly identify the real problems. Not surprisingly, the result has been to waste time and resources and create unnecessary new problems without solving the existing ones.

The first basic requirement is to be clear about the terms and the complexity of the performance issue. The intent is not in any way to discourage attempts to inject a stronger performance orientation. Quite the contrary: to present up-front the full complexity of the performance issue helps lead to clear-eyed adoption of those performance-oriented reforms that have a good chance to last and be effective. Instead, experience shows that it is the introduction of performance-based systems as if they were very easy, simple, and self-evident that is likely to be unsustainable and destroys the credibility of the concept itself.

THE MEANING OF “PERFORMANCE”

Performance: In Terms of What?

Dictionary definitions of “performance” include such alternative terms as “accomplishment,” “achievement,” “realization,” and “fulfillment.” Most of these terms have to do with the *objective* effect of public actions, but some relate to the *subjective* sense of satisfaction experienced as a result of one’s action. Naturally, the economic and public management literature emphasizes the former meaning, not only because of its direct implications for the population, but because subjective satisfaction is extremely difficult to measure and impossible to aggregate.

Accordingly, performance may be defined in terms of *effort* or in terms of *results*. To pay attention to individual effort is often looked at as a “soft and gentle” approach to salve human feelings. It is that. But it is also an eminently practical proposition. Consider what happens if you completely neglect the subjective dimension of performance and focus *only* on objective results. The brighter though lazier persons will be rewarded for their better results, and the less capable but harder workers will be penalized. The former will therefore receive the clear message that underachieving carries no negative consequences; the latter will get the equally clear message that working hard carries no rewards. Both groups being composed of rational individuals, the level of effort will decline across the board and, in time, the entire organization will be populated by underachievers.

Recognizing (even if not rewarding) genuine individual effort can do much for morale and also serve as a demonstrator for others, thus fostering the effectiveness of the organizational unit. More fundamentally, most human beings consider a sense of accomplishment (what Thorstein Veblen called the “instinct of workmanship”) as a strong motivator of their action—

independent of salaries, penalties, or other material incentives. Thus, if public reforms inadvertently remove that motivation, the efficiency of personnel is likely to decline, and the effectiveness of public action along with it.³ The normal human drive to do things right should be harnessed, not disregarded or depreciated. (This is certainly recognized in the more efficient private corporations.) Nonetheless, while we should keep these factors in mind, in order to introduce stronger performance orientation *it is advisable to rely mainly on results, among many reasons, because effort is less easily measurable and is an excellent alibi for lack of results.*

In any event, it is critical to realize that *performance is a relative and culture-specific concept.* Government employees are considered well-performing if they stick to the letter of the rules, in a system where rule compliance is the dominant goal; if they account for every cent of public funds, in a system where rule compliance is the dominant goal; if they obey without question their superiors' instructions, in a strictly hierarchical system; if they compete vigorously for individual influence and resources, in a system where such competition is viewed positively; if they cooperate harmoniously for group influence and cohesion, in a system where conflict is discouraged; and so on.

Must we then infer that all the diverse administrative cultures are equally efficient? Certainly not. Indeed, the objective of institutional reform in public administration is to move from a less efficient to a more efficient set of behavioral rules. But we must recognize that administrative cultures do not come from Mars. They evolve in response to man-made incentive structures and concrete problems. Even when an administrative culture has become obsolete or dysfunctional it is *still* necessary to understand its institutional roots if one wishes to help improve it in a durable way. For example, the practice of advancement by seniority has come under fire as preventing the recognition of individual merit. This is generally true but, as noted in Chapter 10, it must not be forgotten that the seniority principle was originally introduced in public administration largely *as a reform* to insulate the system from the vagaries of political pressures on government employees. Correspondingly, depending on the quality of governance in a country, a change to a "merit-based" system may carry the risk of reopening the door to such pressures. The change may still be desirable, but the reform should also address that risk.

It is an unfortunate reality that many aid-assisted public administration reform programs never took the trouble to assess how and from where today's

problems arose in the first place; it is an unfortunate corollary of that reality that these programs have produced no lasting improvement.

What Sort of “Results”?

The measurement of “objective” performance rests on inputs or on one or more of the following results, or both.⁴ Let’s recapitulate, using the example of law enforcement:

- *Inputs* are the resources used to produce the service—e.g., policemen, prisons, police cars, handcuffs. The social value of inputs is measured by their cost. The performance criterion corresponding to inputs is *economy*, i.e., the timely acquisition of good-quality inputs at lowest cost.
- *Output* is the service itself—e.g., number of arrests. The social value of outputs is approximated by the market price for the same or the closest equivalent service (or, in its absence, by total unit cost). The performance criterion corresponding to outputs is *efficiency*, i.e., minimizing total input cost per unit of output (or maximizing the quantity of output in relation to a given total cost of inputs).
- *Outcome* is the purpose that is achieved by producing the service—e.g., reduction in crime. The social value of outcomes is difficult to assess, except as revealed by public reaction in the political arena. The performance criterion corresponding to outcomes is *effectiveness*, i.e., maximizing outcomes in relation to the outputs produced.⁵
- *Process* is the *manner* in which inputs are procured, outputs produced, or outcomes achieved. The value of good process is high but undetermined. For inputs, good process consists of intelligent compliance with input acquisition and utilization rules and, of course, integrity. In some areas of public activity, as in our example of law enforcement, “due process” has its own independent validity and is a key element of good governance. An increase in arrests achieved by violating civil rights would surely not constitute “good performance.” In other areas, process indicators are a useful proxy for performance when outputs or (more often) outcomes cannot be defined with clarity (e.g., “bedside manner” in health services, “rules for free debate” in policy formulation). Process indicators can be quantitative (e.g., percentage of class time dedicated to student questions) but are usually

qualitative. Even then, they can frequently be transformed into quantitative indicators by feedback from users: e.g., hospital patient satisfaction can be numerically assessed through a patient survey.⁶

Table 18.1 below gives examples of input, output, outcome, and process indicators in various sectors. Some are good indicators; others are bad and likely to reduce or distort performance rather than improve it. The reader should decide which is which and imagine the likely consequences of using one or another indicator. In any case, it is best to avoid relying on any single indicator: a fuller understanding of performance can be gained by using a *combination* of indicators. This means that performance should be evaluated through a dialogue, and not by mechanistic rankings of individual performance indicators.

Table 18.1
Examples of Performance Indicators

Sector	Input	Output	Outcome	Process
<i>General Administration</i>	Number of staff	Number of policy papers	Better decisions	Openness of debate
<i>Education</i>	Student-teacher ratio	Retention rates	Higher literacy	Encouragement of student expression
<i>Judicial system</i>	Budget	Cases heard	Low appeal rate	Assistance for indigent defendants
<i>Police</i>	Number of police cars	Number of arrests	Decline in crime rate	Respect for rights
<i>Corrections</i>	Cost/prisoner	Number of prisoners	Recidivism rate	Prevention of abuse
<i>Health</i>	Nurse-population ratio	Number of vaccinations	Lower morbidity	Bedside manner
<i>Social welfare</i>	Number of social workers	Number of persons assisted	Exits from the system	Dignified treatment

The Link to Accountability

The whole point of measuring results is to improve performance, through the intermediate process of making individuals more accountable for the results of their actions. The hierarchy of results given above suggests a sort of complex production function of public services, whereby the *outcome* of one stage is an *output* of the next stage. Accordingly, in “downstream” activities close to the ultimate user (e.g., urban transport) the output-outcome link is clear and immediate enough to permit using output indicators as a good proxy for outcomes. In “upstream” activities this is not so (e.g., in rule making, where “maximizing” public rules is hardly a desirable measure of public performance).

The above implies an “accountability chain”—with accountability clearest and most immediate by the narrowest performance criterion (i.e., compliance with input allocations), and most ambiguous and diffuse by the broadest performance criterion (i.e., net impact).⁷ For example, it is fairly easy to hold a village nurse strictly accountable for the output of vaccinations, and to reward or penalize him accordingly; it is difficult to hold him responsible for the outcome of improving the health of village children. Yet, his affirmative involvement in household sanitary conditions or nutrition or other health factors may have more influence on the outcome of improving children’s health than a greater number of vaccinations—but *such involvement will not be motivated by an incentive system that focuses only on the outputs.*

Moreover, in the absence of close supervision, it is difficult to prevent immunizations from being performed with less than the recommended quantity of vaccine (the remaining vaccine “leaking” out of the health delivery system). Therefore, abandoning input and quality controls in favor of output indicators may carry substantial practical risks. (Easy “insurance” can be obtained by retaining input and quality controls alongside the output indicators, for as long as may be needed to shake the bugs out of the results-oriented system.)

These considerations are not meant to suggest that outcome indicators are “better” than output indicators, or vice versa. Other things being equal, output indicators are closer to the desired outcomes, and hence more realistic, the closer the activity is to the final user. However, the greater specificity associated with output indicators comes with a loss of relevance, and conversely, it would be difficult to hold public servants strictly

accountable for outcomes; the measurement problem becomes more complex as one proceeds along the scale from input measures through outputs, outcomes, and finally process indicators. Although the quality issue is ever present, there is no great methodological difficulty in defining and measuring outputs (and even less so, inputs): *the issue with output indicators is their relevance*. Similarly, the interpretation of outcomes is rarely in doubt: *the issue with outcome indicators is their feasibility* as a spur to better performance. Outcome indicators are almost always more meaningful, and output indicators almost always more feasible. Combining these two considerations, performance measurement is most appropriate for those government activities where there is a direct and immediate relationship between the government agency's outputs and the desired public outcomes.

The selection of output or outcome indicators (in cases where the introduction of noninput-based performance indicators is appropriate to begin with) is also heavily influenced by data availability and information technology. First, good data and good monitoring permit better definition of outputs and thus justify greater reliance on them as a measure of performance. Conversely, when data are lacking or unreliable (or monitoring is weak), measuring performance by outputs generates only game playing and self-delusion. In such cases, the priority must be to strengthen compliance and responsibility for input use, and to improve the relevant data and monitoring capability—before even considering the introduction of results-based performance elements.

Moreover, data collection costs and, more generally, the transaction costs of introducing performance indicators in a systematic manner can be enormous. These costs must be assessed realistically and weighed against the benefits expected. It is simply wrong to limit a debate on whether to introduce performance indicators only to the benefits expected from doing so. Yet, this is often the case, and performance-based systems have been introduced on the basis of a reasonable expectation that they would improve performance—*but at what cost?*

Finally, as noted earlier, the appropriate choice of performance indicators differs for different countries, times, and sectors. The only valid *general* rule therefore is: *when results measurement is appropriate and cost-effective, performance should be assessed according to that combination of output, outcome, and process indicators that is realistic and suitable for the specific activity, sector, country, and time.*

TEN RULES FOR INTRODUCING PERFORMANCE REFORMS

The above discussion underlined the care, common sense, and direct sectoral knowledge needed to introduce performance indicators successfully. Indeed, the careless introduction of performance indicators has often generated unintended consequences so serious as to provoke a general backlash against all performance measurement. (This explains the apparent paradox of why some officials in developing countries who stand to lose from a new and robust focus on performance generally support the introduction of simplistic performance indicators.) This section offers a variety of analogies, metaphors, and anecdotes to illustrate the issues. The intent is not to add analytical content, but to provide “memory aids” to anchor the issues, and incidentally to take some of the dryness out of the topic of public-sector performance.

The Patton Premise (*Know Where You’re Going Before You Get Moving*)

In the 1970 movie *Patton*, the actor playing General George Patton stumbles on a sleeping soldier while inspecting the sorry state of the US Army after the disastrous defeat in their first battle against the Germans at Kasserine Pass in Tunisia in 1942. The soldier says: “Hey! I am trying to sleep!” then, realizing it’s the new commanding general, mumbles some apology. Patton replies: “Don’t worry, son, you’re the only one in this army who knows what he’s trying to do.”

The first, and most obvious, requirement for strengthening performance is to be crystal clear about the objective of the activity being performed. Yet, often because of the force of fashion, in many countries performance indicators were introduced without defined goals, and in some countries weakened control and accountability systems that had been working reasonably well before.

The Stepsisters’ Predicament (*If the Shoe Doesn’t Fit, Get Another*)

In the original version of the Cinderella story, one of her stepsisters cut off her toe to fit in the glass shoe; the other cut off part of her heel. Both ended up with mutilated feet and—of course—got neither the shoe nor the Prince.

We have stressed throughout the book that all institutional innovations must be viewed in the light of the local cultural, social, and

historical context and—above all—administrative capacity. In some countries, instead of carefully designing the reform to fit local conditions, the approach to improving performance has ignored administrative capacity limitations and other institutional constraints and then—when the approach failed—blamed the very same capacity limitations for the failure. But *administrative capacity, too, is relative—relative to the complexity and ambition of the reform* envisaged. Failing to assess realistically the local context, and designing the reforms without reference to it, will invariably produce a more severe capacity constraint than had existed at the start of the process. To then use this problem as an excuse for the design failure adds, literally, insult to injury.

Regrettably, this is sometimes deliberate. The local capacity constraint engendered by overly complex reforms recommended by external consultants is then used to justify the continued need for the consultants themselves. The eventual outcome is not only a relatively more severe local capacity constraint, but a greater dependence on external advisers—which is the best single tip-off of the failure of the initial institutional innovations.

The Accountability Trade-off (*There's No Free Lunch Here Either*)

In statistical inference, there is an inverse correlation between precision and probability (given the size of the sample). A “point estimate” is highly precise but carries near-zero probability of being right, i.e., corresponding to the true value of the variable. Conversely, a very wide band estimate is highly likely to comprise the true value but is too broad to be useful. Similarly, in the domain of performance there is an accountability trade-off: *accountability can be broad or tight, but not both*. Tight and immediate accountability is by definition narrow accountability; conversely, the link between action and results becomes more ambiguous the broader and the more meaningful the results.

We can either measure very accurately the performance of specific things (and then we can hold those in charge strictly accountable for those specific things), or resign ourselves to getting a rough idea of the important results (and then tolerate the resulting vagueness in attributing responsibility). In practice, it is advisable to do both, in order to get a more rounded sense of the overall results. It follows from the accountability trade-off that performance monitoring through outputs is least appropriate for complex tasks (e.g., mental health services) but can be very effective for simple processes (e.g., trash collection).⁸

The Titanic Warning (*It's What You Can't See That Can Sink You*)

The great ship *Titanic*, which was considered unsinkable, sank in its maiden voyage in 1912 after its below-water compartments struck an iceberg in the North Atlantic. There are two lessons here. The first is that it is not sensible ever to believe that any particular institutional reform is “unsinkable” and bound to succeed. The second lesson, of course, is that the *Titanic* was sunk by the unexposed portion of the iceberg.

The total stock of institutional rules in any society is always much greater than the portion visible as formal rules. Indeed, sometimes the visible formal rules are simply not operative. For example, the Soviet Constitution of 1936 was considered as a model document, protecting individual rights at the same time as it facilitated the achievement of social objectives. In reality, anyone who was foolish enough to act as if this were really so quickly found himself in serious trouble. In reforms intended to encourage stronger performance orientation in government, a design failure to recognize and take into account the key informal rules (which are generally below the surface) is likely to lead to a failure of the reform itself. This does not mean that reforms or external donor agencies are supposed to acquire their own expertise in the inner workings of society, because this would be presumptuous and unrealistic. It does mean that they have a responsibility to identify those who do know and understand the local informal rules, and to get them to participate in the design of the institutional reform—or at least to “kick the tires” before the reform is introduced.

The Heisenberg Dilemma (*Beware the Law of Unintended Consequences*)

In physics, the well-known Heisenberg uncertainty principle states that the actions required to observe a phenomenon themselves alter it. In the context of performance measurement, it is not advisable to be too sure that the actions undertaken will have the effects intended, and *only* the effects intended. Introducing new ways of evaluating the results of human action always leads to changes in behavior (*if* the evaluation is attached to concrete changes in incentives). Of course, it is precisely a change in public servants' behavior—toward the ends rather than the means—that performance indicators are intended to generate. However, attempts to modify behavior often produce unintended behavior that may be at odds with the objective or even defeat it altogether and worsen the initial situation.

Some examples, all from real life, follow:

- When police performance is measured by the number of police officers “on the beat,” important statistical and analytical functions are neglected, with adverse consequences for law and order. If the measure is the overall crime rate, the implicit incentive is for policemen to underreport all crimes. If performance is assessed on the basis of a specific crime, underreporting of that crime and neglect of crime prevention in general are likely.
- If hospital subsidies are based on the length of patients’ waiting lists, hospital managers and doctors will keep noncritical cases waiting as long as possible and will spend inordinate amounts of time on other cases (higher-quality care for some, none for others). If performance is assessed instead on the basis of the number of patients treated, the quality of care may suffer (Williams 1996).
- When an aboriginal tribe in Australia was informed that its sanitation and other subsidies would depend on their performance in keeping sanitary facilities clean, they did so most effectively by thoroughly cleaning the toilets, and then closing them to the public.

The Turkish Evasion (*If It Ain’t Worth Doing, It Ain’t Worth Doing Well*)

A traditional Turkish folk story tells of a man who searched diligently for his purse on the main street because, he said, it was too dark in the back alley where he had lost it. To be sure, it is very difficult to measure performance in meaningful ways, but this can never justify measuring performance in easy but meaningless ways. The “tyranny of the measurable” is in evidence here. Let’s elaborate.

According to a well-known management consulting rule, “what gets measured, gets done”. This may be valid in private sector activity, where the bottom line of profit (or sales, or return on investment, etc.) is both measurable and meaningful. It is much more doubtful in the public sector. There are three obvious conditions for this rule to make sense: (i) the right thing must be measured; (ii) the thing must be measured right; and (iii) there must be consequences if it does not get done. As we have seen earlier, none of these conditions is easy to meet.

Even more of a complication is the obvious corollary of the rule: what does *not* get measured, does *not* get done. The Turkish Evasion warns us that, in the public sector, the least measurable activities may be the most

important ones (e.g., equity or social peace). Finally, as noted, it is never enough to assess the short-term consequences of changes in organization or in incentives, which are usually positive (nor, as stressed earlier, to limit attention to the benefits expected without considering the costs). Both the expected costs and benefits of introducing performance measurement must be considered, and in a long-term rather than immediate perspective.

The Dreedle Illusion (*Better About Right than Exactly Wrong*)

In Joseph Heller's classic antiwar satirical novel, *Catch-22*, the commanding general Dreedle, enamored of "tight bombing patterns," praises a pilot whose raid produced an orderly set of bomb craters in an empty field, and scolds another who destroyed the assigned target with bomb hits scattered all over it.

Clean spreadsheets with neat indicators of clear results and timely monitoring do nothing to stimulate performance if the indicators are not relevant to the outcome sought (or, even worse, if the data themselves are phony). In fact, this false accuracy can result in channeling civil servants' energies toward presenting the data better, or even manipulating the data to make them appear to fit an orderly pattern. Either way, their energies are channeled away from the real objective of their work—which is to improve public service.

The Mechanic's Principle (*If It Ain't Broke, Don't Fix it*)

If the public management function under consideration is performing tolerably well, reformers should be particularly mindful of the risk that changes may worsen the situation. This does not imply the need for passive acceptance of mediocre performance, but simply the need to assess downside risks and address them properly. Symmetrically, however, if the process is dysfunctional or thoroughly corrupt, radical changes may be the only way to improve it.

The Gym Prescription (*Stretch Before Exercise*)

In basic economics, the production possibilities concept makes a distinction between getting actual production closer to the ceiling set by resource and technological constraints, on the one hand, and raising the ceiling itself, on the other. By analogy, it is advisable to first stimulate all improvements possible under the existing regulatory and incentive

framework before introducing new results-based performance indicators and incentives. Second, when the time is right and the right indicators of performance have been chosen, the specific levels to be achieved need to be set. The general principle for setting any performance target is that it must be challenging but achievable. *Both* overambitious and easy targets lead to underachievement. Overambitious targets discourage effort; easy ones don't stimulate better performance. In turn, targets may be set by reference to norms and standards prevailing elsewhere or, better still, by reference to earlier performance in the same country and sector. ("Benchmarking", discussed in the next section, is often used for this purpose; the method has its uses and limitations—see, for example, Powers 1998.)

The Missouri Test (*He Who Lives by the Sword Must Be Willing to Duel*)

The motto of the US state of Missouri, seen among other things on its automobile license plates, is: "Show me". It is inherent in the logic of any performance-based system that *the system*, too, must be subject to a reality test.

Operationally, therefore, it is essential to build into public sector reforms specific provisions for the systematic assessment of the *performance of the performance system itself*. But even before the reforms are introduced, government officials in developing countries, or their public and the media, should demand that the proponent of the reform take the plain but powerful Missouri Test. The test calls for a *demonstration* that the concrete benefits are likely to outweigh the costs, and that there is a good answer to the simple question: How and when will one know whether the model has performed well or badly in this particular country? If the advice is good and the experts are right, they will be able to pass the Missouri Test, and the performance-based reform should be vigorously pursued. If not—inverting the slogan of a well-known athletic shoe company—the only sensible course of action is: Just Don't Do It.

THE "CREAM" OF GOOD PERFORMANCE

Keeping the above warnings in mind, a good performance indicator must be:

- Clear—i.e., precise and unambiguous (not necessarily quantitative);
- Relevant—i.e., appropriate to the objective at hand (not used simply because it is readily available);

- Economic—the data required should be available at reasonable cost;
- Adequate—by itself or in combination with others, the measure must provide a *sufficient* basis for the assessment of performance; and
- Monitorable—in addition to clarity and availability of information, the indicator must be amenable to independent scrutiny.⁹

If any one of these five criteria is not met, formal performance measurement should not be introduced, and other ways of assessing and stimulating good performance are needed—including the old-fashioned method of open dialogue with competent and honest managers. At the same time, however, work should be done toward meeting the CREAM criteria, in order to permit the introduction of good performance measures in the future. In developing countries, data problems and other circumstances are often inimical to the successful introduction of results-based performance indicators. However, it is still possible to assess performance in the delivery of basic social services, through opinion surveys and other means of feedback from those who know the situation best: the users of the services.

SETTING THE TARGETS: BENCHMARKING¹⁰

Benchmarking and performance measurement are closely linked. Performance measurement can be the first step toward improving the performance of a public-sector organization, and, if backed by an appropriate incentive system, it can help shift organizational focus from inputs to outputs and outcomes and thus improve efficiency and effectiveness. However, the real benefits come from using the performance measures as the basis for internal or external comparisons, with the objective of improving the performance of an organization. This is called benchmarking, the technique of comparing business practices and performance levels between organizations to identify opportunities for making improvements in the economy, efficiency, or effectiveness of an organization's activities.

There are two main approaches to benchmarking, i.e., metrics and process benchmarking. Metrics benchmarking focuses on the calculation of numerical performance indicators such as unit costs, response time, and number of customer complaints, which can then be compared with similar data from other organizations in the same field. Metrics benchmarking is a useful diagnostic tool, as it can help an organization to identify the less efficient areas and provide targets it can aim at. However, the performance rankings shown in the so-called "league tables" are of limited use and even misleading, as different organizations are subject to different external

constraints, or which they have little control. More useful is the method of benchmarking the performance of an organization against performance of the *same* organization in the past. Even this, however, does not indicate what improvements can be made and how. For that purpose one has to turn to process benchmarking, which focuses on the comparison of the processes and activities underlying the performance of a function. Thus, metrics allow to identify the problem areas, and process benchmarking help to find ways to deal with the problem.

The first steps in a process benchmarking exercise involve preparing “process maps” for the activities in the selected area of focus, collecting information on resources consumed by those activities, and analyzing the practices, working methods, and policies that determine the performance of those activities. Generally, this stage reveals many obvious inefficiencies in processes, which, if eliminated, can yield significant performance improvements. The next steps are to obtain comparator data, compare the processes, develop recommendations, and implement change. After the changes have been introduced, the new values of the performance indicators provide a measure of the improvements achieved and the basis for starting the next round of benchmarking. Therefore, this technique is often referred to in the literature as benchmarking and continuous improvement.

For the purposes of benchmarking, comparators can be either internal or external. The former refers to a situation where comparisons are made between separate divisions of the same organization where similar processes are performed (e.g., multi-site organizations such as the tax, health, or education department can compare the performance of their offices, hospitals, or schools in different cities). External comparators can be direct competitors, i.e., organizations providing the same product or service. For example, the public sector could benchmark schools or hospitals it runs against those run by the private sector or NGOs in the same area; or by other public-sector bodies performing similar processes, such as land registration and vehicle registration agencies; or by the best organization around, public or private, in the case of similar business processes, such as in the areas of accounting, information systems, procurement, payroll, or customer service. These are the so-called league tables. However, it is often helpful to start with internal benchmarking, i.e., comparing performance measures between different offices or sites, understanding the processes and methods that explain the differences in the measures, and deciding what is the best internal practice, before going to outside comparators.

Benchmarking can yield additional benefits in the public sector by introducing a form of competition. If the results are publicized and general recognition, promotions, and career opportunities of public-sector managers are linked to the relative performance of their offices, divisions, or ministries, it can be a powerful force for improvement in the public sector. However, the league table approach can be demotivating for those at the bottom of the league. Motivation is better fostered by focusing on the gap between the individual unit and the best unit, and, as noted, the changes over time. Benchmarking also enables meaningful and realistic performance targets to be set, and can help to increase the client-orientation of the organization.

Some of the problems encountered in the application of the benchmarking techniques are necessarily the same as those for performance measurement, i.e., capturing the important attributes of the product in question, agreeing on what is to be measured and how to measure it, and ensuring the comparability of performance between organizations. In addition, because benchmarking is a resource-intensive technique, the scope of any single benchmarking exercise must be restricted to the key areas, i.e., those that account for the largest component of costs or where the performance gaps are widest, or both. It is also important to avoid excessive detail in collecting data or mapping processes as it could divert effort from the primary purpose of benchmarking, i.e., identifying better practices and implementing the lessons learned. Finally, a critical success factor in benchmarking is the commitment of the senior management to improving the performance of the organization. (Box 18.1 gives highlights of Hong Kong, China's experience with benchmarking.)

Box 18.1

Benchmarking in Hong Kong, China: Mass Transit Railway Excels in Worldwide Industry Study

The Mass Transit Railway Corporation (MTRC) carries 2.4 million customers daily throughout Hong Kong, China. The organization is consistently rated as one of the best in Hong Kong, China.

MTRC conducts performance benchmarking annually through its three-year-old benchmarking project, Community of Metros (CoMET). The benchmarking group spent considerable time defining the 18 key performance indicators listed below. In half of the indicators, MTRC was classified as best in class but weak in staff efficiency and incident management. To improve these areas, MTRC set up special task forces that met regularly, and made site visits to the best performers. However, it was seen that some industry challenges still stand in the way of certain improvements, and factors like social system, culture, and purchasing power make it difficult to adopt lessons learned from the study.

Category	Key Performance Indicator	Best Practice	MTRC's Score	MTRC's Ranking
Financial	1. Total cost per passenger	0.16	0.42	2
Performance	2. Operating cost per passenger	0.09	0.24	2
	3. Maintenance cost per revenue car operating km	0.59	1.30	5
(expressed in BCU)	4. Fare revenue per passenger	1.15	0.61	4
	5. Total commercial revenue/Operating cost (including maintenance)	6.19	6.19	1
	6. Operating cost/Revenue car operating km	0.20	1.05	3
	7. Total cost/Revenue car operating km	0.77	4.12	2
Efficiency	8. Passenger journey/Total staff + contractor hours	64	62	2
	9. Revenue capacity km/Total staff + contractor hours	2,253	1,960	2
	10. Revenue car km/Total staff hours	14	6	7
Asset Utilization	11. Passenger km/Capacity km	34.07%	23.64%	3
Reliability	12. Capacity km/Track km	596,583,750	596,583,750	1
	13. Revenue car operating hours between incidents	12,074	12,074	1
	14. Car operating hours/Total hours' delay	50,978	50,978	1
	15. Trains on time/Total trains	99.79%	99.79%	1
	16. Revenue car operating km/Total incidents	361,718	361,718	1
Service Quality	17. Total passenger-hours' delay per 1,000 passenger journeys	2.38 min	2.38 min	1
	18. Passenger journeys on time/Total passenger journeys	99.80%	99.80%	1

continued on next page

Box 18.1 (cont'd.)

On the basis of the benchmarking findings, among other actions MTRC redeveloped its supplier selection criteria and created a fully computerized purchase ordering system. These changes achieved substantial cost savings. MTRC has also cut in half its damage/shortage/rejection rate in the past two years and received better pricing because of quantity discounts and better shipping arrangements.

Source: Powers (1998).

KEY POINTS AND DIRECTIONS OF IMPROVEMENT

Key Points

Introducing a stronger performance orientation in public administration is important in most countries. However, this task is neither simple nor self-evident. Experience shows that mistakes and counterproductive results in this area have usually originated from neglect of the complexity and implementation difficulties of performance measurement. Many of these mistakes could have been avoided by identifying the real administrative problem, defining correctly the objective of intervention, and being realistic about actual monitoring and implementation.

Performance is a complex issue partly because the concept itself is not simple—with a subjective dimension in terms of individual effort and an objective dimension in terms of results. While it is important not to neglect entirely the subjective effort, and recognize it in appropriate ways, performance should be measured mainly in terms of results. However, the results themselves need to be defined carefully. They can be specified in terms of inputs (the resources used to produce a public service); or outputs (the service itself); or outcomes (the purpose achieved by producing the service); or good process. The performance criteria for inputs, outputs and outcomes are, respectively, economy, efficiency, and effectiveness.

Each of these performance indicators has advantages and disadvantages. Exclusive focus on good procurement and utilization of inputs leads to a means mentality that neglects the purposes for which the resources are obtained. Output indicators are more appropriate for activities close to the ultimate user but not for upstream public activities such as policy analysis.

Outcome indicators are generally more relevant, but also less useful for allocating responsibility. And attention to due process, which is essential for the long term, becomes sterile formalism if it is viewed as an end in itself.

Because outputs are more quantifiable but more narrow, while outcomes are broader but also influenced by factors outside the control of the civil servant or organization in charge, there is an accountability trade-off—by which accountability can be either tight but narrow (through output indicators) or broad but loose (through outcome indicators). Consequently, it is important to use a combination of indicators of performance, and never to rely exclusively on any single indicator. In general, a good performance indicator must meet the “CREAM” criterion—that is, be Clear, Relevant, Economic, Adequate and Monitorable. If any of these conditions is not met, formal performance measurement should not be introduced, and other ways of assessing and stimulating good performance should be considered.

If and when performance indicators are introduced, appropriate target levels need to be set. Targets that are too easy or too ambitious both lead to underperformance. The setting of challenging but achievable targets can be helped by benchmarking, i.e., comparison with standards of performance in similar organizations or for the same organization at different times. In general, comparisons with other organizations or other countries are problematic because the circumstances are rarely the same. For example, when evaluated by student achievement, schools in poor neighborhoods typically compare unfavorably with schools in rich neighborhoods, but for reasons that are not necessarily related to the performance of teachers or administrators. More reliable are time comparisons, provided that resources and other basic parameters have not changed between the two periods being compared. Finally, an interesting option is process benchmarking, which compares the performance of similar organizations in terms not of outputs but of the soundness of the procedures followed.

Directions of Improvement

Injecting formal performance-related elements into public management requires extreme care, both because better performance orientation is critical for improving public administration and because there are many wrong ways of pushing it and only a few ways of doing it right. The suitability of performance measurement and the specific indicators themselves depend among other things on the sector in question.

Concerning the overall approach to fostering better administrative performance, the lessons of international experience are the following:

- Never confuse the end of better performance orientation with any one of the specific means for achieving it. There are many ways to foster performance. Depending on circumstances, performance improvements may justify the use of quantitative indicators; or call for qualitative indicators; or rely on dialogue, moral suasion, peer pressure, and other means that do not entail overt performance indicators of any sort. When performance indicators are appropriate, they should normally be introduced without wholesale changes in administrative or budgetary systems.
- Consider the probable impact of introducing performance indicators on individuals' behavior, especially in multi-ethnic societies and very small economies, and take compensatory or "insurance" measures.
- Understand the different uses and limitations of input, output, outcome and process indicators of performance, and tailor the use of each to the specific sector and issue in question. Whenever possible, use a combination of indicators to assess performance, rather than any single one.
- Assure robust monitoring of performance, with swift and predictable consequences.
- Performance indicators can be used in the dialogue between the line ministries and the central ministries (or the public), but direct and mechanical links to procedures, personnel, or budgetary appropriations should generally be avoided in developing countries.
- Build-in provisions for the systematic assessment of *performance of the performance system itself*. It is inherent in the logic of any performance management system that it, too, must be subject to a reality check, and to periodic proof that its concrete benefits outweigh the cost.

Beyond the above caveats, it is important to constantly be on the lookout for any possibility to expand the service awareness of government administration; raise the rewards (not necessarily monetary) for good performance and the sanctions for unsatisfactory performance; and keep

under constant review the possibility of introducing the various tools for measuring and monitoring performance. In all these tasks, systematic feedback from the employees, the service users, and the public at large is invaluable, and so is an informed and aggressive free media.

The actual process of introducing performance indicators into the public administration can consist of the following stages:

- pick one or two government departments that provide services directly to the public;
- introduce simple performance measures at an acceptable cost (including transactions cost);
- monitor closely the functioning and impact of the measures;
- debug the measures and adjust as needed;
- gradually expand the application of performance measures to other governmental areas as and when appropriate; and, most importantly,
- stop upon reaching the point of diminishing returns.

NOTES

- ¹ See Bird (1971). There is evidence that Wagner's Law may have ceased to be operative in developed countries in the early 1990s, but it is still very much in evidence in developing countries, where the size of government is closely related to the country's per capita income (Schiavo-Campo, 1998).
- ² See Hood (1991) for an exposition of the NPM, and Borins (1995) and Savoie (1995) for a summary of the arguments for and against the NPM approach, respectively.
- ³ In the ancient adage: Man does not live by bread alone. A reductionist view of human nature risks, in time, sharply reducing public sector effectiveness, and increasing the risk of corruption.
- ⁴ These measures are discussed in a variety of sources, e.g., Beeton, ed. (1988) and Fédération des Experts Comptables Européens (1991).
- ⁵ *Impact*, often used as a synonym for outcome, is more properly defined as the *value added* from the activity, i.e., the gross outcome minus the contribution from other entities or activities. The notion is important in that it takes some account of favorable or unfavorable circumstances beyond the control of those responsible. However, impact (in this sense of value added) is nearly impossible to measure, and is not discussed further in this chapter.
- ⁶ Note that not all useful data concerning a public service are necessarily performance indicators. For example, the percentage of arrests stemming from citizens' direct complaints is a useful statistic for law enforcement, but says little, if anything, about the performance of the law enforcement apparatus.

- ⁷ For process indicators, accountability can be stronger or weaker depending on the nature of the public activity.
- ⁸ Perrow (1986), quoted in Laking (1998).
- ⁹ Shand (1998) lists many more requirements for performance indicators. However, several are desirable but not mandatory, and others are in fact different dimensions of the five requirements listed above.
- ¹⁰ We are indebted to Naved Hamid and Gie Villareal for this section.

