

Performance Measurement: A Technical Perspective

Clyde Tucker
Washington, DC

Customer satisfaction, public policy evaluation, data quality, technical expertise, trend analysis

I. Introduction

Performance-based evaluation, such as that required in the Government Performance Review Act (GPRA) of 1993 in the U.S., is an important step toward government accountability. Measures of performance are not new to the government, but GPRA and other legislation like it, puts a spotlight on them, not only making these measures mandatory but also subject to higher review. When it comes to creating these measures, however, the devil is in the details. So what I want to focus on in this paper are the technical problems associated with measuring governmental performance, in the U.S. and elsewhere.

Much of program evaluation research relies on measures of customer satisfaction. Advocates of the Total Quality Method (TQM) are advocates of such measures, and they are one important component of a performance measurement system. Yet, customer satisfaction with government programs is not a sufficient indicator. It can be, especially in government, only an imperfect indicator of the output resulting from a long process. After all, manufacturers would not abandon control charts just because customer satisfaction ratings were good.

The public's view of any consumer product is a clouded one. The public only sees the finished product and cannot possibly know the ins and outs of the manufacturing process. The quality measures used by experts throughout that process, hopefully to assure customer satisfaction in the end, depend upon a great deal of research and product knowledge. Thus, it falls to technicians to develop and implement good internal performance measurement. This is no easy task, even in the private sector, and it is clearly more difficult in the public sector.

The next section highlights some of the problems with developing performance measurement systems in the government, in particular, democratic ones. The third section proposes some solutions to these problems, and the final section looks to the long-term future of government performance measurement.

II. The Problems

1. In the public sector, more often than not, someone wins and someone loses.

This is a political fact, and it can affect performance measurement. Governmental programs are a reflection of value judgments that are arrived at through the political process. Thus, there is always a constituency (maybe not a majority, but often a significant minority) which opposes a program or a decision. In this environment, even objective measures of performance can be politicized. This is complicated by the fact that there can be any number of performance measures. How the U.S. is doing in Iraq depends on which measures you think are important. Although the U.S. Census count in 2000 was evaluated according to technical guidelines, ultimately the judgment of the accuracy of the evaluation depended on whether you were a winner or a loser. Thus, even sophisticated measurement may not be accepted in some quarters.

2. Managing the government is not easy

To begin with, changes in political administrations can produce significant restructuring and new sets of priorities. Changes in management teams in the private sector also can cause disruption, but, at least, the goal (profit) usually remains the same. Differences in the political philosophies of the major political parties may lead to different goals in the public sector. Furthermore, the direction that the executive wants to go may not coincide with that of the appropriator, legislatures.

Of course, there are provincial and local governments, which the national government may not be able to influence like a company's national headquarters can its subsidiaries.

Even the executive departments have their separate cultures, and, like the other parts of government, their own constituencies. The existence of these powerful and often opposing constituencies at all levels of government can provide a shield that a government entity can use to avoid instituting changes, such as the development of an adequate performance measurement system.

3. Determining the market value for government products is difficult.

How do you value the intangible benefits resulting from government action? Government outputs cannot always be lined up and counted. How do you determine the value of human life (or the quality of life), which can be affected by government programs ranging from defense, to health care, to national parks? Perhaps, customer satisfaction is one measure, but government actions are often so nebulous that average citizens have trouble grasping what the particular costs and benefits are for them individually. Furthermore, the government outputs tend to be unique. They can't be obtained in the open market. So no standard of comparison exists for weighing the relative costs and benefits of similar products. This means customers must rely on comparisons over time, which can be a complicated cognitive exercise.

4. The calculation of future costs and benefits is complicated.

Because the benefits and costs of government programs often will occur in the distant future, it is difficult to evaluate them within the first few years after they are implemented. This is especially true for new programs. In the U.S., witness the current debate over the potential benefits and costs of health care reform. Doesn't the fate of the U.S. Census Bureau's American Community Survey depend, to some extent, on its perceived long term costs and benefits relative to the Census long form?

5. Multiple indicators necessitate a balancing act.

Many government programs either have multiple outputs or a single output that can be measured in several ways. Obtaining agreement on performance goals or objectives becomes difficult (Of course, this also can be true in the private sector.). For instance, in the U.S. government statistical agencies, program managers often focus on the timely release of data products as the most important goal, but statisticians tend to make data quality the highest priority. This situation requires us to develop a set of weights (usually subjective) for the multiple outputs or multiple indicators.

6. Resources have to be found to conduct performance measurement.

The goal of performance measurement is clearly to improve government programs, but it is difficult to separate the resources for program improvement from those for measuring the improvement. They tend to go hand in hand. The private sector has to find the resources, too, but they have a better way to justify them—the profit margin. The first resource to consider is time. It takes time to develop a performance measurement system and, ultimately, program improvements. Part of the process involves the time needed to experiment with and test alternative measures and improvements. This then becomes a problem where technical staff must work with program staff for an extended period of time, sometimes for several years.

Of course, we cannot ignore the importance of money in this endeavor. Money is not only needed to pay for staff time to develop quality measures and product improvements, but also to pay for, at least in the U.S. government agencies, the extra data collection or other research activities taking place outside the normal production functions of the agency.

Then, we come perhaps to the most precious resource of all—the skill set of the agency staff. Fortunately, major U.S. statistical agencies have dedicated research staff in both substantive and statistical matters available to do much of the technical work. The substantive experts (e.g., economists, medical doctors and various natural scientists) have the theoretical background to

design relevant quality measures and sound improvements. Statisticians bring to this their expertise in sampling, estimation and analysis, and social scientists are knowledgeable about measurement error and the design of data collection procedures for implementing improvements. All of these experts, in one way or another, contribute to the development of performance measures.

Some agencies do not possess this technical expertise. Since they must rely on outside contractors, they can be at a disadvantage. The agency may not have the technical knowledge to monitor the contract effectively, and the communications between agency staff and contract staff may not be optimal. This is true even when the work is being done by another government agency. In addition, contractors may use performance measures that minimize the shortcomings of their designs. Without in-house expertise that can evaluate the contractor's product, agencies may accept what the contractor has done and find out only later about the problems.

Good technical staff is hard to come by and harder to keep. They are in great demand in the private sector, and the rewards are greater there (e.g., the pharmaceutical industry in the case of statisticians). Secondly, the government is not as attractive an employer in many countries compared to a generation or more ago. Even for those who want to enter the public service, the hiring process can be daunting. Furthermore, besides senior management, the current surge in retirements will hit the technical areas the hardest. And no one may be coming in behind them.

7. Besides the quality of the product, we must be concerned with the quality of the measures of performance.

Performance measurement is only as good as the measures you use. It may be difficult to measure the most important outcomes, and program managers naturally gravitate towards those outcomes that are easily measured. Going back to an earlier BLS example, timeliness is much easier to measure than data quality. Although we have well-defined measures of one side of data quality, that is, the precision of our statistical estimates, the accuracy of these estimates is much more difficult to assess. Determining the accuracy of statistical estimates relies on advances in the relatively new field of nonsampling error research. Discoveries are being made all the time, but there is no coherent body of knowledge, as in sampling theory, to lead us to good measures. Furthermore, there are relatively few experts in nonsampling error in government agencies. Academicians have done most of the work. Until the field is more developed, managers will use measures such as timeliness and response rates to measure the performance of their statistical programs.

Outside statistical agencies, adequate performance measures also are difficult to come by. Take, for example, the development of weapons systems. There are four technical issues involving performance measurement in this case that I will mention. The first is that these systems are developed over many years so it's hard to evaluate them on a year-by-year basis. By the time evaluation is possible, billions of dollars have been spent. Only a few contractors exist in this environment, so competition is limited. Add to that the fact that the systems are so expensive bids can only be evaluated at the "drawing board" stage. No attempt is made to develop and compare parallel systems. It also is expensive, and sometimes dangerous, to test a system once it is completed. So few tests may be run, and it is quite unlikely the systems will be tested under battlefield conditions until much later. Finally, one way of evaluating weapons systems is to compare the costs of the systems to the "perceived" benefit, which, as I already said, is difficult to measure. But, just on the cost side, the picture has not been a pretty one. Getting a detailed accounting of the costs is not easily accomplished. Added to that is the fact that there are often large cost overruns that are difficult to contain once development is underway, and the reasons for these overruns are murky at best. This situation, by the way, is not unique to weapons systems. Have you ever heard of a highway project being completed within budget? Without reliable measures of cost, what good are the quality measures?

III. Possible Solutions

1. Improve communications

I believe the communications between all the players in performance measurement must be improved. More cross-agency discussion of the technical aspects of creating and using a performance measurement system is needed. These discussions should include such topics as prioritizing goals, indicator selection and construction, methods of quantitative analysis of the measures, and the design of and experimentation with improvement options. The different countries should work more closely together to see that state-of-the-art performance measurement and analysis is carried out by their agencies. More interaction between technical staff and field staff (including those in local government) is needed to make sure the use of the performance measures in the administration of programs throughout their countries takes place. Finally, the public could use more information about performance measurement and what it is designed to do.

2. International panel

To focus awareness on the technical aspects of performance measurement I suggest holding a series of seminars or a conference sponsored by Eurostat, the U.N., or the International Statistical Institute). A preliminary workshop might serve as a good start.

3. Cost-benefit analysis

Improvements in cost accounting for the development of performance measurement systems and the quality improvements that result from their use should be sought. Also, quantitative assessments of the benefits derived from performance measurement are needed. All of this information will serve as input to a cost-benefit model of performance measurement to justify resource allocation decisions.

4. Cadre of technical experts

A cadre of technical experts (both in measurement and analysis) should be assembled and given appropriate rewards to consult with agencies on the development and use of performance measurement systems. These experts would be drawn from a variety of technical agencies and even across countries (possibly the U.N. Statistical Office could facilitate this) and would not necessarily be expected to be subject-matter experts. Instead, they would consult with subject-matter experts in the different programs about the development of performance measurement systems tailored to the programs. These technical experts also should be made available to local governments.

5. Trend analysis

The most useful information from the analysis of performance measures likely will come from the establishment of a baseline followed by trend analysis. The agencies should be encouraged to include this type of analysis in their plans. Of course, time series experts should be among the assembled cadre of technical experts already discussed.

6. Customer satisfaction

More research is needed into ways to improve the measurement of customer satisfaction. In particular, there should be an emphasis placed on creating rigorous measures of the value of investment in specific government programs. This won't be easy, because government program managers, let alone the public, don't think in these terms. Perhaps, the measurement of these values is the most difficult task of all.

IV. Looking to the Future

The development and use of performance measurement systems is necessarily a long-term commitment. For one thing, it will take some time to produce meaningful trend data. So agencies must be expected to allocate sufficient resources to this effort for years to come. Furthermore, resources have to be devoted to developing the expertise needed to make improvements in the system and analyze the results over time.

On the technical side, methods for not only measuring and analyzing performance but also evaluating the quality of this work will be needed. Expertise in nonsampling error measurement will be important here. Finally, an ultimate goal should be to develop a system of measures that can take into account not only the quality of government performance but its true value to the public.