The Performance Measurement Puzzle

There is a simple and persuasive proposition that is quite common in government policy and practice: better measurements of performance will lead to overall improvements in government. That proposition is fundamental to any notion of governing as rational decision making, from at least as far back as the Program Planning and Budgeting Systems (PPBS) and government accountability movements in the 1960's, up to the emergence of ComStat-style programs currently operating in many agencies. Performance measurement is central as well to the President's Management Agenda for improving U.S. federal agency operations, and many similar initiatives that can be found in state agencies. In spite of this long history of concern with performance measurement, however, it remains a puzzling problem for governments at all levels.

Recent work here at CTG addressed some aspects of that puzzle and provided us with some reflections on government performance measurement. Those reflections involve three questions that are closely related, but speak to different parts of the overall puzzle:

1. What to measure. To be useful, measurement must probe beneath the general performance goals of government to employ specific indicators and data elements. However, identifying and agreeing on these can present daunting challenges.
2. How to conduct valid measurement and analysis of the results. Measurement issues are central to the feasibility of performance assessment as well as its credibility.
3. How to link the measurements to both operations and the longer term outcomes of government programs. Measuring outcomes alone is necessary, but not sufficient.

CTG explored these questions in projects at different levels of government and with varying goals and scopes of operation. They included work with one local government in New York State which sought to define and measure performance in a particular area of government: policing. That case clearly demonstrated why performance measurement is never neutral, with its potential to affect many aspects of government operations and stakeholder interests.

Another project addressed the many questions involved in identifying and collecting the valid data needed for comprehensive performance measurement at the national level, across government agencies. In this project, the Turkish Ministry of Finance joined CTG in a workshop to help develop their governmentwide performance management program. The workshop focused on frameworks for linking budget-making to cost and operational data from government agencies and to evidence of the results they are intended to achieve.

A third recent project took on the question of how to expand the scope of performance measurement. That effort focused on ways to include the public value of government IT investments; the social, economic, and political returns. Lessons from each of these projects help to fill in pieces of the performance measurement puzzle.

WHAT TO MEASURE?

The scope of performance measurement can be problematic for several reasons. For any particular government program or activity there are sure to be multiple goals, stakeholders, and possible indicators of effectiveness. Some level of consensus about goals and priorities is necessary to mobilize the support and resources needed for a measurement effort. In the local policing project, the main issues were not availability, but usefulness of data. As the project report noted:

The critical question … is not just can the department develop a set of categories, indicators, and measures that they believe will be useful in assessing their performance, but can Town management and the PD come to some consensus about these elements and agree to use them as the foundation of future examinations of department priorities, practices, and outcomes.

This question was ultimately resolved by identifying eleven broad performance categories aligned with the performance goals [see box below]. Indicators and measurements for each were then identified through broad participation of government managers and staff. The result was a performance measurement framework with broad support and a realistic set of measurements that could be collected and used without major disruption to existing operations.

This list illustrates a rather expansive response to the question of what to measure, in particular, how the performance of a police department or any other government unit can be perceived from different perspectives. The list includes measures relevant from the point of view of internal department operations, like officer safety and

© 2008 The Research Foundation of State University of New York

morale, along with others relevant from the point of view of the community at large, like public safety and responsiveness. When the scope of performance measurement is opened to this latter, public perspective, many more potential indicators and measurement problems are revealed as well.

Exploring some of these public value measurement problems was the focus of a different CTG project that developed a framework for assessing public returns on government IT investment. The performance perspective employed in that project identified performance goals in terms of a public value proposition, i.e., the value to the public returned from government operations or investments.

That value proposition must be broadly conceived to do justice to the scope of government and how it affects individuals, groups, and both public and private organizations. The research in that project revealed an expanded way to describe public value in terms of six kinds of impacts governments can have on the interests of public stakeholders:

- **Financial**—on income, asset values, liabilities, entitlements, and other kinds of wealth or risks to any of them.
- **Political**—impacts on the ability to influence government actions or policy, or to participate in public affairs as a citizen or official.
- **Social**—impacts on family or community relationships, opportunity, status, or identity.
- **Strategic**—impacts on economic or political advantage or opportunities for future gain.
- **Ideological**—impacts on beliefs, moral or ethical values, or positions.
- **Stewardship**—impacts on the public's view of government officials as faithful stewards in terms of public trust, integrity, and legitimacy.

Expanding the scope of government performance in this way brings into focus two distinct but equally important types of public value. One is performance in terms of the delivery of benefits directly to citizens. The other is performance that enhances the value of government itself as a public asset. Actions and programs that make government more transparent, more just, or a better steward have added public value, a non-financial but nonetheless important aspect of performance. This framework describes how to include both in public value assessments.

Such an expanded scope of performance measurement has both positive and negative implications. More things to measure means more cost and complexity in the measurement process. Increasing the scope of goals and measures can also greatly increase public expectations for government performance, with greater risk of disappointment and failure. Those concerns are discussed in more detail below. On the positive side, however, the greater the value potential of a government program or investment, the stronger the argument can be for mobilizing public support and resources. Neglecting an expanded view of public value propositions in performance measurement can result in lost opportunities to increase support and enthusiasm for government programs.

HOW TO MEASURE?

The method issues in performance measurement for government are as diverse and complex as the functions of government itself. Even when there is clear consensus on and specification of goals and indicators, the problems of data validity, access, quality, and interpretation remain daunting. One performance category for the police department in the example above was "responsiveness to community needs." While a laudable goal, the department could not rely on standard ways to identify, prioritize, or assign numbers to community needs or even to how "responsive" individual police actions might be. Similar problems inhabit most government performance goals and criteria. While hardly solving a wide range of these problems, the CTG project provided some valuable insights about measurement issues.

One important insight is that improving performance measurement is a systemic process. In both the local police department and the Turkish Government projects, the measurement initiatives touched all parts of the governments. Changing data collection and reporting processes had human resource and business process impacts. Most existing data collection and reporting requirements remained in place, resulting in increased work loads or shifts in work processes. Existing information systems were not fully adequate to the new tasks. Establishing new information flows within and across organizational units can encounter many technical, managerial, and political barriers. Overcoming these barriers and constraints will require effective collaboration, strong managerial support, and close attention to what is feasible, as opposed to ideal, in terms of new data and analyses.

A second insight is that the work of performance measurement improvement should be seen as ongoing, rather than a one-time project with a fixed end date. Because of the complexity and cost of performance measurement initiatives, it is usually best to build them in phases. That will provide opportunities to adjust and adapt the design

© 2008 The Research Foundation of State University of New York

to what is learned along the way. The progress of CompStat and CityStat programs in several cities has been uneven and subject to development along the way, in spite of significant successes. The reinventing of the U.S. federal government, begun in the early 1990's, and several follow-on initiatives have gone through modifications and will almost certainly continue to evolve. The Turkish Government's performance management program is planned for phased deployment, with provisions for learning and adjustment over a multi-year period. As the capabilities and demands on government change, so must the mechanisms for performance measurement.

It is also important to recognize that performance measurement has consequences. The results can be used to reward, to punish, to change work practices, affect careers, and shift political power relationships. How measurement is designed and conducted is consequently of much more than just technical interest. Therefore the validity and integrity of performance measurements and their underlying data resources are always at risk. Mitigation of those risks is then an essential part of a good performance measurement design.

LINKING PERFORMANCE MEASUREMENT TO OPERATIONS AND OUTCOMES

The linkage problems of interest here are bidirectional. That is, they involve the way performance measurement methods link in one direction to the operations and business processes within the government program, and in the other to the outcomes that represent performance. Measuring the outcomes alone is necessary but not sufficient. Without the linkages into the operations and business processes, there is no way to know where the results came from or how to intervene to improve them. Cost-effectiveness measures, for example, require knowing what resources went into creating a particular outcome as well as the value of the outcome itself. Thus one set of linkages extends into the operations and information resources of the government programs, the other into the environment where the results can be detected and measured. Each presents a different set of challenges to performance measurement.

The challenges related to the internal operations of government are typically a mix of resource constraints, inadequate data, and conflicts of interest. Expanded performance capability in a government agency requires new or re-allocated resources, often of a significant magnitude. The Turkish Government's performance based management initiative, for example, called for new data collection and reporting procedures to eventually be implemented across all national government agencies. There were similar efforts included in the studies CTG conducted for the public value assessment project: governmentwide ERP implementations in Israel and the Commonwealth of Pennsylvania, and one in the Ministry of Finance in Austria. All were multi-million dollar, multi-year initiatives that included major performance measurement components. Even though on a much smaller scale, the performance measures that emerged in the police department project described above included some substantial new data collection and reporting procedures.

The need for these investments points out the importance of expanded data resources to track the processes that influence, generate, and document performance. The extensive cost accounting, process analysis, and activity reporting capabilities needed for performance measurement are seldom fully developed in governments. Financial management systems and management information systems may require major overhauls to produce the needed data.

That same challenge applies to the assessment of outcomes. Consider the performance assessment issues faced in a program to improve the nutritional health of a city's homeless population. It may be relatively straightforward to count the number of meals served, the costs incurred, and the number of clients the program engages. But none of those measures directly reflect the nutritional health of the participants. That would require knowing much more about the health status and nutritional habits of the homeless population than is feasible to collect. Crude, indirect measures may be all that's available. This necessity to often rely on problematic inferences to gauge performance is unfortunately common in most human service programs and represents a threat to the credibility and validity of many outcome measures.

A more serious threat to the validity of performance measurement can result from vulnerability of the data to manipulation, particularly when the measurement is linked to budgets or personnel evaluation. The risk of manipulation exists anytime a government worker reports, collects, or otherwise handles data in which they have a personal interest. Therefore, performance management systems typically go to considerable lengths to eliminate or control that kind of data tampering. Colleges that use student questionnaires to evaluate teaching, for example, do not allow the professor involved to administer or handle the results. In many cases, however, performance measurement systems rely on reporting from the workers whose performance is being evaluated. Those situations call for monitoring or auditing systems to preserve the integrity of the information resources.

These problems and challenges make clear that performance measurement in government will never be an exact science. There will almost certainly be contention both inside and outside government about any performance

© 2008 The Research Foundation of State University of New York

assessment, with many valid questions about the value of its results. However there is also great promise in the efforts to improve performance measurement capabilities. They can shed valuable new light on areas where real improvements are possible and where more efficient use can be made of public resources. Though less than perfect, these kinds of measurement initiatives can be very valuable learning experiences as a foundation for government improvements.

Anthony M. Cresswell, Interim Director, Center for Technology in Government

© 2008 The Research Foundation of State University of New York