

## Discussion

Miron L. Straf

National Academy of Sciences--National Research Council

### Introduction

As I read these excellent papers, two thoughts come to mind for a title for this session. Reading Kirkendall's paper, I would name the session, "Statistician: Measure thyself," reading the paper by Collins, Rapoport, and Tupek, I would name it, "Let's win one for the GPRA."

I may have been asked to comment on these papers about performance measures for research and statistical agencies, because of a relevant study of the Committee on National Statistics (CNSTAT). The Committee has convened a panel to assess and recommend performance measures for local, state, and federal officials to use in appraising progress in meeting the objectives of public health Performance Partnership Grants in several areas, including mental health, substance abuse, HIV, chronic diseases, and preventive health services. The panel, sponsored by the Department of Health and Human Services, is chaired by CNSTAT member Edward B. Perrin; Jeffrey J. Koshel is study director. My remarks are informed by the development of this study and discussions with the panel, but are those of my own and not necessarily those of the panel, Committee, or the National Academy of Sciences--National Research Council.

In fact, it's ironic that the Academy is involved at all in improving performance. The Academy is a unique institution, quasi-government and quasi-academic, and, as a result, is the only institution that is run with *both* government efficiency and academic effectiveness.

The quasi academic side approaches performance measurement and the Government Performance and Results Act of 1993 (GPRA) with cynicism, epitomized by Sidney Winter:

The producers of management advice operate in an economic environment where the tides of fad and fashion run strong. Frameworks, slogans, and buzzwords are brought forth in great profusion with attendant fanfare and claims of novelty. Although large rewards often accrue to successful fashion leaders, it is open to question whether organizations actually perform much better as a result of this activity. To the jaded eye, the latest widely acclaimed insight often looks suspiciously like a fancy repackaging of some familiar platitude or truism. Alternatively, it may be that this year's fashionable ideas are genuinely valuable--but largely because they help to correct a misallocation of attention that was itself produced by an excess of enthusiasm for ideas fashionable in the recent past. (Winter, 1994)

The quasi-government side, on the other hand, approaches GPRA and performance measurement with even more cynicism. They've seen it all before. The history was recounted by David Mathiasen at a planning meeting on improving the effectiveness of government convened at the Academy by the Commission on Behavioral and Social Sciences and Education. What follows is my excerpt and slight paraphrase of Mathiasen's remarks:

In the Kennedy-Johnson administration, we had PPB: Planning-Programming-Budgeting. Systems analysis enabled policy makers to tell policy implementers how to do their job, with the result that policy making and policy implementation became disconnected.

In the Nixon administration, we had MBO: Management by Objectives. General management concepts, such as super agencies, were rejected by the Congress. Their analog in the private sector (conglomerates) began to fail.

In the Carter administration, we had ZBB: Zero Based Budgeting. OMB tried to use it to identify trade-offs to be made between programs in different functional areas and discovered (1) no one knew how to do it and (2) politics.

In the Reagan administration, we had no solutions. Remember, government was the problem, not the solution.

In the Bush administration, we tried to reinstitute MBO, but gave up after one year. In response to the HUD scandal, OMB set up management SWAT teams. They tackled such biggies as the Railroad Retirement Commission and the Bureau of Indian Affairs. (Mathiasen, 1994)

So what are we to make of GPRA and performance measurement? I have come to think that there is a context in which performance measures and the major goals of GPRA are a good thing for research and statistical agencies. It is that context I would like to develop and, in which, discuss the papers.

### **Performance Measures and Outcomes**

The three most important things to consider in developing or selecting a performance measure are (1) outcomes, (2) outcomes, and (3) outcomes. "Outcomes are results expressed in terms of the real difference federal programs make in people's lives . . ." (Bowsher, 1996). Examples are a reduction in infant mortality or the percentage of those completing job training who are placed in jobs for which they were trained.

Why measure outcomes? Because the major purpose of GPRA is to focus an agency's attention away from process and strategies and toward outcomes. In his testimony on GPRA, Comptroller General Charles Bowsher listed a number of questions Congress should ask of agencies. First and foremost was how well is the agency measuring outcomes. "Striving to measure outcomes," he said, "will be one of the most challenging and time-consuming aspects of GPRA." OSHA, for example, can strengthen enforcement and provide innovative incentive programs. But this strategy means nothing unless it leads to safer and healthier workplaces.

Another reason to measure outcomes is that, through them, the public can hold our government accountable. And it is through accountability for results, that GPRA is to increase public trust in government.

Outcomes that tell us that our government is doing the right things, not how right we are doing things:

On the one level, . . . performance-based management is a way of thinking about how to improve what government *does*. On another level, it is a way of thinking about what government *is*. (Kettl, 1994)

What if we can't tell the extent to which a program or agency has caused the observed outcomes? The answer is that outcome measures alone are not supposed to do that: they are no substitute for program evaluation. As Harry Hatry and Joe Wholey, two founders of the GPRA movement, say:

Only when substantive in-depth program evaluations have been undertaken, or when the linkage between the performance indicators and agency actions is otherwise established, can reasonable confidence be obtained as to the extent to which the agency caused the measured results. (National Academy of Public Administration, 1994)

What if outcomes are impossible to measure? We then typically settle for measures of intermediates, such as outputs or process. But we should not abandon the concept of outcomes. Rather, we should select those intermediate performance measures from which we can predict an effect on outcomes, at least in theory if not in practice. And, most important, we should seek to learn more about the linkages between the intermediate measures and outcomes.

When the model or relationship among intermediate measures of process, output measures, and an outcome measure is known, then we can rely on the intermediate measures. An example is the case of vaccination against a disease. The outcome is incidence or prevalence of the disease. The intermediates are the numbers vaccinated per day (process) or the vaccination rate in the population (output).

If outcomes are the thing to measure, why don't people do it? Two reasons are typically given. One is that they don't have the data:

Outcome monitoring systems that work well in reporting, tracking, and using a small number of outcome-oriented measures *must* ride on substantial data collection and reporting systems that are quite extensive. (Affholter, 1994)

The other reason often given for not measuring outcomes is that agencies are reluctant to report outcomes over which they have little or no control. But this attitude reflects a perceived or real fear of misuse of these measures. Frank Fairbanks, City Manager of Phoenix, admitted the proclivity for misuse:

We also erred in using measures in a negative manner. Management, supervisors and employees came to see measures only as a control system with negative consequences. When the data was good, nothing happened. When the data was bad, management sent a jolt of retribution to the people in the work unit. Everyone began to hate the system. And soon, everyone was working hard to learn how to control

and even falsify the data. Little or no energy was focused on improving performance. (Fairbanks, 1996)

Affholter notes that "This is perhaps the greatest pitfall for outcome monitoring. Legislators and some agency heads will want to know, simply and solely, what programs or contractors should be cut or ended because money spent on them is money wasted. Outcome monitoring systems can tell those key stakeholders where to begin more in-depth examinations, not which conclusions will be justified." (Affholter, 1994)

My thesis is that we must either measure outcomes or be guided by them in what we do measure. We need to marshal or develop the data systems for this purpose. And we need the research to link other measures to outcome measures. Moreover, we must interpret and communicate how these outcome measures can be properly used. They may be used to recognize which programs have an effect, but not the extent of the effect. That can only be determined with evaluation programs or further research.

Thus, three cornerstones of outcome-based performance measures are (1) data, (2) research, and (3) interpretation and communication. Isn't filling these requirements just what research and statistical agencies are all about?

### **Performance Measures in the National Science Foundation**

I turn to the two excellent papers, the first on the National Science Foundation (NSF). NSF struggles to find meaningful performance measures related to outcomes. It starts with deriving goals for science from principal purposes of government: improving our environment, national security, and quality of our lives. An intermediate goal is to enhance the connection between fundamental research and national goals.

I would be unabashed in adding, to the goals of NSF, research for its own sake. Discoveries and the search for new knowledge are part of the human spirit. And NSF embodies that spirit. The search may be as important to us as the results obtained. Seeking just to tie the product of research to other national goals is not, for example, going to explain why we need so much of astronomy--at least not until the next big bang, and that performance measure is a long way off.

The paper identifies many of the problems of performance measures: research results cannot be predicted, applications of them are often unanticipated, their impacts occur far into the future, and the outcomes attributed to NSF are impossible to isolate. But these problems should not stop us from outcome measurement. By documenting how research has benefitted society, we give credence to our model that intermediate measures of research are related to outcomes.

As one example of an additional intermediate outcome measure, I would suggest that NSF follow up not only those who receive NSF grants but also the near winners for information on how the NSF role has affected their research productivity. That was done for the Advance Technology Program (ATP) with some surprising results. For example, about half of the applicants surveyed who did not receive ATP funding continued their projects using other funding sources. (U.S. General Accounting Office, 1996).

It is important for NSF in developing performance goals to connect them to outcomes through theories and models. By and large the paper has sought to do that, through the intermediate goals of world leadership, new knowledge, and excellence in education, leading to a focus on the following strategies:

- (1) promoting partnerships in science,
- (2) developing intellectual capital,
- (3) building the infrastructure for science, and
- (4) integrating research and education.

I am concerned, however, with the further step back to focus on the authors' set of precursor measures of research, facilities, education, and administration, and would suggest further exploration to connect more strongly how measures of these precursors relate to strategies and intermediate goals.

### **Performance Measures in the Energy Information Administration**

The second paper describes EIA's development of an excellent practical guide for improving performance. It is in the tradition of the TQM (Total Quality Management) focus on continuous improvement. My major concern is that, however useful many of the measures, outputs, and outcomes are for improving performance and management, they are not related to the ultimate purposes of the agency. Many of the outcomes in the paper, I would label as outputs.

For a statistical agency in a cabinet department, the outcomes of the agency are the outcomes of all the programs in the department, including outcomes at state and local levels as well. A major role of a statistical agency is to measure these outcomes and to provide the data for other programs to measure their outcomes. For this reason, for example, a statistical agency would be led to intermediate measures to improve the quality and credibility of its data. I would suggest that EIA, not only a forerunner in this GPRA movement, but also an agency that has advanced the practice of both data and model validation, help us to further develop measures in these directions.

### **Conclusion**

Focusing our government on outcomes presents an exciting opportunity for research and statistical agencies. A major role of a statistical agency is to measure these outcomes and to provide the data for other programs to measure their outcomes. The analogous role for a research agency is to generate the knowledge of what our government could do, in theory if not in practice, to improve the health, environment, and well-being of our society; to establish the cause and effect relationship between what we could do and the eventual outcomes; and to provide the methods of evaluation research that can be applied in conjunction with outcome measures.

These roles alone mean that research and statistical agencies are the means through which our government can focus on outcomes and be guided in making a real difference in people's lives. With GPRA, now is the moment to seize the opportunity.

## REFERENCES

- Affholter, Dennis P. (1994). Outcome monitoring, in *Handbook of Practical Program Evaluation*, Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, eds. San Francisco: Jossey-Bass.
- Bowsher, Charles A. (1996). Achieving GPRAs Objectives Requires Strong Congressional Role. Testimony before the Committee on Governmental Affairs, U.S. Senate, and the Committee on Government Reform and Oversight, House of Representatives, Washington, D.C.: U.S. GAO (GAO/T-GGD-96-79), March 6, 1996.
- Fairbanks, Frank (1996). Testimony before the Committee on Governmental Affairs, U.S. Senate, and the Committee on Government Reform and Oversight, House of Representatives, March 6, 1996.
- Kettl, Donald F. (1994). Performance measurement in the federal service: Issues and prospects, in *Measuring Performance When There is no Bottom Line*, John J. DiIulio, Jr. ed., report of a April 19, 1994 Brookings Institution conference. Center for Public Management, Working Paper 94-1. Washington, D.C.: The Brookings Institution.
- Mathiasen, David (1994). Notes on the history of public management reform in the United States since 1961, remarks prepared for a planning meeting on improving the effectiveness of government, National Academy of Sciences--National Research Council, May 18-19, 1994.
- National Academy of Public Administration (1994). Toward useful performance measurement: Lessons learned from initial pilot performance plans prepared under the Government Performance and Results Act, Report by the National Academy of Public Administration Advisory Panel on Improving Government Performance, Harry P. Hatry and Joseph W. Wholey, Co-Chairs. National Academy of Public Administration, November 1994.
- U.S. General Accounting Office (1996). *Measuring Performance: The Advanced Technology Program and Private-Sector Funding*, Washington, D.C.: U.S. GAO (Letter report, January 11, 1996, GAO/RCED-96-47)
- Winter, Sidney G. (1994). "Organizing for Continuous Improvement: Evolutionary Theory Meets the Quality Revolution," in *Evolutionary Dynamics of Organizations*, Joel A.C. Baum and Jitendra V. Singh, eds. New York: Oxford University Press, 1994.