

An unprecedented challenge to the *evaluation* profession **has** arisen from the increasing *demand for documentation of results of public and nonprofit programs*. The current *status of performance measurement efforts* is described in *this chapter*, and *many challenges facing program managers* and other users of *performance data* are identified.

Using Performance Measurement to Improve Programs

Kathryn E. Newcomer

A variety of initiatives undertaken by federal, state, and local elected officials during the last two decades have required that public managers provide evidence that their programs work. Program managers around the world, from Sydney, Australia to Sunnyvale, California, have been asked to document the results of their work. Although program effectiveness has been questioned before, current demands have been more focused, requesting both specific performance measures and targets. Funders in the nonprofit world have similarly become more insistent and focused in their requests for documentation of results. An unprecedented challenge to the evaluation profession **has** arisen out of **this** severe imbalance between demand for and supply of performance data.

Performance measurement is the label typically given the many efforts undertaken within governments and in the nonprofit sector to meet the new demand for documentation of results (Wholey and Hatry, 1992). Different stakeholders involved with programs funded by public and nonprofit sources hold different ideas about what constitutes satisfactory performance, but much of the guidance provided to program managers indicates that the intended outcomes of the programs are what should be monitored (Government Performance and Results Act of 1993; U.S. Office of Management and Budget, 1996; and Hatry, H., van Houten, T., Plantz, M. C., and Greenway, M. T., 1996.)

Assessment of service delivery at the local level of government is not new, but linking the measures, or indicators, to program mission; setting performance targets; and regularly reporting on the achievement of target levels of performance are new features in the performance measurement movement sweeping across the public and nonprofit sectors in the United States. With

program managers rushing to meet deadlines regarding development of performance measures and collection systems, questions about the effectiveness of these efforts are appropriate. What is needed to design and support an effective performance measurement system? What does effective performance measurement entail? Why is demand so far outrunning supply in this blossoming field of program evaluation?

This chapter reviews the state of the art in performance measurement. The relationship between performance measurement and program evaluation is discussed. Some of the most pressing issues and challenges facing program managers and all users of performance data are identified.

Measuring the Performance of Public and Nonprofit Programs

Programmatic performance is not an objective reality to be measured and evaluated. Performance is a socially constructed reality (Berger and Luckmann, 1967). Criteria for defining and rating a program's performance must be assembled and agreed upon by a group of officials with both the authority and responsibility for the task. With public and nonprofit programs there are many stakeholders with interests in the performance of the program, from those receiving benefits to those allocating funding, and they may value potential criteria differently.

Traditionally, *inputs*, or the resources allocated to the delivery of programs, have been measured to track programs. Traditional process-oriented management techniques in both the public and nonprofit sectors have traced budgetary and staffing figures to assess how programs are faring. Evaluation activities during the last twenty years have gone beyond basic budgetary monitoring to count program *outputs*, such as products and services delivered to clients or activities funded. Implementation evaluation has focused attention, for example, on counting outputs and assessing whether or not the outputs have been delivered as intended in legislative and regulatory guidance.

Measurement of some outputs, such as counting the number of children immunized, captures the intended result of the program. More typically, counting outputs does not convey whether or not persons, households, schools, or other institutions have been affected in the manner intended in the program's mission. *Outcome* is the term reserved for the result desired by the program's designers.

Program outcomes are the intended results of programs that many of those currently demanding evidence of the value added by programs want to see. *Outcome monitoring*, or the routine measurement and reporting of indicators of outcome-oriented results of programs, is another term that has been used to express the measurement of the outcome facet of program performance (Affholter, 1994). Outcomes may be usefully differentiated by whether or not they are *intermediate* or *end outcomes*. Intermediate outcomes include immediate reactions of clients, for example, satisfaction with training or with medical

care, and actions taken by other levels of government or by those in the non-profit or private sector. For example, firms implementing pollution control procedures upon intervention by the federal or state environmental protection agency may constitute an intermediate outcome of that agency.

End outcomes, or impacts, are the intended results of the program upon persons or society. For example, a person's or group's health and the air and water quality affected by regulated firms are end outcomes of the health and environmental protection programs. However, such things as health and air and water quality are affected by many things out of the control of the government, so it is the net impact of the program upon such phenomena, when other influences are discounted, that constitutes the intended program outcome.

It should be noted that community-level or social indicators such as teenage pregnancy rates and crime rates sometimes serve as performance measures for state and local governments. For example, Oregon's milestones and benchmarking effort has used social indicators to set state performance targets (U.S. General Accounting Office, 1994). However, the vast majority of performance measurement efforts focus at the level of an organization or program in identifying indicators of performance.

Selecting What to Measure

Performance measurement is a fairly inclusive term that may refer to the routine measurement of program inputs, outputs, intermediate outcomes, or end outcomes. Decisions about what to measure reflect two key factors: the *intended use* of the performance data and the *value priorities* of those stakeholders who choose what to measure. Performance data can be used to support a variety of decisions, and perceptions about which decisions will be affected are critical to those charged with selecting performance measures. The relative priority given to values such as efficiency, equity, and service quality, among others, also affects the selection of indicators.

Reflecting the signals emanating from citizen revolts and legislative and executive demands for evidence that programs work, most governmental calls for performance measures suggest that the measures will inform resource allocation decisions. At the federal level of government, for example, the Office of Management and Budget (OMB) has requested that agencies provide performance measures with their budgetary submissions for the past five years. OMB has sent messages through many of their initiatives that performance measures must support budgetary requests and that they can be used defensively in budget-cutting exercises. The Government Performance and Results Act of 1993 (the Results Act) has also sent the message that through planning and performance measurement government managers can "improve the confidence of the American people in the capability of the federal government by systematically holding federal agencies accountable for achieving program results, and . . . improve federal program effectiveness and public accountability by

promoting a new focus on results, service quality and customer satisfaction.” The Results Act also adds that internal management should be improved through performance measurement, but the focus on accountability, accompanied by signals from OMB, indicates that resource allocation is the primary, if not driving, use for the performance data.

Pushing performance measurement as a tool to assess managerial accountability actually implies both that managers should use the performance data to demonstrate that the money allocated to their programs has been well spent, and to improve the operations of the programs they manage. But can the same performance data be used for both purposes?

There is a significant dilemma posed by attempting to identify and collect performance data for two vastly different uses. In the evaluation profession the distinction between formative, or process, evaluation and summative evaluation is pertinent. Formative evaluations collect data about the extent or nature of program implementation. Summative evaluations collect data to assess whether or not the program has the intended impact (Scriven, 1967; Scheirer, 1994). When questions are directed toward improving program administration, certain data will be relevant. However, very different data will be relevant for decisions regarding whether or not a program should continue to receive funding at a steady level—or at all. Evaluators interested in maximizing the use of evaluation data long have stressed the importance of matching the data collected to the intended use (Patton, 1986).

Managers pressed to choose performance measures during lean budgetary times, when they typically have very limited resources for performance measurement, face a tough choice. Strategically selecting measures that will make a program look effective to funders is a likely outcome. In many cases the sorts of measures that might effectively inform program improvement decisions may provide data that managers would not find helpful for resource allocation purposes. Given a choice about the use of limited resources to develop performance measurement systems, collecting data that will support budgetary requests will probably win out over collecting internally useful, but less resource supportive, data.

Another interesting dilemma currently facing public and nonprofit managers is that the intended use for the performance data is not clear. Given the mixed messages typical of political pronouncements that performance measurement should be used to improve both the programs and the budgetary process, there may be a lack of clarity about potential use, with resulting anxiety among program managers about what to measure. Political requestors, such as mayors and city councils, may press for data that will show the public what they are getting, whereas city managers and other career managers call for internally useful data. Two or more clear signals can become confusing when line managers are trying to adhere to conflicting guidance.

Many stakeholders are involved in selecting performance measures, and they bring different values to the table. Programs delivered in both the public and nonprofit sectors have multiple constituencies and, typically, multiple

objectives. Programs are designed to achieve specific goals, such as improving the health or employability of beneficiaries, but they are also intended to be delivered in an efficient, equitable, customer-friendly, environmentally sound, and fair manner. In other words, there are many values attached to service delivery in the public and nonprofit sectors, and different stakeholders will prioritize these values differently as they choose what to measure.

Simply forging a consensus among diverse stakeholders with the authority to choose what to measure is no small feat. For example, within public agencies and nongovernmental organizations and foundations, staff in budget, financial management, and development offices view the world differently than line managers, yet they may share responsibility for envisioning performance, and appropriate measures of it, with the managers.

Donald Kettl (1996) has aptly noted that one key benefit to be derived from implementing performance measurement in the federal government is improved communication. Envisioning program missions, strategic objectives, and performance measures that will measure the extent to which objectives are met requires that many stakeholders within agencies and across governmental organizations, such as OMB, congressional committees, and the agencies, come to agreement on appropriate measurement strategies. Such horizontal and vertical communication is not easy to facilitate but is rewarding for programs bolstered by stakeholder agreement.

Performance Measurement as Program Evaluation

Program evaluation professionals might well be thrilled at the widespread calls for measurement of programmatic results to inform decision making. In fact, when initially passed, the federal Government Performance and Results Act was dubbed the “full employment act for program evaluators.” If government agencies had reinvigorated evaluation offices and used program evaluators to help design their performance measurement systems, all parties would have benefited. This typically has not been the case.

How does performance measurement relate to program evaluation? Program evaluation is clearly the more inclusive of the two approaches to reviewing programs in the public and nonprofit sectors. Program evaluation consists of the systematic description and judgment of programs and, to the extent feasible, systematic assessment of the extent to which they have the intended results (Wholey, Hatry, and Newcomer, 1994). Evaluation efforts may focus on program inputs, operations, or results.

Program evaluation is retrospective, collecting and analyzing information on existing programs. The specific evaluation strategy employed reflects who wants to know what about a program as well as who collects the information. There is great variety in the types of evaluation questions raised by program management and oversight bodies, ranging from problem-based investigations to performance assessments to evaluations of the net impact of programs (Newcomer, 1996). Evaluation efforts are currently undertaken in many

inspector general offices at the federal level, as well as by legislative audit offices and comptroller offices at state and local levels of government in the United States, and by evaluation consultants within and outside of government.

Evaluators can and often do address challenging “why” and “how” questions with program evaluation methods. Why are programs not delivering the expected results? Why does implementation of the same program vary across sites? How do specific program components contribute to outcomes achieved? Why are unintended negative and positive results occurring? Performance measurement typically captures quantitative indicators that tell what is occurring with regard to program outputs and perhaps outcomes but, in itself, will not address the how and why questions. Certainly, in some circumstances budgetary oversight bodies may not be interested in the how and why questions; they just want to know how many or how much has been delivered. But if internal management wants to know how to improve program operations, they must venture beyond the performance data.

Useful comparisons of disaggregated performance data can instruct management on program operations. Comparative analyses of client data, as well as case studies of delivery sites, are also important services program evaluators can provide. Addressing questions about quality assurance for services contracted out can and should be undertaken by program evaluators; current trends toward more privatization of governmental services make oversight of such services even more important.

Can program evaluation professionals support performance measurement efforts? Absolutely! Will these complementary efforts evolve naturally? Probably not. Only if those charged with using performance measurement systems recognize what they can learn and what they cannot learn from the performance data themselves will program evaluation be drawn on to support and strengthen performance measurement systems.

Challenges to Effective Use of Performance Measurement

The demand for performance data to demonstrate program results in agencies and nonprofits offering all sorts of services is high. What impediments currently appear to constrain the supply of useful performance data? Some of the political and communications challenges in choosing what to measure have been discussed here. More issues are raised in the chapters to follow. Exhibit 1.1 summarizes the major challenges that currently constrain the effectiveness of performance measurement in improving programs.

Abundant political support and resources are essential to ensure that performance measurement systems are designed with adequate input from key stakeholders and with technical expertise to ensure useful systems. Consultation with stakeholders in oversight bodies, service beneficiaries, and internal staff takes time and resource support. Adequate technical assistance to support

Exhibit 1.1. Challenges to Using Performance Measurements to Improve Programs in the Public and Nonprofit Sectors

1. What should be measured?
 - What decisions will the measures inform?
 - Whose decisions will the measures inform?
 - Who will select the measures?
 - What values will be served by the measures?
2. Where are the resources to support measurement systems?
 - Where will the performance measurement system be maintained?
 - Where are the resources for system design and maintenance?
 - Will top-level management ensure continued resource support?
 - Where are the individual and program-level incentives to support and use performance measurement systems?
3. How will program evaluation be used to support performance measurement?
 - Is there political will to support efforts to learn why program results were or were not achieved?
 - Is there political will to support efforts to learn how to improve programs?
 - Where are the resources to support program evaluation efforts?
4. Why are discrepancies between expectations and outcomes of performance measurement likely?
 - What sorts of technical assistance to facilitate design of performance measurement systems are needed and available?
 - Where is there potential for mixed signals about the use of performance measurement in decision making?
 - What are likely consequences of mixed signals about the use of performance measurement in decision making?
 - Why might performance measures inadequately capture program results?

both design and implementation is essential. The availability of both political and technical support should not always be assumed.

Expectations that performance measurement is attainable and will be useful are fairly high. It is certainly feasible that reasonable indicators of program outputs, and perhaps intermediate outcomes, can be routinely collected. Success stories about the use of performance measurement to support resource requests and to convince the public of the value of local government services in the United States abound (Epstein and Olsen, 1996; also see Chapter Three of this volume). Why might outcomes of measurement efforts fail to meet these raised expectations?

Factors that may hinder effective utilization of performance measurement to support decision making are political and communication challenges with which program evaluators are quite familiar. So, what advice might the evaluation profession offer to managers charged with devising useful performance measurement systems?

First, designing a performance measurement system is an extremely time- and resource-consuming process that should be undertaken with clear expectations among relevant parties about what is needed. Political will from the top of the pertinent organizations, whether that be a city or county council or a department secretary, must be in place to secure the necessary resources and

political commitment from stakeholders. In the development process, multiple stakeholders involved with programs must work together to come to agreement on programmatic mission and objectives, on the potential use of performance measures, and on the set of performance indicators that will be most useful to them.

Second, defining performance is an inherently political process. It is also an inherently governmental process. Elected representatives and their agents create and authorize programs. Knowledge of legislative history and program intent is essential to ensure that program objectives and performance measures accurately reflect the intent of the program creators. There are no correct answers when it comes to selecting performance measures. The “right” measures are so defined by those stakeholders who hold the most influence over the process. Smart consultants can help facilitate useful development processes, but they will not be able to identify performance measures. Sharp trainers cannot guarantee that their trainees will be better able to select the correct measures, either. Knowledge of the political context is more valuable than methodological expertise in this endeavor, though both are necessary skills. Concepts of performance and the most useful measures of it will evolve. Flexibility to change how performance is measured must be assured, so that systems can be upgraded and improved over time.

Third, location of a performance measurement system matters. Financial management, policy, and planning and budget offices all have stakes in the development of performance measures, but they should probably not manage the system. Line managers should be empowered to develop and use the systems. They need to take ownership over the systems so that they will be willing and able to use them. There must be adequate financial support so that performance measurement does not get relegated to line managers without resources. As additional yet uncompensated work, data collection to support performance measurement will simply not get done. Authority and resources must accompany responsibility for performance measurement.

Fourth, there must be clear communication about the use of performance measures within the relevant political organizations. Key political stakeholders, such as city councils and city manager offices, departmental budget offices, and congressional committees, must be brought together to discuss use before commitments to specific measures are made. Mixed signals or fear of punitive use can undercut line management commitment to performance measurement quickly. Even rumors of punitive use of performance measures can wreak quick damage.

Fifth, program evaluators with substantive expertise in service areas can add value to both design and use of performance measurement. Evaluators can educate potential users on pertinent distinctions. They can offer program logic modeling to help distinguish outputs from intermediate outcomes and from end outcomes. They can demonstrate the problems in drawing causal linkages between programs and end outcomes affected by many factors out of the control of the programs. They can identify the costs and benefits of different data

collection approaches for capturing programmatic outputs and outcomes. They can point to likely implementation traps, such as not providing adequate political leadership and technical support for system maintenance. And they can help moderate expectations to help reduce disappointment when measurement supporters find that performance data are not used directly in decision making. Twenty years of discussion about how to enhance utilization of evaluation has taught evaluators many valuable lessons quite pertinent to current developments in performance measurement. Our knowledge about such things as the criticality of adequate consultation **with** users and the involvement of stakeholders in every phase of evaluation efforts—from design to use—and the importance of audience-oriented presentation can certainly inform current performance measurement efforts. Performance measurement is but one facet of program evaluation, and it can be well served by the evaluation profession's institutional memory about enhancing utilization of strong performance data.

References

- Affholter, D. P. "Outcome Monitoring." In J. Wholey, H. P. Hatry, and K. E. Newcomer (eds.), *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass, 1994.
- Berger, P. L., and Luckmann, T. *The Social Construction Of Reality*. Garden City, N.Y.: Anchor Books, 1967.
- Epstein, J., and Olsen, R. T. "Lessons Learned by State and Local Governments." *The public Manager*, 1996, 25 (3), 41–44.
- Hatry, H. P., Sullivan, J. M., Fountain, J. R., Jr.; and Kremer, L. (eds.). *Service Efforts and Accomplishment Reporting: Its Time Has Come—An Overview*. Norwalk, Conn.: Government Accounting Standards Board, 1990.
- Hatry, H., van Houten, T., Plantz, M. C., and Greenway, M. T. *Measuring Program Outcomes: A Practical Approach*. Alexandria, Va.: United Way of America, 1996.
- Kettl, D. F. "Implementation of the Government Performance and Results Act of 1993: Testimony before a joint hearing of the House Committee on Government Reform and Oversight and the Senate Committee on Governmental Affairs, Mar. 6, 1996, 104th Congress.
- Newcomer, K. E. "Evaluating Public Programs." In J. L. Perry (ed.), *Handbook of Public Administration*. (2nd ed.) San Francisco: Jossey-Bass, 1996.
- Patton, M. Q. *Utilization of Focused Evaluation*. Newbury Park, Calif.: Sage, 1986.
- Scheirer, M. A. "Designing and Using Process Evaluation." In J. Wholey, H. P. Hatry, and K. E. Newcomer (eds.), *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass, 1994.
- Scriven, M. "The Methodology of Evaluation." In R. W. Tyler, R. M. Gagne, and M. Scriven (eds.), *Perspectives on Curriculum Evaluation*. American Educational Research Association. Monograph Series on Curriculum Evaluation, no. 1. Chicago: Rand McNally, 1967.
- Shadis, W. R., Cook, T. D., and Leviton, L. C. *Foundations of Program Evaluation*. Newbury Park, Calif.: Sage, 1991.
- U. S. General Accounting Office, *Managing for Results: State Experiences Provide Insights for Federal Management Reform*. GAO/GGD-95-22. Washington, D.C.: General Accounting Office, 1994.
- U. S. Office of Management and Budget. *Preparation and Submission of Budget Estimates*. Circular A-11. Washington, D.C.: Office of Management and Budget, 1996.
- Wholey, J. S., and Hatry, H. P. "The Case for Performance Monitoring." *Public Administration Review*, 1992, 5 (6), 604–610.
- Wholey, J. S., Hatry, H. P., and Newcomer, K. E. (eds.). *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass, 1994.

14 USING PERFORMANCE MEASUREMENT TO IMPROVE PROGRAMS

KATHRYN E. NEWCOMER is professor *and* chair in the Department of *Public Administration*, The George Washington University, Washington, D.C.