

# Formative and Summative Evaluation: Related Issues in Performance Measurement

JOSEPH S. WHOLEY

## ABSTRACT

Wholey first defines "performance measurement" and then explains how it can serve both formative and summative evaluation functions. He also offers the view that formative evaluation is typically more useful than summative for governmental purposes, and that performance measurement is more useful than one-shot evaluations (of either formative or summative stripe). Challenges to evaluators wishing to contribute to performance measurement methodology are outlined.

## INTRODUCTION

Scriven (1991, p. 60) suggests important, exciting purposes for our work as evaluators: "The battle for quality, efficiency, equity, and accountability ... is a battle against forces of incompetence and corruption, and a battle for the life and leadership role of this society." This paper discusses performance measurement issues related to the formative and summative roles of evaluation: formative evaluation for performance improvement, and summative evaluation for accountability, policy and budget decisionmaking, and other purposes beyond performance improvement. Like more elaborate forms of evaluation, performance measurement can serve both formative and summative purposes.

In this volume, Chen and Patton emphasize the importance of context. In the United States and in many of the other industrialized democracies, we suffer from deficits in government performance, from continuing budget deficits, and from a deficit in public trust. Our context today is one of increasing demands for effective public services, limited resources for government agencies and programs, declining public confidence in and support for government, and increasing demands for accountability by those providing public services. On a hap-

---

Joseph S. Wholey • School of Public Administration, University of Southern California, Los Angeles, California 90007.

*Evaluation Practice*, Vol. 17, No. 2, 1996, pp. 145-149.  
ISSN: 0886-1633

Copyright © 1996 by JAI Press Inc.  
All rights of reproduction in any form reserved.

---

pier note, our context is also one in which there is great interest in easing the pressures of resource constraints by providing increased management flexibility in return for accountability for results.

This paper first explores the concepts of performance and performance measurement and then suggests that performance measurement is likely to serve both formative and summative purposes at least as well as “one-shot” evaluation studies. The paper concludes with two challenges: that evaluators assist in needed performance measurement work, and that evaluators assist in learning how (and to what extent) performance measurement can help improve government performance and help restore public confidence in government.

## Performance

“Performance” is an interesting concept. “Performance” is not an objective reality out there somewhere waiting to be measured and evaluated. “Performance” is socially constructed reality (Berger & Luckmann, 1966). “Performance” exists in people’s minds if it exists anywhere at all. We have to define what “performance” means before attempting to measure performance.

“Performance” may include *inputs*; in particular, agency and program costs, such as dollars or staff-years expended. “Performance” may include *outputs*, such as products or services provided to the public. “Performance” may include *intermediate outcomes*, such as customer satisfaction and actions taken by lower levels of government or by private firms (e.g., to control pollution). “Performance” may include *end outcomes*, such as changes in environmental quality and changes in health status. “Performance” may include *net impacts*: i.e., what difference a program has made. “Performance” may include the production of *unintended outcomes*, such as costs incurred by firms or consumers in response to environmental programs. “Performance” may relate to *economy, efficiency, effectiveness, cost-effectiveness, or equity*.

Opinions will differ as to what are the key dimensions of performance. Before attempting to develop specific performance indicators, therefore, agencies should inform themselves of the expectations and priorities of key stakeholders. To the extent possible, agencies should define “performance” broadly enough to capture the dimensions that are of greatest interest to the agency’s key stakeholders—that is, they need to identify the inputs, outputs, and outcomes that are of greatest interest to intended users, such as: authorizing committees, appropriations committees, policy officials, program managers, those delivering program services, those served by the program, and others affected by or interested in the program.

## Performance Measurement

*Performance measurement (or performance monitoring)* is the periodic measurement of program performance (inputs, activities, outputs, or outcomes). Performance measurement is typically done annually to provide public accountability and assist budget decisionmaking. It can be done even more frequently to support program management and program improvement efforts. Performance measurement often includes the comparison of performance levels with prior performance or with performance goals which (for example) can be operationalized as target levels of performance for a fiscal year or some other designated time period. From the perspectives of many policymakers and managers, performance measurement *is* evaluation, although performance measurement does not typically provide information on the net impacts of policies or programs.

The first and most difficult step in performance measurement is reaching agreement on the set of performance dimensions that are important to capture and the quantitative or qualitative performance indicators that best represent those dimensions. Once “performance” has been fully operationalized in terms of reliable and valid performance indicators, those data can be collected through a variety of means, such as agency and program data systems, surveys, or ratings by trained observers. Performance goals (targets) should typically be established only after performance indicators have been pilot-tested and baseline data have been collected.

### **Playing Formative and Summative Roles**

Performance measurement can serve both formative and summative evaluation purposes. The stated purposes of performance measurement under the Government Performance and Results Act of 1993 (P.L. 103-62), for example, include *both* improving service delivery and program effectiveness *and* improving public accountability and congressional decision-making (U.S. Congress, 1993).

The question of whether summative evaluation or formative evaluation is more important is intriguing. In my view, formative evaluation takes precedence over summative evaluation. Recognizing the importance of summative evaluation for public accountability and improved policy and budget decisionmaking, I find many more needs and opportunities for formative evaluation work. Only rarely will (summative) evaluations be sufficiently timely, relevant, and conclusive to affect policy or budget decisionmaking. Formative evaluation can be used—and often is used—to assist in the much more frequent policy and management decisions that result in incremental changes designed to improve existing policies and programs. Evaluability assessment and related approaches can help agencies clarify long-term goals and objectives and improve program management and performance, for example. Evaluability assessment, process evaluation, performance measurement, case studies, and other (formative) evaluation approaches help agencies improve service quality, efficiency, and equity.

As input to policy or budget decisionmaking, annual performance measurement will often be sufficient. For program improvement, more frequent performance measurement will often be required. For program improvement, it will often be desirable to disaggregate performance information by geographic area or by type of client served (see Hatry, 1989; Hatry & Fountain, 1990).

In my view, performance measurement is more likely to serve both formative and summative purposes than are “one-shot” evaluation studies. Over the years, both evaluators and intended users have noted that evaluation studies often go unused. Policymakers and managers often find one-shot evaluation studies to be neither timely nor relevant. Opportunities to “fine tune” performance measurement systems should make performance measurement products more relevant to policymakers’ and managers’ concerns. Annual or more frequent performance measurement should provide products that are more timely from policymakers’ and managers’ perspectives.

Furthermore, performance measurement can serve as a bridge to more sophisticated formative and summative evaluation studies. It can do this by facilitating consensus on valid performance indicators, by providing (time series) data on program outcomes, and by identifying appropriate opportunities to use qualitative evaluation to explore factors contributing to performance variations over time, or performance variations among subordinate units.

## Challenges for Evaluators

Evaluators can and should help program managers, policy analysts, and others in developing appropriate performance measurement systems. Further, evaluators should help us learn how to better use performance measurement data to improve government performance and credibility. Three examples of the kinds of contributions needed from evaluators are listed below.

*Performance measurement.* Agencies need help as they prepare to measure the performance of their programs. Evaluators can and should help in clarifying key stakeholders' expectations and priorities, in identifying possible performance indicators, in assessing the feasibility and likely utility of alternative performance measurement systems, and in selecting the performance indicators in terms of which performance will be assessed. The evaluability assessment process—in particular, the use of logic models—can help agencies to meet these challenges (see Scheirer, 1994; and Wholey, 1994).

Once quantitative or qualitative performance indicators have been selected, evaluators can help in assessing performance; in particular, in assessing outcomes achieved (see Affholter, 1994; Caudle, 1994).

*Case studies.* To give policymakers and managers a sense of what it takes to do performance measurement, and to throw light on the extent to which performance measurement is, in fact, useful for improving performance, case studies of performance measurement are needed. Much of the content for performance measurement training should come from the best that agencies have accomplished to date (see Koskinen, 1995).

What we need at this point is a network of managers and evaluators working to produce case studies of the use of performance measurement. These case studies should explore the context for the development of indicators (in particular, the development of outcome indicators); describe how and why performance indicators were developed; present performance indicators that are in use; trace the use and impact of performance information; document the political, bureaucratic, and financial costs of the development and use of performance indicators; suggest "lessons learned" from the performance measurement experience; and describe planned next steps in performance measurement and the use of performance information. The American Society for Public Administration and the National Academy of Public Administration are among those supporting efforts to develop the needed case studies (see Olsen, 1995).

*Cross-case analyses.* A key issue will be whether the cost of performance measurement is justified by its use. Does it improve management, improve agency and program performance, improve budget decisionmaking, help restore public confidence in government? Through case studies and cross-case analyses, evaluators can help answer the big questions of public management and public administration; for example, how to use performance measurement to improve program performance (Behn, 1995), and what instruments of collective action best achieve which societal goals (Kirlin, 1995). We need case studies and cross-case analyses to identify the circumstances under which performance measurement can contribute to "the battle for quality, efficiency, equity, and accountability" (Scriven, 1991), and thus contribute to the preservation of government and civilization in these difficult times.

## REFERENCES

- Affholter, D. P. (1994). Outcome monitoring. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation*, pp. 96–118. San Francisco, CA: Jossey-Bass.
- Behn, R. B. (1995). The big questions of public management. *Public Administration Review*, 55(4), 313–324.
- Berger, P. L. & Luckmann, T. L. (1966). *The social construction of reality*. Garden City, NY: Doubleday.
- Caudle, S. L. (1994). Using qualitative approaches. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation*, pp. 69–95. San Francisco: Jossey-Bass.
- Hatry, H. (1989). Determining the effectiveness of government services. In J. Perry (Ed.), *Handbook of practical program evaluation*, pp. 469–482. San Francisco, CA: Jossey-Bass.
- Hatry, H., & Fountain, J. R. (1990). *Service efforts and accomplishments reporting: Its time has come*. Norwalk, CT: Governmental Accounting Standards Board.
- Kirlin, J. J. (October, 1995). *The big questions of public administration*. Unpublished manuscript.
- Koskinen, J. A. (December, 1995). *Strategic planning and performance measurement: Learning as we go*. Memorandum to members of the President's Management Council. Washington, DC: Office of Management and Budget.
- Olsen, R. (December, 1995). *GAA status report and action items*. Memorandum to members of the Government Accomplishment and Accountability Task Force. Washington, DC: American Society for Public Administration.
- Scheirer, M. A. (1994). Designing and using process evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation*, pp. 40–68. San Francisco, CA: Jossey-Bass.
- Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. D. Phillips (Eds.), *Evaluation and education: At quarter century*, pp. 19–64. Chicago: University of Chicago Press.
- U.S. Congress, Senate, Committee on Governmental Affairs (1993). *Government Performance and Results Act of 1993*, Report 103–58, 103rd Congress, 1st sess.
- Wholey, J. S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation*, pp. 15–39. San Francisco, CA: Jossey-Bass.