

# Evaluation And Utilization In The United States

Joseph S. Wholey  
School Of Public Administration  
University Of Southern California

## ABSTRACT

Cet article traite de l'évolution de l'évaluation de programme aux États-Unis depuis le début des années sixantes. Il utilisant l'exemple de six programmes menacés de disparition, l'auteur discute de l'utilisation de l'évaluation dans la prise de décision quant au maintien ou à l'expansion d'un programme. L'auteur suggère aussi des politiques et des pratiques avec lesquelles les décideurs, les questionnaires et les évaluateurs peuvent agir pour maintenir ou améliorer la performance des gouvernements.

This paper offers some reflections on where we are in evaluation and in the utilization of evaluation. We will examine what we and others can do, not only to enhance evaluation and its utilization, but also to enhance the performance of government itself.

The paper draws on the perspectives of policy-makers and managers outside the evaluation profession. It has also been influenced by the perspectives of several who were part of the 1985 evaluation meeting in Toronto: Michael Rayner, who discussed operational and strategic uses of evaluation; Michael Scriven, who first distinguished between formative and summative evaluation; Gerald Barkdoll, who emphasizes the roles of evaluators in helping managers to improve their programs; Leonard Rutman, who has helped so many in Canada and the United States to appreciate the value of program evaluation; and Leonard Bickman, who made many useful suggestions as this paper developed.

## Significant Developments in Evaluation in the United States

In the 1985 evaluation conference, Michael Scriven reminded us that evaluation has a long history: Evaluation includes curriculum evaluation, product evaluation, and personnel evaluation, as well as program evaluation. If we define *program evaluation* as the comparison of actual program performance with some standard of expected program performance and the drawing of conclusions about program effectiveness and value, the United States has been hard at work at program evaluation at least since the early 1960s. Program evaluation has been an important component of federal employment and training programs since the Manpower Development and Training Act of 1962, an important component of federal antipoverty programs since the Economic Opportunity Act of 1964, an important part of federally-funded research and demonstration programs since the mid-1960s, and an important part of federally-funded education programs since the Elementary and Secondary Education Act of 1965. Under an amendment offered by Senator Robert F. Kennedy, for example, the Elementary and Secondary Education Act has led to thousands of evaluations at local, state, and federal levels: a revolutionary expansion of educational research and evaluation.

Program evaluation was further stimulated when President Lyndon B. Johnson directed that the Defense Department's Planning Programming Budgeting System be extended to domestic agencies beginning in 1966. The ensuing attempts to do policy analysis revealed a widespread lack of information on the effectiveness of federal programs. In the Child Health Act of 1967 and the Public Health Amendments of 1968, the Administration therefore proposed and the Congress agreed to give the Secretary of Health, Education, and Welfare authority to use small percentages of program funds for evaluation of federally-funded health programs and to make corresponding reductions in the funds available for program operations. Such "evaluation set-asides" were later inserted into the legislation authorizing a number of other programs; for example, the Agriculture Department's Special Supplemental Food Program for Women, Infants, and Children (the WIC Nutrition Program). Beginning in the late 1960s, Congress also appropriated many millions of dollars each year for evaluations of employment and training programs, education programs, crime control programs, housing programs, energy programs, and other domestic programs. In the 1960s and 1970s the federal government authorized and funded many social experiments, including sophisticated evaluations, in such areas as income maintenance, health insurance, education vouchers, and housing allowances. For at least twenty-five years, Congress has strongly encouraged program evaluation.

In the late 1970s and the 1980s, some disenchantment with evaluation became apparent. Led by the Senate Appropriations Committee, Congress began to set limits on contracts for consultant services, including contracts for evaluation. "It seems as though, year after year, the same programs get re-evaluated, yet never change," the Committee complained (Senate, 1979:25). Beginning in 1981, the Reagan Administration made further substantial cuts in federal programs and overhead functions, including evaluation. When ideology and anecdotal information were sufficient for policy decisions to eliminate programs, neither summative nor formative evaluations were desired. The Reagan Administration's program of fiscal federalism, combining large numbers of categorical federal programs into block grants to the States, further reduced federal requirements and funds for evaluation at the federal level.

There are still major federal efforts in evaluation, however. Each year hundreds of federal staff years and tens of millions of federal dollars are invested in program evaluation. Such agencies as the Food and Drug Administration and the Federal Bureau of Investigation have large inhouse evaluation staffs, which they use to review and improve their programs. The General Accounting Office, which has become the world's largest evaluation organization, each year performs hundreds of program evaluations for the Congress. Though the Inspector General Act of 1978 has been primarily negative both in concept (an expansion of auditors' efforts to combat "fraud, waste, and abuse" in federal programs) and in implementation, the inspectors general are performing some evaluations designed to improve program efficiency and effectiveness.

State and local governments and nonprofit organizations have greatly in-

creased their understanding and use of evaluation over the past twenty-five years. More recently, the audit profession has developed an expanded concept of auditing that in many cases includes program evaluation as part of a financial audit.

### **Perspectives on Evaluation Utilization**

Over the past ten to fifteen years, there has been growing concern over the issue of evaluation and on working with managers to help improve their programs. Others have stressed the need for evaluation planning and design work to identify the types of evaluation information that will be relevant and useful to managers and policymakers. A third group has focused on the value of synthesizing and clarifying what has been learned from the many studies in specific program areas. Still others have emphasized the use of evaluation in the "enlightenment" of policymakers and managers, even when evaluations do not appear to result in specific policy or management decisions.

Meanwhile, policymakers and members of the public have evidenced a growing concern over the costs and performance of government. In times when government resources have been constrained and will be further constrained, evaluation will have to demonstrate its usefulness in fairly specific ways if it is to continue as an important government function. To justify its continuation on any large scale, evaluation may have to prove its value in helping managers to improve the efficiency, quality, effectiveness, and credibility of their organizations or in helping policymakers to decide whether specific programs should be curtailed or taxes levied to support program continuation.

In this context, let us focus our attention on two specific uses of evaluation that directly relate to government performance: (1) Evaluation has *operational* (or *formative*) use if management has a system that uses evaluation information to improve organizational or program performance (e.g., to improve program efficiency, quality, or effectiveness). (2) Evaluation has *strategic* (or *summative*) use if policymakers allocate resources on the basis of, and in the direction suggested by, evidence of a program's value. A management system that used evaluation information and intangible incentives to stimulate higher program performance would be an example of operational use; a specific budget decision might be an example of strategic use. A performance-based resource allocation system could provide examples of both kinds of use.

I am currently exploring these two types of evaluation use in terms of a framework that includes the context of the evaluation, characteristics of the evaluation, primary uses of the evaluation, and outcomes of the evaluation (see Table 1).

While examining how five federal agencies fared during the first three years of the Reagan Administration, I was interested to discover that Congress had maintained or expanded a number of programs whose effectiveness had been demonstrated through timely, credible evaluations or ongoing performance monitoring systems, while in the same agencies Congress had agreed to eliminate or sharply reduce many other programs generally

recognized as ineffective or inappropriate. Knowing from a performance monitoring system, for example, that the Community Health Centers program efficiently provides needed health services, Congress maintained the Community Health Centers Program as a categorical federal program (a "block grant" that only one State has accepted) and maintained program funding, while agreeing to reduce or eliminate other health services programs. Similarly, while agreeing to cut other nutrition programs, Congress maintained and then expanded the WIC Nutrition Program, a program that many evaluations and a timely evaluation synthesis had demonstrated improves nutrition and health status and saves money in other health care programs. While agreeing to reduce or eliminate other employment and training programs, furthermore, Congress maintained Job Corps funding and enrollment, using a preliminary evaluation that showed that the Job Corps' economic benefits to society outweigh the program's high costs. While agreeing to reduce or eliminate other mental health programs, Congress maintained the Community Support Program (CSP), using evaluation evidence demonstrating that the program helped mobilize State and local funding for the chronically mentally ill and demonstrating that CSP clients had less need for hospitalization as a result of the program (Wholey, 1984). When I examined some of the "winners" in detail, the evidence of evaluation utilization became even more impressive (Wholey, 1986).

**Table 1:**

**A Framework For Exploring Evaluation Utilization**

1. Context of the Evaluation

- a. Political and bureaucratic support of the organization or program
- b. Political and bureaucratic constraints
- c. Resources made available for evaluation

2. Characteristics of the Evaluation

- a. Type evaluation
- b. Stage in the evaluation when the evaluation was used
- c. Evaluation findings

3. Primary Uses of the Evaluation

- a. Agree on ambitious but realistic program objectives and program performance measures
- b. Change program activities
- c. Communicate program performance

4. Outcomes of the Evaluation

- a. Improve program performance
- b. Improve program credibility
- c. Change program resources in the direction suggested by the evaluation (in particular, maintain or expand an effective program)

At State and local levels as well, there is impressive evidence that evaluations can lead to policy decisions to maintain and expand effective programs even in times of budget restraint. Tennessee initiated a statewide prenatal program, for example, after a preliminary evaluation had demonstrated that prenatal care had reduced the incidence of low birth-weight infants in disadvantaged populations in Tennessee (Smith, 1986). In New York, the State Department of Mental Hygiene maintained the Harlem Valley Psychiatric Center (previously a candidate for closure) after internal performance monitoring and external evaluation showed that the Center efficiently provided high quality psychiatric care and other needed services in both inpatient and community setting (Levine, 1980). These examples of evaluation utilization are documented in *Performance and Credibility*, a volume based on the 1984 evaluation conference, and in an earlier work, *Evaluation and Effective Public Management* (Wholey, 1983). In these examples of evaluation utilization, the following points are significant:

1. In six programs whose futures all were in doubt, ongoing or interim evaluations produced findings that credibly documented high program performance: program efficiency, program effectiveness, and in some cases production of savings in other programs.
2. The programs differed along many dimensions: In some but not all cases, program managers had used earlier evaluation activities to help in establishing realistic, measurable objectives and in stimulating improved program performance. In some but not all cases, the programs enjoyed a good deal of political support. In some but not all cases, advocacy groups used evaluation findings in their support of the program.
3. In every case, once the program's high performance had been demonstrated, managers used that information in debates over the program's future existence and higher-level policymakers were persuaded to maintain or expand the program.

### **Some Policies and Practices That May Improve Government Performance**

From the points of view of government managers, policymakers, and advocacy groups interested in maintaining and expanding public programs, examples like these suggest a number of hypotheses worth further exploration:

1. To achieve and demonstrate high performance, those in charge of public organizations and programs must accomplish three key leadership functions:
  - Get policy and management agreements on, and commitments to, ambitious but realistic definitions of "performance" for their organizations and programs; in particular, agreements on the outcomes to be achieved and agreements on quantitative or qualitative measures of the outcomes to be achieved
  - Stimulate high organizational/program performance
  - Credibly communicate the value of the organization's activities, within and outside the organization.

2. Evaluators have analytic and behavioral skills that can help policymakers and managers to agree on appropriate definitions of "performance" for their organizations and programs. These definitions of "performance" will include policy and management agreement on:
  - Ambitious but realistic objectives for their organizations and programs, including but not limited to intermediate outcome objectives not fully under the control of management
  - Important potential side-effects to be minimized
  - Valid, reliable measures of performance.

The agreed-on performance measures may be quantitative or qualitative, and may include measures of input, process, efficiency, equity, or outcome. Outcome measures will typically be needed to demonstrate the value of an organization's activities to those outside the organization.

3. By providing timely, valid information on within-program performance variations and on variations in performance over time, and by assisting in the implementation of recommended changes, evaluators can help policymakers and managers to stimulate and achieve higher performance in the programs for which they are responsible.
4. By providing timely, valid, understandable information on organizational performance, evaluators can help policymakers and managers to credibly communicate the value of the organization's activities within and outside the organization.
5. By providing timely, valid, credible information on program performance (in particular, by providing information on program efficiency, program effectiveness, and production of savings in other programs), evaluators can help managers and policymakers to maintain and expand effective programs.
6. A series of partial evaluations is likely to be more timely and useful than a single comprehensive evaluation.
7. Verbal and graphical communication of evaluation information tends to promote utilization.

## Conclusion

In the United States, the separation of powers between legislative and executive branches of government has stimulated evaluation and enhanced the utilization of evaluation. Evaluation and evaluation utilization have been hindered, however, by the widespread lack of incentives for high performance in government organizations and programs.

To the extent that policymakers, the media, and members of the public can create incentives for high performance in government organizations and programs, the skills and products of evaluators can be useful. To enhance evaluation and its utilization in government, policymakers and managers need to change incentives in ways that will stimulate higher performance within and among government programs. The General Account-

ing Office could create new incentives for high program performance in the federal government, for example, by annually classifying each major federal program as either "relatively effective" or "relatively ineffective" in meeting its objectives and in saving money in other programs.

It appears that evaluation is coming of age, both in the United States and in Canada. Even in times of government cutbacks, evaluators can help managers to improve the performance of their programs, help managers to gain needed resources, and help policymakers to separate good programs from bad. Now is the time for evaluators to demonstrate high performance in their own work!

## References

- Levine, Murray. *From State Hospital to Psychiatric Center* (Lexington, Mass. D.C. Heath and Company, 1980).
- Smith, Jean D. "Communicating the Value of Tennessee's Prenatal Program." In Joseph S. Wholey, Mark A. Abramson, and Christopher Bellavita, Eds., *Performance and Credibility: Developing Excellence in Public and Nonprofit Organizations* (Lexington, Mass. D.C. Heath and Company, 1986), pp. 257-270.
- Wholey, Joseph S. *Evaluation and Effective Public Management* (Boston: Little, Brown and Company, 1983).
- Wholey, Joseph S. "Executive Agency Retrenchment." In Gregory B. Mills and John L. Palmer, Eds., *Federal Budget Policy in the 1980s* (Washington, D.C. The Urban Institute Press, 1984), pp. 295-332.
- Wholey, Joseph S. "WIC. Positive Outcomes for a Demonstrably Effective Program" and "The Job Corps: Congressional Uses of Evaluation Findings." In Joseph S. Wholey, Mark A. Abramson, and Christopher Bellavita, Eds., *Performance and Credibility: Developing Excellence in Public and Nonprofit Organizations* (Lexington, Mass. D.C. Heath and Company, 1986), pp. 271-283 and 245-255.
- United States Senate, Committee on Appropriations, Senate Report 92-247, "Departments of Labor and Health, Education, and Welfare and Related Agencies Appropriations Bill, 1980," July 13, 1979.