



An Evaluation Roadmap for a More Effective Government

**Prepared by
AEA Evaluation Policy Task Force
February, 2009**



The American Evaluation Association is an international professional association of evaluators devoted to the application and exploration of program evaluation, personnel evaluation, technology, and many other forms of evaluation. Evaluation involves assessing the strengths and weaknesses of programs, policies, personnel, products, and organizations to improve their effectiveness. AEA has over 5500 members representing all 50 states in the US as well as over 75 foreign countries.

Mission: *To improve evaluation practices and methods
Increase evaluation use
Promote evaluation as a profession and
Support the contribution of evaluation to the generation of theory
and knowledge about effective human action.*

Evaluation Policy Task Force

The goal of the American Evaluation Association's Evaluation Policy Task Force is to promote evaluation policies that are critically important to the practice of evaluation.

The members of the Task Force are:

- William Trochim, Chair
- Eleanor Chelimsky
- Leslie Cooksy
- Katherine Dawes
- Patrick Grasso
- Susan Kistler
- Mel Mark
- Stephanie Shipman
- George Grob, Consultant

The term "evaluation policy" encompasses a wide range of potential topics that include (but are not limited to): when systematic evaluation gets employed, and on what programs, policies and practices; how evaluators are identified and selected; the relationship of evaluators to what is being evaluated; the timing, planning, budgeting and funding, contracting, implementation, methods and approaches, reporting, use and dissemination of evaluations; and, the relationship of evaluation policies to existing or prospective professional standards. To deal with the broad potential scope of this effort, the Task Force will, during its first two years, concentrate on evaluation policies in the United States Federal government, in both the legislative and executive branches.

An Evaluation Roadmap for a More Effective Government

The question we ask today is not whether our government is too big or too small, but whether it works - whether it helps families find jobs at a decent wage, care they can afford, a retirement that is dignified. Where the answer is yes, we intend to move forward. Where the answer is no, programs will end. And those of us who manage the public's dollars will be held to account - to spend wisely, reform bad habits, and do our business in the light of day - because only then can we restore the vital trust between a people and their government.

*President Barack Obama
Inaugural Address
January 20, 2009*

The Challenge

The United States faces a national debt in the trillions of dollars, annual deficits in the hundreds of billions, and uncertainties about financial institutions and the economy. At the same time, significant concerns remain about national security, health care, education, energy development, and many other facets of American life. Underlying all of these problems lies the central overarching challenge of how to address these problems effectively, how to determine what works and what doesn't.

Why Program Evaluation Is Essential

The effectiveness of Federal programs depends on timely and accurate feedback and analysis that assesses what works, how it works, and why. Without such feedback, Federal agencies are essentially driving while blindfolded, unable to sense accurately whether they are heading in the wrong direction. Ineffective programs get promulgated, resources get squandered, and learning and progress cannot occur.

Program evaluation can provide the needed feedback function in the management and study of any program. Evaluation uses systematic data collection and analysis to address questions about how well government programs and policies are working, whether they are achieving their objectives, and perhaps most important, why they are effective or not. It provides evidence that can be used to compare alternative programs, guide program development and decision-making, and reveal effective practices. By its very nature, program evaluation provides the publicly accessible evidence that is at the heart of greater transparency.

Since the inception of modern program evaluation in the era of the Great Society programs of the 1960s, the U.S. Federal Government has failed to achieve a coordinated and coherent evaluation capability as an integral part of Federal program management. Many program evaluations have been conducted, to be sure, and their results applied to make reasoned program decisions. But for the most part evaluations have been sporadic, inconsistently applied, and inadequately supported.

Federal agencies are often reluctant to undertake evaluation and the units that are formed to accomplish it too often are short-lived and under-resourced. Training and capacity building for evaluation is inconsistent across agencies and insufficient to raise the level of evaluation to what is needed. In the past eight years, the Office of Management and Budget attempted to institute consistent evaluation requirements across all Federal programs through its Program Assessment Rating Tool (PART) program. However, that effort, while a step in the right direction, was not adequately resourced, was inconsistently applied, was too narrow relative to the options that are suitable across the life cycle and circumstances of different programs, and did not provide the high-quality training and support for agencies to accomplish evaluation effectively. While significant advances in the use of evaluation have occurred in the Federal Government since the 1960s, the commitment needed to consistently ensure that decisions are informed by evaluation has not yet been made.

The Obama administration has a unique opportunity to advance its broad policy agenda by integrating program evaluation as a central component of Federal program management.

The time is especially right for such a bold move. The breadth and seriousness of the challenges we face provide a political climate that could support a commitment to a major advance in Federal program management. The lessons that have been learned in those agencies that have experience in applying evaluation constitute a solid knowledge base upon which we can build. And, the field of evaluation has evolved to a point where it is more capable than ever before to support a significant expansion in the scope of Federal efforts.

The new administration would benefit significantly by using program evaluation to

- address questions about current and emerging problems
- reduce waste and enhance efficiency
- increase accountability and transparency
- monitor program performance
- improve programs and policies in a systematic manner
- support major decisions about program reform, expansion, or termination
- assess whether existing programs are still needed or effective
- identify program implementation and outcome failures
- inform the development of new programs where needed
- share information about effective practices across government programs and agencies

The key is to make program evaluation integral to managing government programs at all stages, from planning and initial development through start up, ongoing implementation, appropriations, and reauthorization. In short, what is needed is a transformation of the Federal management culture to one that incorporates evaluation as an essential management function.

Recommendations

We recommend that each Federal agency adopt the following framework to guide the development and implementation of its evaluation programs:

Scope and Coverage

- Invest in a body of evaluative work that covers public programs and policies throughout their life cycle and utilizes evaluation as a tool for improving programs and assessing their effectiveness
- Evaluate Federal programs and policies in a manner that is appropriate for program stewardship and useful for decision-making
- Build into each new program and major policy initiative an appropriate framework to guide the conduct of evaluations throughout the life of the program or initiative
- For existing programs, assess what is already known and develop evaluation plans to support future decision making

Management

- Assign senior, experienced officials to administer evaluation centers or coordinate evaluation functions at appropriately high levels of government agencies
- Prepare annual and long term evaluation plans to guide decision-making about programs
- Provide sufficient and stable sources of funds to support professional evaluation activities
- Coordinate and communicate about evaluation efforts across agencies with overlapping or complementary missions
- Develop written evaluation policies across and within Federal agencies that can guide evaluation efforts and help assure quality
- Assure that evaluation units and staff receive high-level, public and consistent support for their evaluation efforts

Quality and Independence

- Develop and adopt quality standards to guide the evaluation functions
- Promote the use and further development of a robust set of appropriate methods for designing programs and policies, monitoring program performance, improving program operations, and assessing program effectiveness and cost
- Safeguard the independence of the evaluation function with respect to the design, conduct, and results of evaluations
- Preserve and promote objectivity in examining program operations and impact

Transparency

- Consult closely with Congress and non-Federal stakeholders in defining program and policy objectives, operations to be assessed, and definitions of success
- Disseminate evaluation findings relating to public accountability to policy makers, program managers, and the public at large
- Create clearinghouses for sharing information about effective and ineffective program practices

In this paper we develop these concepts more fully. First, we describe the general principles that should guide a government-wide effort to strengthen evaluation functions. Then we propose some broad administrative steps that can be taken to institutionalize evaluation in Federal agencies. Finally, we discuss how the Executive Branch and the Congress can collaborate in making the most effective and efficient use of evaluation as a staple of good government.

General Principles

Following are some general principles to guide an initiative to integrate evaluation with program management.

Scope. Evaluation should span the life cycle of programs and policies, making it integral to planning, developing, and managing government programs at all stages of their development. Evaluation activities should be used to:

- Make sure that program designs make sense
- Identify problems encountered during start up and correct them before they become permanent features of programs
- Identify and share promising approaches that emerge during program implementation
- Assess the extent to which programs and policies are being implemented as intended
- To the extent feasible, establish expectations and performance standards at the inception of the program; involve stakeholders to refine them as programs mature
- Develop appropriate and efficient data collection and reporting systems and information technology support to provide a continuing flow of evaluative information to policy makers and program managers
- Examine the extent to which programs are having the reach they are expected to have
- Periodically examine selected program features to improve their effectiveness and efficiency
- Periodically assess program results and service quality
- Examine systematically whether an apparently successful program can be transferred to another setting before making the considerable investment in scaling it up

Coverage. In general, Federal programs and policies should be subject to evaluation. Evaluation is a responsible and necessary part of good management.

Analytic Approaches and Methods. The choice of analytic approaches and methods depends on the questions being addressed, the kind of program being evaluated, the status of its implementation, when the evaluation results are needed, what they are needed for, and the intended audience.

There are no simple answers to questions about how well programs work, and there is no single analytic approach or method that can decipher the complexities that are inherent within the program environment and assess the ultimate value of public programs. Furthermore, definitions of what constitutes “success” may be difficult or contested. A range of analytic methods is needed, and often it is preferable to use several methods simultaneously, including both quantitative and qualitative approaches. Some evaluation approaches are particularly helpful in the early developmental stages of a program, whereas others are more suited to situations in which the program has become more routinized and regularly implemented. The broader policy and decision making context also can influence the approach that is most appropriate. Sometimes information is needed quickly, requiring studies that can either use existing data or rapid methods of data collection, while at other times more sophisticated long-term studies are required to understand fully the dynamics of program administration and beneficiary behaviors.

Over the years, the evaluation field has developed an extensive array of analytic approaches and methods that can be applied and adapted to various types of programs, depending upon the circumstances and stages of the program's implementation. For example, surveys are among the bedrock tools for evaluation. But there are many ways in which they can be used, and this method, just like all the others, has evolved to address new and emerging policy interests. The fairly recent use of web surveys is a case in point.

Fundamentally, all evaluation methods should be context-sensitive, culturally relevant, and methodologically sound. A complete set of evaluation approaches and methods would include but not be limited to:

- case studies
- surveys
- quasi-experimental designs
- randomized field experiments
- cost-benefit and cost-effectiveness analyses
- needs assessments
- early implementation reviews
- logic models and evaluability assessments

Resources. Sufficient resources should be made available for evaluation, including stable annually recurring sources of funds and special one time funds for evaluation projects of interest to Executive Branch and congressional policy makers. The stable recurring evaluation funds, where they exist, should generally be made available through specific appropriations or set-asides of program funds to be pooled for evaluation activities. These methods can also be combined to support viable evaluation programs. The overall approach should be to authorize and require periodic evaluations of each program throughout its life so that a rich source of evaluative information is available and will be known to policy makers during the annual appropriations and cyclical reauthorizations and amendments that are typical of public programs today.

The cost of maintaining and using large cross cutting surveys and performance monitoring data bases that are available to support policy analysis, evaluation, management, and research should be funded apart from the stable evaluation budget.

Professional Competence. Evaluations should be performed by professionals with evaluation training and experience appropriate to the evaluation activity in which they are engaged (such as performing studies, planning evaluation agendas, reviewing evaluation results, performing statistical analyses.) Evaluation is an interdisciplinary field that encompasses many areas of expertise. Many evaluators have advanced degrees in, and often work collaboratively with colleagues in allied fields, such as economics, political science, applied social research, sociology, psychology, policy analysis, statistics, and operations research, to name just a few.

Federal agencies should be encouraged to recognize the multi-disciplinary nature of evaluation and assure that the diversity of disciplines is appropriately represented in both internal and independent consulting teams.

Evaluation Plans. Each Federal agency should require its major program components to prepare annual and multi-year evaluation plans of the studies and other evaluation activities that they will undertake. The plans should be updated annually. The planning should take into account the needs of evaluation results for informing program budgeting, reauthorization, agency strategic plans, ongoing program development and management and responses to critical issues that arise concerning program effectiveness, efficiency, and waste. These plans should include an appropriate mix of short and long term studies to ensure that evaluation results of appropriate scope and rigor are available when short or long term policy or management decisions must be made. To the extent practical, the plans should be developed in consultation with program stakeholders who are involved in or affected by the programs.

Evaluation questions can spring up unexpectedly and urgently, emerging from, say, a changed political or social context, and or a sudden need for information to support a Presidential initiative or to respond to questions raised by the Congress. Therefore evaluation plans should leave room for these contingencies by setting priorities that allow for some flexibility in the scheduling of evaluations.

Dissemination of Evaluation Results. The results of all evaluations related to public accountability should (except where this is inconsistent with the Freedom of Information Act or Privacy Act) be made available publicly and in a timely manner. They should be easily accessible through the internet with user friendly search and retrieval technologies. Similarly, evaluations of promising and effective program practices should be systematically and broadly disseminated to potential users in all Federal agencies and the public. Evaluation data should to the extent feasible and with sufficient protections for privacy be made available to the public and professional community to enable secondary analysis and encourage transparency.

Evaluation Policy and Procedures. Each Federal agency and its evaluation centers or evaluation coordinators (discussed below) should publish policies and procedures and adopt quality standards to guide the conduct of evaluations within its purview. Such policies and procedures should identify the kinds of evaluations to be performed, and the criteria and administrative steps for: developing evaluation plans and setting priorities; selecting the appropriate evaluation approaches and methods to be used; consulting subject matter experts; ensuring the quality of the evaluation products; publishing evaluation reports; ensuring the independence of the evaluation function; using an appropriate mix of staff and outside consultants and contractors; and promoting the professional development of evaluation staff.

Independence. While the heads of Federal agencies and their component organizations should participate in setting evaluation agendas, budgets, schedules, and priorities, evaluation managers should be independent with respect to the design, conduct, and results of their evaluation studies.

As noted above under “Resources,” we recognize that the establishment of performance indicators, along with the tracking and reporting of program results through such performance indicators, is a part of evaluation that is properly the function of program administrators. Nevertheless, evaluators may be called upon to provide independent advice on performance measurement, and in some cases to administer complex performance measurement systems on behalf of management.

Institutionalizing Evaluation

Significant progress has been made in gradually establishing evaluation as an integral component of the management of government programs. However, additional steps are needed to take advantage of evaluation as a management tool.

Current Status of GPRA and PART

The most significant evaluation-related initiatives of the last 15 years have been the enactment of the Government Performance and Results Act of 1993 (GPRA) and, more recently during the George W. Bush administration, OMB's Program Assessment Rating Tool (PART).

Generally, GPRA encourages a strategic, agency wide, mission view and also focuses on whether government programs achieve results in terms of the goals and objectives for which they are established. Evaluation was defined in GPRA as addressing the "manner and extent to which" agencies achieve their goals, thus addressing both implementation and results. In practice, it has been implemented in a way that emphasizes the use of performance indicators and measurement to see whether a goal has been reached or not, with less attention being paid to evaluation studies that might shed light on the role the program played in reaching the goal, on why programs do or do not meet their goals and objectives, and on how programs might be improved. As a result, there is less information through this process that can guide programmatic or policy action.

PART focuses on programs' effectiveness and efficiency, especially on their impact. It draws on GPRA in terms of its analysis of whether programs meet their performance goals. However, it recognizes that some programs can meet their goals and still fail to have meaningful impact because of shortcomings in their designs or their goals. It attempts to assess whether programs are free from design flaws that prevent them from being effective. It introduces evaluation, and even calls for a body of independent evaluations for programs. For the most part, however, it emphasizes the use of evaluation as a way to determine program impact. While possibly not intended, it has had the effect of over-advocating for one particular type of impact evaluation, namely, randomized controlled trials, as a "gold standard" for the measurement of program success. This has tended to limit its ability to recognize success in programs for which randomized controlled trials are not a suitable method for assessing effectiveness or improving performance.

Some distrust PART results because they believe the goals and objectives upon which its analyses are based may be driven by political ideologies. In particular, Congress has distanced itself from PART. Some have noted that PART excludes policies like tax expenditures and focuses on discrete programs when multiple activities that cut across agency boundaries may contribute to achievement of goals.

OMB has moved to address some of these perceived shortcomings by initiating a pilot test of alternative impact assessment methodologies. In addition, in November 2007 President Bush signed an Executive Order on Improving Government Program Performance. It creates the position of Performance Improvement Officer in each Federal agency and establishes a

government wide Performance Improvement Council under the direction of OMB to more systematically promote performance assessments of programs.

GPRA and PART have made the use of performance measurement and management as a staple of government program management. But they fall considerably short of what is needed to address the problems our country faces.

Going Beyond GPRA and PART

What we are proposing is a vision for transforming the view of what agency heads and the Congress can do to benefit from program evaluation and systematic analysis both to improve the design, implementation, and effectiveness of programs and to assess what works and what doesn't, and why. This vision is a comprehensive one that recognizes that evaluation is more than simply "looking in the rearview mirror" and that it needs to be utilized earlier in the life of a program, as an integral part of managing government programs at all stages, from initial development through start up, ongoing implementation, appropriations, and reauthorization.

For this approach to work, action will be needed by both the Executive Branch and the Congress.

In the Executive Branch

The infrastructure and practice of program evaluation in Federal agencies is somewhat of a mixed story. Some agencies have well developed stable evaluation offices; others do not. The same can be said for evaluation funding, scope, policies, planning, and dissemination.

There are considerable differences in the nature and size of programs and agencies, in the kinds of evaluation information needed by program managers, and in the maturity of agency evaluation programs. Under various circumstances, the evaluation function itself may be broadly or narrowly defined. In addition, the evaluation function may appropriately be associated organizationally with various related functions such as management, planning, research, and policy development, including legislative or regulatory development. For example, it is not uncommon to find offices of planning and evaluation, research and evaluation, and monitoring and evaluation. Some Inspectors General have offices of inspections and evaluations. Sometimes the evaluation function is highly centralized in an agency or within a large program area of an agency. Elsewhere, the evaluation function is scattered throughout an agency in small offices connected with individual programs or small groups of them.

There is no single best practice for organizing evaluation offices and functions. All of the arrangements described above have emerged, for one reason or another, in response to circumstances related to such factors as substantive area, kind of agency, or type of evaluation focus. They may or may not be the most effective models to use under current circumstances. Whatever model is chosen, the functions and attributes described above in the section on general principles (coverage, scope, methods, resources, competence, planning, dissemination of public accountability results, policy and procedures, and independence) need to be established and nourished regardless of the organizational structure that is employed.

Based on the general principles discussed in the previous section, for the Executive Branch we propose two organizational frameworks as desirable for supporting the practice of evaluation within Federal agencies:

Centers of Evaluation. Agencies could establish one or more Centers of Evaluation to provide stable organizational frameworks for planning and conducting evaluations. The organization of such centers should be such that every program of the agency falls within the purview of one of the centers with respect to program evaluation. The heads of these evaluation centers should report directly to the senior executive of the organizational component in which they reside. Each of these centers would:

- have stable budgets sufficient to plan and carry out an appropriate level of program evaluation over several years
- strategically plan a body of evaluation work for the agency, each operational component, or program group for which it has evaluation responsibility
- hire professional evaluators or engage consultants or contractors with the diversity of skills necessary to plan and execute (or procure) independent evaluation studies
- publish the results of evaluations related to public accountability of the programs within their jurisdictions
- consult with agency program and budget offices, and, in concert with the agency's legislative liaison office, with the Congress in the development of evaluation plans
- issue policies and procedures to guide the conduct of evaluation work within its purview, including guidance on appropriate methods for conducting both formative and summative evaluations
- share information about effective programs and evaluation methods across government agencies
- promote and facilitate the ongoing training and professional development of evaluators within the center

Evaluation Coordinators. Agencies that choose to scatter their evaluation offices, associating small evaluation offices with individual programs or small collections of programs, should appoint one or more senior officials who will be responsible for coordinating the strategic planning of evaluation activities across the agency or across larger agency components or groups of programs. Such officials need not necessarily supervise the production of evaluations. Their role would be one of coordination, advice, and facilitation. Evaluation coordinators would:

- advise the agency head or senior officials on matters pertaining to evaluation
- ensure that each program operating division or major program group of the agency has a current annual evaluation plan
- promote, facilitate, and coordinate the development of evaluation plans for programmatic issues that cut across agency lines
- facilitate the preparation of evaluation budgets
- establish appropriate standards, frameworks, and procedures for evaluation activities within the agency
- facilitate the development and efficient and effective production of evaluation products
- facilitate the dissemination of evaluation reports related to public accountability

- share information about effective programs and evaluation methods
- promote and facilitate the ongoing training and professional development of evaluators within the agency

Combined Approach. Federal agencies may find it advantageous to use both approaches—evaluation centers for large programs, program groupings, and overall evaluation support and evaluation coordinators for scattered evaluation offices—to ensure the viability of the evaluation function.

In the Congress

Evaluation and analysis for the Congress is carried out through the Government Accountability Office (GAO), the Congressional Research Service (CRS), the Congressional Budget Office (CBO), the National Academies, temporary commissions, and statutory requirements of agencies. To a lesser extent, studies or investigations are conducted under the auspices of congressional committees or subcommittees. We are not proposing to change these organizational structures. Instead, we focus on strengthening the connection between evaluations and the laws that Congresses passes. This can be done by building evaluation expectations into authorizing legislation and making sure that adequate resources are explicitly set aside for evaluation.

Authorizing legislation. The authorization of new programs and periodic reauthorization of existing programs provide opportunities for Congress to establish frameworks for systematic evaluation of new and continuing programs. Congressional committees can, through authorizing legislation, provide guidance on or stipulate such things as:

- Early implementation reviews to identify management start up problems, such as scheduling, contracting, and grant making issues and to correct them before things get out of hand
- Requirements for the development of evaluation plans
- Evaluation of promising approaches to be shared among program implementers
- Development of performance indicators and the means to collect meaningful data on them once the program gets going
- Studies assessing program effects and identifying why programs are or are not effective
- Specific topics of interest to the Congress to be evaluated, with the results to be reported to the Congress for its oversight functions, appropriations, and during the next reauthorization of the program
- The establishment, expansion, or amendment of ongoing major surveys or other data collection mechanisms that the Congress wishes become a permanent continuing source of reliable data well into the future of the program
- Establishment of evaluation centers or evaluation coordinators as described above
- Funding for evaluation activities

Collaboration Between Executive Branch and the Congress

Evaluation results may be more useful if the Congress and the Executive Branch work together to include broad evaluation expectations and concerns in authorizing statutes and appropriations.

This collaboration is not always easy to achieve, and under some political circumstances, may even be impossible. Nevertheless, when this has happened in the past, it has been highly productive, both for fine-tuning evaluation designs and for the eventual policy or program use of the evaluation outcomes.

Federal Evaluation in the 21st Century

The fiscal and other challenges that our nation and the world faces in the coming century are staggering. Addressing them will require a Federal evaluation initiative commensurate to the level of challenge that we face. Program evaluation is essential both for prospective planning of programs and for retrospective assessment of effectiveness. Evaluation is the systematic, public, and transparent process of providing feedback to the Federal system. The Obama administration has a unique historical opportunity to transform the government through the integration of evaluation as an inherent part of the thoughtful and responsible management and oversight of programs. A significant commitment to this transformation can be a major legacy of this administration. And, it is essential both for achieving the policy and program goals of this administration and for passing onto future generations a more coherent and effective system to enable our society to learn what works, to address problems well, and to advance the dream of a better world.