The Institute of Brazilian Issues – IBI

The Minerva Program - Fall 2010

Performance Measurement of US Federal Programs: The Use of "The Program Assessment Rating Tool"

Author: Alzira Ester Angeli

Advisor: Christopher Yukins

Washington, D.C. – December 2010

Table of Contents:

Performance Measurement of US Federal Programs: The Use of "The Program Assessment Rating Tool"

I. Overview

The concept of government accountability in past decades has focused only on accountability for complying with laws, including the proper use of funds, so that public resources are used strictly for legal purposes. However, in recent years, an additional focus has been on accountability for results.

Indeed, during the last three decades of the 20th century a number of managers at all levels of government in many countries around the world developed and implemented performance measurement systems to address demands that they document government performance.

Demands for data documenting program performance have been originated mainly from citizens, reflecting their widespread desires for more transparent, entrepreneurial, and efficient government. Citizens urged to check what they receive back for the taxes paid, in terms of public services and goods delivered to the society.

The increased demand for data documenting the performance of public programs and agencies is in part a sign of the information age and obviously information technology has facilitated the processing of unprecedented amounts of program data more efficiently than ever before (Newcomer, 1997).

Faced with those ardent demands, a large number of local governments in the United States started measuring performance in the mid-1970s. Many state and local governments now require their departments and agencies to prepare performance measures. The requirements may originate in statutes, ordinances, rules, or executive directives. Often, performance measures are required in budget submissions.

The use of performance evaluation as a management tool fosters many local government managers realize that performance measures are also useful to inform decision making in a wide variety of administrative arenas, including strategic planning, budgeting, and contract monitoring. Within the federal executive branch of the United States government, evaluation takes place in order to avoid waste and inefficiency in public policies and programs, which undermine citizens' confidence in government and reduces its ability to address adequately vital public needs.

On the other hand, it is important to highlight that implementing and using performance measurement systems require much more of managers than technical measurement skills. In fact, experience in many government jurisdictions in the United States and abroad has shown that the effective use of performance measurement to improve program management presents managers with complex communication, analytical, political, and measurement challenges (Newcomer, 2003).

A. Performance measurement of US federal agencies and programs

Although governments in the United States have traditionally evaluated their performance at the federal, state and local levels, the implementation of performance measurement as a systematic management tool by the U.S. federal government took place later than in some other countries, such as United Kingdom, Canada, Australia and New Zealand.

While in those countries public programs and agencies were assessed every three or five years since the 1980's, the methodical use of performance evaluation was eventually adopted by the U.S. federal government in 1993, under The Government Performance and Results Act – GPRA.

In the late 1990s, managers of federal agencies in the United States were coping with demand and supply challenges regarding performance measures. Since then, their struggles have provided instructive lessons for public managers at other levels of the public sector.

1. Government Performance and Results Act of 1993

The Government Performance and Results Act of 1993 – GPRA is the Public Law 103-62, which was enacted to: hold federal agencies accountable for achieving program results; improve internal management, program effectiveness, public accountability, and delivery of services; and provide information on program effectiveness and efficiency.

The GRPA was issued in result of the findings of the Congress of the United States of America, according to which:

"(1) waste and inefficiency in Federal programs undermine the confidence of the American people in the Government and reduces the Federal Government's ability to address adequately vital public needs;

(2) Federal managers are seriously disadvantaged in their efforts to improve program efficiency and effectiveness, because of insufficient articulation of program goals and inadequate information on program performance; and

(3) congressional policymaking, spending decisions and program oversight are seriously handicapped by insufficient attention to program performance and results."

Therefore, the GRPA purposes are to:

"(1) improve the confidence of the American people in the capability of the Federal Government, by systematically holding Federal agencies accountable for achieving program results;

(2) initiate program performance reform with a series of pilot projects in setting program goals, measuring program performance against those goals, and reporting publicly on their progress;

(3) improve Federal program effectiveness and public accountability by promoting a new focus on results, service quality, and customer satisfaction;

(4) help Federal managers improve service delivery, by requiring that they plan for meeting program objectives and by providing them with information about program results and service quality;

(5) improve congressional decision making by providing more objective information on achieving statutory objectives, and on the relative effectiveness and efficiency of Federal programs and spending; and

(6) improve internal management of the Federal Government."

The GPRA established a coordinated framework for accountability of federal programs and agencies due to strategic planning and result evaluation. In this sense, the establishment of that framework was feasible as a consequence of the introduction of four integrated elements: strategic planning, performance plans and reports, managerial accountability and flexibility, and performance budgeting (budget addressed to results).

Together those components impose that annually every federal executive agency presents its strategic and operational plans, as well as its performance report, in order to demonstrate the quantified objectives and goals established; the performance indicators applied; the human, financial and technological resources necessary to accomplish with the results presented; and the comparison of these results against the expected ones.

In the last fifty years one of the major concerns related to public budget elaboration in the federal branch of the United States government has been the pursuit of feasible alternatives for the allocation of public scarce resources considering the results achieved by well-succeeded programs. First of all, the challenge was to measure the inputs and outputs. In the recent years, the efforts have been concentrated in measuring the outcomes (Joyce, 1997).

Indeed, the GPRA focus on the evaluation of effective outcomes. In this sense, one significant concern explicit in the GPRA is the dissemination of "Performance-Based Management" among the federal agencies, so that the adoption of performance indicators may help public managers along the decision process. A close analysis of the GRPA text demonstrates its pragmatism on this issue.

Under the GRPA, it is mandatory that every federal executive agency develops a five-year minimum period strategic plan for program activities, which is submitted to the Office of Management and Budget – OMB and to the Congress for approval. Moreover, each strategic plan has to be updated and revised at least every three years.

According to the GRPA, the strategic plan has to contain:
"(1) a comprehensive mission statement covering the major functions and operations of the agency;
(2) general goals and objectives, including outcome-related goals and objectives, for the major functions and operations of the agency;
(3) a description of how the goals and objectives are to be achieved, including a description of the operational processes, skills and technology, and the human, capital, information, and other resources required to meet those goals and objectives;
(4) a description of how the performance goals included in the plan required by 'section 1115(a) of title 31' shall be related to the general goals and objectives in the strategic plan;
(5) an identification of those key factors external to the agency and beyond its control that could significantly affect the achievement of the general goals and objectives; and
(6) a description of the program evaluations used in establishing or revising general goals and objectives, with a schedule for future program evaluations."

Besides the strategic plan, an annual performance plan is demanded by the GRPA, so that every agency is in charge to prepare once a year a performance plan covering each program activity set forth in its budget.

Also according to the GRPA, the performance plan has to:
"(1) establish performance goals to define the level of performance to be achieved by a program activity;
(2) express such goals in an objective, quantifiable, and measurable form unless authorized to be in an alternative form under 'subsection (b)';
(3) briefly describe the operational processes, skills and technology, and the human, capital, information, or other resources required to meet the performance goals;
(4) establish performance indicators to be used in measuring or assessing the relevant outputs, service levels, and outcomes of each program activity;
(5) provide a basis for comparing actual program results with the established performance goals; and
(6) describe the means to be used to verify and validate measured values."

At last, under the GRPA every federal executive agency has to prepare and submit to the President and the Congress, a report on program performance for the previous fiscal year, which sets forth the performance indicators established in the agency performance plan along with the actual program performance achieved compared with the performance goals expressed in the plan for that fiscal year.

In accordance with the GRPA, the performance report has to:
"(1) review the success of achieving the performance goals of the fiscal year;
(2) evaluate the performance plan for the current fiscal year relative to the performance achieved toward the performance goals in the fiscal year covered by the report;
(3) explain and describe, where a performance goal has not been met (including when a program activity's performance is determined not to have met the criteria of a successful program activity under 'section 1115(b)(1)(A)(ii)' or a corresponding level of achievement if another alternative form is used);
(a) why the goal was not met;
(b) those plans and schedules for achieving the established performance goal; and
(c) if the performance goal is impractical or infeasible, why that is the case and what action is recommended;
(4) describe the use and assess the effectiveness in achieving performance goals of any waiver under 'section 9703 of this title'; and
(5) include the summary findings of those program evaluations completed during the fiscal year covered by the report."

Considering the analysis of the United States Government Accountability Office on the GRPA (GAO, 1998), the following conclusions have emerged:
- The improvement of government actions depends fundamentally on the evaluation of their effects, which are the outcomes.
- Performance evaluation is a complex process, which often demands a long implementation period.
- A successful program is supported both by the executive and the legislative powers. This integration is essential for public accountability.

2. The Office of Management and Budget

Agency performance and program evaluation are overseen within the U.S. federal executive branch by The Office of Management and Budget – OMB, the largest component of the Executive Office of the President.

The core mission of The Office of Management and Budget – OMB is to serve the President, to whom it reports directly, as the implementation and enforcement arm of presidential policy government-wide. In order to accomplish with its mission, the OMB helps a wide range of executive departments and agencies within the federal branch, concentrating its efforts through five critical processes: budget, management, regulation, information policy, legislative information, as follows:

Budget — development and execution of the budget, a significant government-wide process by which the President implements decisions, policies, priorities, and actions in all areas.

Management — oversight of agency performance, federal procurement, financial management, and information technology.

Regulation — coordination and review of all relevant federal regulations issued by executive agencies, so that they reflect presidential priorities and ensure that economic and other impacts are assessed as part of regulatory decision-making.

Information Policy — address of Executive Orders and Presidential Memoranda to agency managers. Those are the mechanisms by which the President directs specific government-wide actions by executive branch officials.

Legislative Information — clearance and coordination of all agency communications with Congress, including testimony and draft bills, to ensure consistency of agency legislative views and proposals with presidential policies.

Concerning the management side of the OMB, its structure includes the following offices:

The Office of Federal Financial Management – OFFM, which provides strategic direction to: improve financial management, reporting, and systems; reduce improper payments; improve grants management; and "right-size" federal real property.

The Office of Federal Procurement Policy – OFPP, which is responsible for setting government-wide acquisition priorities and policies, that shape the full range of federal executive agency procurement practices. OFPP strengthens the acquisition workforce and fosters cost-saving and risk-reduction practices. OFPP also works to ensure that agencies rely on and manage contractors in fiscally responsible ways.

The Office of E-Government and Information Technology - *E-Gov*, which develops and provides direction in the use of internet-based technologies to streamline citizens and businesses interact with the federal government.

The Office of Performance and Personnel Management - OPPM, which drives mission-focused performance gains across the federal executive government. It is within the management side of the OMB that agency performance and program evaluation are overseen.

a. The Office of Performance and Personnel Management

The Office of Performance and Personnel Management - OPPM is the component of the OMB that coordinates the administration's goal-setting and performance review process for agencies' high priority performance goals and guides agency strategic and annual planning, performance reviews and performance reporting. Agencies' management of programs and resources are overseen to determine their net effects, success or failure, and how the agencies respond to these findings by making management improvements and developing new budget and policy proposals.

The OPPM also functions as the Resource Management Office for the U.S. Office of Personnel Management - OPM and guides federal personnel policy, working closely with the OPM to implement effective personnel policies and practices.

B. Key definitions and concepts

Performance Measurement

"Performance measurement is the label given to routine measurement of program inputs, outputs, and outcomes undertaken by agencies within governments and the nonprofit sector to meet demands for documentation of program performance" (Wholey and Hatry, 1992).

Based on that definition, performance measurement is an ongoing process that involves the selection, definition and application of performance indicators, which addressed to program activities and services delivered quantify their efficiency, efficacy and effectiveness based on the inputs, outputs and outcomes identified.

Therefore, performance measurement applies typically quantitative information to indicate to "what" extent the program is accomplishing with its objectives and goals and whether the expected results are being achieved. For that reason, performance data must be intimately linked to the mission, objectives and goals from the highest organizational level downward to the operational levels.

In fact, performance measurement is a powerful tool to improve management as it helps managers by providing them information on how resources and efforts should be allocated to ensure effectiveness. In this sense, program managers must have appropriate understanding of the indicators applied and how the activities performed affect each indicator's behavior.

Specifically in the public sector, not only performance measurement keeps program partners focused on the key goals of a government program, but also supports development and justification of budget proposals by indicating how taxpayers benefit.

Most performance measures can be grouped into one of the following six general categories even though certain organizations may develop their own categories as appropriate depending on their mission:

1. Effectiveness: A process characteristic indicating the degree to which the process output (work product) conforms to requirements (Are we doing the right things?)
2. Efficiency: A process characteristic indicating the degree to which the process produces the required output at minimum resource cost (Are we doing things right?)
3. Quality: The degree to which a product or service meets customer requirements and expectations.
4. Timeliness: Measures whether a unit of work was done correctly and on time. Criteria must be established to define what constitutes timeliness for a given unit of work. The criterion is usually based on customer requirements.
5. Productivity: The value added by the process divided by the value of the labor and capital consumed.
6. Safety: Measures the overall health of the organization and the working environment of its employees.

All types of indicators, as to be useful management tools, must have the attributes of any ideal unit of measure, which are:

- reflect the customer's needs,
- provide an agreed upon basis for decision making,
- apply broadly,
- are understandable,
- are interpreted uniformly,
- are compatible with existing sensors,
- are precise in interpreting the results,
- are economical to apply.

Hatry, Gerhart, and Marshall pointed that actually the following benefits are likely to be achieved as a result of using performance measurement within public agencies and programs (Hatry, Gerhart and Marshall, 1994):

- Foster dialogue to clarify program logic. Thinking about performance measures causes policy-makers, managers, and staff to ask "why are we doing this?" and to sometimes challenge traditional methods and assumptions. This benefit is often more evident when outsiders, less familiar with the program, participate along with insiders in formulating the measures.

- Focus policy discussion on results. Policy discussions are constrained by the type, quality, and amount of information available. Where good information on program performance is lacking, there is an unfortunate tendency for these discussions to rely on speculation and anecdotes and to orient toward process issues and implementation details. Good performance measures can lift the considerations of policy-makers to larger questions of program design, outcomes, and efficiency.

- Identify opportunities for improvement. If performance shortfalls are identified early, the agency can take timely corrective actions and evaluate the effect of the actions.

- Guide management in allocating resources. Good performance measurement can provide valuable input for budget and planning processes. It can assist policymakers and managers to judge where they may get the best, or at least better, returns on investment.

Program Evaluation

Wholey and Hatry define program evaluation as "the systematic assessment of the extent to which programs have achieved the intended results" (Wholey and Hatry, 1992). Upon this definition, program evaluations can and do address questions about "why" and "how" programs achieve or fail to achieve their established objectives.

According to the definition provided in the GPRA, program evaluation is the "assessment, through objective measurement and systematic analyses, of the manner and extent to which federal programs achieve intended objectives". Thus, under the GRPA program evaluation is clearly viewed as a complementary analytical tool to support performance measurement in government agencies and programs.

1. Performance Measurement x Program Evaluation

The European Commission defines performance measurement and program evaluation, respectively, as "a continual process carried out during the execution of the program with the intention of immediately correcting any deviation from operational objectives. Evaluation, on the other hand, is specifically conducted at a discrete point in time in the life cycle of a program, and consists of an in-depth study" (European Commission, 1997).

But, "how does performance measurement relate to program evaluation?". Newcomer states that program evaluation is clearly the more inclusive of the two approaches to reviewing programs in the public sector. Program evaluation is retrospective, collecting and analyzing information on existing programs. Its specific evaluation strategy can also assess the compliance of programs` implementation with statutory guidance as well as cost-benefit analyses. There is great variety in the types of evaluation questions raised by program management and oversight bodies, ranging from problem-based investigations to evaluations of the net impact of programs (Newcomer, 1997).

Although program evaluation, like performance measurement, may also focus on programs inputs, operations, and results, evaluators often 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?

On the other hand, performance measurement typically captures quantitative indicators, which show "what" is occurring with regard to program outputs and perhaps outcomes but, in itself, will not address the "why" and "how" questions. In fact, in many circumstances budgetary oversight bodies may not be interested in the "why" and "how" questions; they just want to know "how many" or "how much" has been delivered. However, if internal managers are committed to improve program operations, they must venture beyond the performance data.

In short, both forms of assessment, performance measurement and program evaluation, aim to support resource allocation and other policy decisions, but performance measurement surely provides earlier warnings to managers due to its ongoing nature.

Concerning practical matters, Julnes and Holzer argue that the application of performance indicators in public agencies and programs faces several dilemmas. The first dilemma is the difficulty in defining what are the outputs and the outcomes. The second one is the difficulty in measuring the results achieved: are they originated only in consequence of the program`s activities or they were also influenced by other factors like external non-expected ones? Besides, a third dilemma emerges: the fragmentation and overlap of government actions hinder the establishment of each agency organizational mission, objectives and goals, related to the specific programs it is accountable for (Julnes and Holzer, 2001).

Consequently, implementing and using performance measurement systems require much more of public managers than technical measurement skills. Considering those dilemmas, Newcomer highlights that the use of performance measurement to improve government programs' effectiveness presents managers with "complex communication, analytical, political, and measurement challenges" (Newcomer, 2003).

Relating to the first challenge considered by Newcomer, public managers responsible for the performance measurement system have to communicate clearly and frequently with all stakeholders involved in the processes, no matter who they are: other public managers, non-government managers, or the beneficiaries themselves. For instance, to keep the system aligned with the strategic objectives, program managers have to communicate constantly with the organizations' high administration.

"The analytical capacity to map program logic accurately and to conceptualize appropriate outputs or outcomes to measure is a second fundamental challenge for those charged with measuring performance of public programs" (Newcomer 2003). Hence, not only with a systemic analysis of the organization, focusing on its mission, objectives and goals, but also with the definition of the best fitted indicators, is it possible to identify what has to be measured and how to measure it.

Newcomer poses the following about the political aspect of public performance measurement: "performance measurement efforts will only be successful if there is sufficient political capital to involve the pertinent stakeholders, and to convince the pertinent political leaders that the performance data should be used to guide managerial decision making. Political capital is also needed to ensure the ongoing resource commitment for support and maintenance of performance information systems" (Newcomer, 2003).

Politicians may not consider that performance measurements are necessary as long as outcomes are routinely reported. Besides it, information provided by performance measurement is only part of the set of information that public managers and policy officials need to improve their decision process. For that reason, performance measurement often has to be combined with evaluation data to precisely demonstrate why results occur and what value a program adds. Indeed, as the measures by themselves seldom provide definitive answers, a major purpose of performance measurement is to raise fundamental questions and, due to its nature, it cannot replace data on program costs, political judgments about priorities, and innovative solutions.

Finally, Newcomer argues on the challenge of the measurement itself: “the ability to design and use relevant and sufficient measures to capture program performance depends on analytical capacity but extends further. Staff responsible for performance measurement must also know how to verify the accuracy and reliability of performance data” (Newcomer, 2003). The concern is that those in charge of selecting performance measures do not adopt only measures that are accessible and relatively inexpensive disregarding how their collection may affect the information gathered.

2. Common performance measures

In the Government Performance and Results Act of 1993 – GRPA the following definitions are presented:

- Outcome Measure: means an assessment of the results of a program activity compared to its intended purpose.

- Output Measure: means the tabulation, calculation, or recording of activity or effort and can be expressed in a quantitative or qualitative manner.

- Performance Goal: means a target level of performance expressed as a tangible, measurable objective, against which actual achievement can be compared, including a goal expressed as a quantitative standard, value, or rate;

- Performance Indicator: means a particular value or characteristic used to measure output or outcome;

- Program Activity: means a specific activity or project as listed in the program and financing schedules of the annual budget of the United States Government; and

- Program Evaluation: means an assessment, through objective measurement and systematic analysis, of the manner and extent to which Federal programs achieve intended objectives.

Each program adopts outcome, output, and efficiency measures and uses them to track their progress towards goals. Taken together, a program's measures should tell a comprehensive story of its performance. In this sense, as important as the definition of the “performance indicators” to measure the “outputs” and the “outcomes” is the accurate establishment of the “performance goals” to be pursuit, otherwise the figures may lead the managers to a distorted diagnosis.

a. output indicators

An output refers to the *internal* activities of a program, such as the products or services delivered. Thus, output indicators provide quantitative information regarding to materials, policies, procedures and legislation developed. Outputs can include process measures (paper flow, adjudication), attribute measures (timeliness, accuracy, customer satisfaction), and measures of efficiency.

An output answers the question "what does the program do to achieve its goal or purpose?" Exemplifying, a job training program may provide classes to teach someone the skills necessary to find a job. An output measure in this case may be the number of people who complete a job training program.

In general, programs are more likely to be assessed against outputs rather than outcomes. This is because output data is collected and reported more frequently, and outputs more typically correspond to activities and functions being directly controlled, as opposed to focusing on results. Nevertheless, outputs should help track a program's progress toward reaching its outcomes.

b. outcome indicators

An outcome refers to the events or conditions of direct importance to the beneficiary that are *external* to the program. Outcomes indicators measure short and/or long-term program effectiveness, which considers, relating to the program's objectives, changes in beneficiaries' conditions in consequence of the benefits obtained.

An outcome answers the question "what is the program's goal or purpose?" An outcome measure in the example taken, a job training program, may be the number and percent of people employed within six months of completing the job training program or how much their income increased.

Public managers are strongly encouraged to the use of outcome measures because they are much more meaningful to the public than outputs, which tend to be more process-oriented or means to an end. Outcomes may relate to the society as a whole or to the specific beneficiaries of public programs, depending on the complexity and range of the program assessed.

c. efficiency indicators

Efficiency measures capture a program's ability to carry out its activities and achieve results, outcomes and/or outputs, regarding the resources allocated. Performance measures may include the consideration of inputs, resources used to produce outputs and outcomes, particularly in the context of cost-efficiency or unit costs. Indeed, a very useful efficiency indicator is the one that addresses the cost of achieving a unit of outcome.

Managers are stimulated to consider the most meaningful level of input measures. For example, cost-efficiency measures based on outputs per dollar will typically be more useful than measures of output per unit of personnel. Similarly, social costs may be more meaningful than federal budget costs when evaluating effectiveness of regulatory programs. Inputs from state and local partners may be relevant in assessing the effectiveness of some programs matched by federal assistance.

C. Impact evaluation

Impact evaluation is known as the systematic identification of the effects, positive or negative, intended or not, on individual households, institutions, and the environment caused by a given development activity such as a program or project (World Bank Independent Evaluation Group).

It is an important component of the armory of evaluation tools and approaches and integral to global efforts to improve the effectiveness of aid delivery and public spending more generally in improving living standards. Originally more oriented towards evaluation of social sector programs in developing countries, notably conditional cash transfers, impact evaluation is now being increasingly applied in other areas such as the agriculture, energy and transport.

According to the World Bank Independent Evaluation Group - IEG, the term “impact evaluation” has taken various meanings over time, mainly during the last twenty years, and continues to be used in different approaches. The following have been the most common meanings:

- An evaluation which looks at the impact of an intervention on final welfare outcomes, rather than only at project outputs, or a process evaluation which focuses on implementation;
- An evaluation concerned with establishing the counterfactual, i.e. the difference the project made - how indicators behaved with the project compared to how they would have been without it;
- An evaluation carried out a period of time, in general five to ten years, after the intervention has been completed so as to allow time for impact to appear; and
- An evaluation considering all interventions within a given sector or geographical area.

The World Bank IEG states that these four definitions are not mutually exclusive. In fact, many IEG impact evaluations in the 90s were carried out five or more years after the intervention closed and also tried to establish a counterfactual. But nor need they coincide. Participatory impact evaluation follows the first definition, reporting beneficiary perspectives on how the intervention changed their lives, but makes no formal reference to a counterfactual.

IEG's current approach combines the first two definitions, under what impact evaluation is an assessment of the impact of an intervention on final welfare outcomes. It is noticed that this particular definition used by the World Bank IEG – counterfactual analysis of final welfare outcomes – is one which implies the need for a rigorous approach to establishing the counterfactual.

1. Objectives

In contrast to outcome monitoring, which examines whether targets have been achieved, impact evaluation is structured to answer the question: "how would outcomes such as beneficiaries' well-being have changed if the intervention had not been undertaken?". This involves counterfactual analysis, that is, "a comparison between what actually happened and what would have happened in the absence of the intervention" (World Bank Independent Evaluation Group). Impact evaluation aims to answer key questions for evidence-based policy making: "what works?", "what doesn't?", "where?", "why?" and "for how much?".

2. Approaches

Impact evaluation, even though being only one among a number of evaluation tools, has received increasing attention due to public policy making in recent years. According to the World Bank IEG, this is a tendency strengthened in both Western and developing country contexts. The results agenda has forced public agencies to demonstrate that the money they spend is effectively contributing to the improvement of the taxpayers` standard of living, thereby increasing demand for impact evaluation. In the current environment, calls for increased aid spending are only credible if it can be shown that current spending is actually contributing toward the attainment of the established objectives and goals.

Impact evaluation is approached in various ways. The current emphasis on rigor is important but should not lead to the neglect of aspects of evaluation design which ensure its policy relevance. Adopting a theory-based approach is the best way of ensuring such relevance, since it will yield information on how the program is working not just if it is working. Application of such an approach implies a mixed-methods evaluation design, i.e. one that combines quantitative and qualitative data collection. The time required for proper survey design should not be under-estimated, and can provide an opportunity for the evaluation team to spend time in the field which will prove invaluable when it comes to writing a well-contextualized report.

3. Models

The most useful of the several methods or models of impact evaluation are:

- Rapid assessment or review, conducted *ex post.* This method can encompass a range of approaches to endeavor to assess impact, such as participatory methods, interviews, focus groups, case studies, an analysis of beneficiaries affected by the project, and available secondary data;

- *Ex-post* comparison of project beneficiaries with a control group. With this method, multivariate analysis may be used to control statistically for differences in attributes between the two groups — this is one way of estimating the counterfactual situation;

- *Quasi*-experimental design, involving the use of matched control and project *(beneficiary) groups.* This method involves the use of a "non-equivalent" control group to match as closely as possible the characteristics of the project population – either through propensity score matching or using a multivariate regression approach. This method often involves the use of large scale sample surveys, and sophisticated statistical analysis; and

- Experimental design (Randomized design). Under experimental evaluations the treatment and comparison groups are selected randomly and isolated both from the intervention, as well as any interventions which may affect the outcome of interest. These evaluation designs are referred to as randomized control trials. In experimental evaluations the comparison group is called a control group. When randomization is implemented over a sufficiently large sample with no contagion by the intervention, the only difference between treatment and control groups on average is that the latter does not receive the intervention. The experimental approach is often held up as the 'gold standard' of evaluation and it is the only evaluation design which can conclusively account for selection bias in demonstrating a causal relationship between intervention and outcomes.

- Non-experimental design. This method is so-called because it does not involve a comparison group which does not have access to the intervention. The method used in non-experimental evaluation is to compare intervention groups before and after implementation of the intervention. Non-experimental designs are the weakest evaluation design, because in order to show a causal relationship between intervention and outcomes convincingly, the evaluation must demonstrate that any likely alternate explanations for the outcomes are irrelevant. However, there may be cases where non-experimental designs are the only feasible impact evaluation design, such as universally-implemented programs or national policy reforms in which no isolated comparison groups are likely to exist.

II. The Program Assessment Rating Tool – PART

A. What is it?

The Program Assessment Rating Tool, or PART, for short, is a management system developed in 2002 by The Office of Management and Budget – OMB (The Office of Performance and Personnel Management – OPPM) in order to evaluate the performance of United States federal programs. Based on those evaluations, the OMB/OPPM addresses recommendations directly to program managers, so to improve the results expected by the beneficiaries of public policies.

The PART diagnosis helps public managers in identifying a program's strengths and weaknesses, supporting funding and management decisions aimed at making the program more effective. It looks at all factors that affect and reflect a program's performance including its purpose and design; performance measurement, evaluations, and strategic planning; program management; and program results and accountability.

“By linking the GPRA and the PART processes, the federal government has moved a long way toward assessing the performance of program activities and focusing on their contribution to an agency’s achievement of its strategic and program performance goals. There is now much greater emphasis on developing performance measures according to outcome-oriented standards” (GAO, 1998).

The PART uses questionnaires designed by the OMB to determine public programs` overall effectiveness. Depending on the type of the program assessed, the PART has customized questions, related to the seven PART categories: Direct Federal, Competitive Grant, Block/Formula Grant, Regulatory, Capital Assets and Service Acquisition, Credit, and Research and Development.

Direct Federal **-** Programs where services are provided primarily by employees of the federal government, such as the State Department's Visa and Consular Services program.

Competitive Grant **-** Programs that provide funds to state, local and tribal governments, organizations, individuals and other entities through a competitive process, such as Health Centers at the Department of Health and Human Services - HHS.

Block/Formula Grant **-** Programs that provide funds to state, local and tribal governments and other entities by formula or block grant, such as Department of Energy's – DOE, Weatherization Assistance program and HHS' Ryan White HIV/AIDS program.

Regulatory - Programs that accomplish their mission through federal rules that implement, interpret or prescribe law or policy, or describe procedure or practice requirements, such as the U.S. Environmental Protection Agency's Mobile Source Air Pollution Standards and Certification program.

Capital Assets and Service Acquisition - Programs that achieve their goals through development and acquisition of capital assets, such as land, structures, equipment, and intellectual property, or the purchase of services like maintenance and information technology. Program examples include Navy Shipbuilding and the Bonneville Power Administration.

Credit - Programs that provide support through loans, loan guarantees and direct credit, such as the Export Import Bank's Long Term Guarantees program.

Research and Development - Programs that focus on knowledge creation or its application to the creation of systems, methods, materials, or technologies, such as DOE's Solar Energy and NASA's Solar System Exploration programs.

The following definitions are used in the PART:

Strategic goals: are statements of purpose or mission that agencies may include in a strategic plan. Strategic goals might not be easily measurable. For example, a strategic goal for a weather program might be protecting life and property, and promoting commerce and the quality of life, through accurate forecasts. To the greatest extent reasonable, the PART encourages agencies to use their strategic goals to develop specific, operational performance goals.

Performance goals: are the target levels of performance expressed as a measurable objective, against which actual achievement can be compared. Performance goals can be stated as either outcomes or outputs, but to be complete they should incorporate targets and timeframes into a performance measure. Two sort of performance goals are required: - long-term performance goals, which addresses performance that is generally several years or more in the future (an annual performance goal in the future and the cumulative effect of annual activities); and - annual performance goals, which should be stated in yearly increments.

Performance measures: are the indicators or metrics that are used to gauge program performance. Performance measures can be either outcome or output measures.

Targets: are the quantifiable or otherwise measurable characteristics that tell how well a program must accomplish a performance measure. In summary, together with the performance measure, the targets and timeframes establish a performance goal.

B. How programs are assessed?

1. Method

The Program Assessment Rating Tool - PART contains several questions about the quality of a program's long-term and annual performance goals and measures as well as questions about the actual performance compared to those established goals. Together, these are the primary issues that determine an overall PART rating.

A standard questionnaire is used to accomplish with a consistent series of analytical questions. Each PART questionnaire includes 25 significant, yet common sense, questions about a program's performance and management.

Those 25 PART questions are divided up into four sections:

Section I - Program Purpose and Design: the first section of questions asks whether a program's purpose is clear and whether it is well designed to achieve its objectives.

Section II - Strategic Planning: the second section involves strategic planning and weighs whether the agency establishes valid annual and long-term goals for its programs.

Section III - Program Management: the third section rates the management of an agency's program, including financial oversight and program improvement efforts.

Section IV - Program Results and Accountability: the fourth section of questions focuses on results that programs can report with accuracy and consistency.

For each PART analytical question, a short answer is required ("yes", or "large extent", or "small extent," or "no"), but always supported with consistent evidence besides a detailed explanation. The set of all answers determines a program's overall rating.

After working with the OMB to come up with the right answers to questions, federal agencies have the opportunity to formally appeal the answers with which they still disagree. In the appeals process, agency appeals are considered and adjudicated by a five-person panel comprised of members of the President's Management Council and a group of deputy secretaries responsible for management issues at the respective agencies.

a. Questionnaire

Extracted from The PART, the following are the questions applied to programs in 2008:

Section I - Program Purpose and Design:

1.1. Is the program purpose clear?
1.2. Does the program address a specific and existing problem, interest, or need?
1.3. Is the program designed so that it is not redundant or duplicative of any other federal, state, local, or private effort?
1.4. Is the program design free of major flaws that would limit the program's effectiveness or efficiency?
1.5. Is the program design effectively targeted so that resources will address the program's purpose directly and will reach intended beneficiaries?

Section II - Strategic Planning:

2.1. Does the program have a limited number of specific long-term performance measures that focus on outcomes and meaningfully reflect the purpose of the program?
2.2. Does the program have ambitious targets and timeframes for its long-term measures?
2.3. Does the program have a limited number of specific annual performance measures that can demonstrate progress toward achieving the program's long-term goals?
2.4. Does the program have baselines and ambitious targets for its annual measures?
2.5. Do all partners (including grantees, sub-grantees, contractors, cost-sharing partners, and other government partners) commit to and work toward the annual and/or long-term goals of the program?
2.6. Are independent evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?
2.7. Are budget requests explicitly tied to accomplishment of the annual and long-term performance goals, and are the resource needs presented in a complete and transparent manner in the program's budget?
2.8. Has the program taken meaningful steps to correct its strategic planning deficiencies?

Specific Strategic Planning Questions by Program Type:

2.RG1. Are all regulations issued by the agency or program necessary to meet the stated goals of the program, and do all regulations clearly indicate how the rules contribute to achievement of the goals? (Regulatory)

2.CA1. Has the agency or program conducted a recent, meaningful, credible analysis of alternatives that includes trade-offs between cost, schedule, risk, and performance goals, and used the results to guide the resulting activity? (Capital Assets and Service Acquisition)

2.RD1. If applicable, does the program assess and compare the potential benefits of efforts within the program and (if relevant) to other efforts in other programs that have similar goals? (Research and Development)

2.RD2. Does the program use a prioritization process to guide budget requests and funding decisions? (Research and Development)

Section III - Program Management:

3.1. Does the agency regularly collect timely and credible performance information, including information from key program partners, and use it to manage the program and improve performance?

3.2. Are federal managers and program partners (including grantees, sub-grantees, contractors, cost-sharing partners, and other government partners) held accountable for cost, schedule and performance results?

3.3. Are both federal and partner funds obligated in a timely manner, spent for the intended purpose, and accurately reported?

3.4. Does the program have procedures (e.g., competitive sourcing/cost comparisons, information technology improvements, appropriate incentives) to measure and achieve efficiencies and cost effectiveness in program execution?

3.5. Does the program collaborate and coordinate effectively with related programs?

3.6. Does the program use strong financial management practices?

3.7. Has the program taken meaningful steps to address its management deficiencies?

Specific Program Management Questions by Program Type:

3.CO1. Are grants awarded based on a clear competitive process that includes a qualified assessment of merit? (Competitive Grants)

3.CO2. Does the program have oversight practices that provide sufficient knowledge of grantee activities? (Competitive Grants)

3.CO3. Does the program collect grantee performance data on an annual basis and make it available to the public in a transparent and meaningful manner? (Competitive Grants)

3.BF1. Does the program have oversight practices that provide sufficient knowledge of grantee activities? (Block or Formula Grant)

3.BF2. Does the program collect grantee performance data on an annual basis and make it available to the public in a transparent and meaningful manner? (Block or Formula Grant)

3.RG1. Did the program seek and take into account the views of all affected parties (e.g., consumers; large and small businesses; state, local, and tribal governments; beneficiaries; and the general public) when developing significant regulations? (Regulatory)

3.RG2. Did the program prepare adequate regulatory impact analyses if required by Executive Order 12866, regulatory flexibility analyses if required by the Regulatory Flexibility Act and SBREFA, and cost-benefit analyses if required under the Unfunded Mandates Reform Act. Did those analyses comply with OMB guidelines? (Regulatory)

3.RG3. Does the program systematically review its current regulations to ensure consistency among all regulations in accomplishing program goals? (Regulatory)

3.RG4. Are the regulations designed to achieve program goals, to the extent practicable, by maximizing the net benefits of its regulatory activity? (Regulatory)

3.CA1. Is the program managed by maintaining clearly defined deliverables, capability/performance characteristics, and appropriate, credible cost and schedule goals? (Capital Assets and Service Acquisition)

3.CR1. Is the program managed on an ongoing basis to assure credit quality remains sound, collections and disbursements are timely, and reporting requirements are fulfilled? (Credit)

3.CR2. Do the program's credit models adequately provide reliable, consistent, accurate, and transparent estimates of costs and the risk to the government? (Credit)

3.RD1. For research and development programs other than competitive grants programs, does the program allocate funds and use management processes that maintain program quality? (Research and Development)

Section IV - Program Results and Accountability:

4.1. Has the program demonstrated adequate progress in achieving its long-term performance goals?

4.2. Does the program (including program partners) achieve its annual performance goals?

4.3. Does the program demonstrate improved efficiencies or cost effectiveness in achieving program goals each year?

4.4. Does the performance of this program compare favorably to other programs (including government and private programs) with similar purpose and goals?

4.5. Do independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results?

Specific Results Questions by Program Type:

4.RG1. Were programmatic goals (and benefits) achieved at the least incremental societal cost and did the program maximize net benefits? (Regulatory)
4.CA1. Were program goals achieved within budgeted costs and established schedules? (Capital Assets and Service Acquisition

Types of Programs:
RG = Regulatory
CA = Capital Assets and Service Acquisition
RD = Research and Development
CO = Competitive Grants
BF = Block or Formula Grant
CR = Credit

C. Programs' ratings

1. Rating range

The answers to the questions in each of the four sections result in a numerical score for each section, ranging from 0 to 100, 100 being the best score. Because reporting a single weighted numerical rating could suggest false precision, or draw attention away from the very areas most in need of improvement, numerical scores are translated into qualitative ratings.

The bands and associated ratings adopted in the PART are as follows:

PART Rating Range	
Effective	85-100
Moderately Effective	70-84
Adequate	50-69
Ineffective	0-49

The PART ratings are grouped into the broad categories "Performing" and "Not Performing". Programs that are performing have ratings of “effective”, “moderately effective”, or “adequate”:

Effective - The highest rating a program can achieve. Programs rated “effective” set ambitious goals, achieve results, are well-managed and improve efficiency.

Moderately Effective - In general, a program rated “moderately effective” has set ambitious goals and is well-managed. However, moderately effective programs likely need to improve their efficiency or address other problems in the programs' design or management in order to achieve better results.

Adequate - This rating describes a program that needs to set more ambitious goals, achieve better results, improve accountability or strengthen its management practices.

Programs categorized as “not performing” have ratings of “ineffective” or “results not demonstrated”:

Ineffective - Programs receiving this rating are not using public resources effectively. Ineffective programs have been unable to achieve results due to a lack of clarity regarding the program's purpose or goals, poor management, or some other significant weakness.

Results Not Demonstrated - Regardless of overall score, a rating of “results not demonstrated” is given when programs do not have agreed-upon performance measures or lack baselines and performance data.

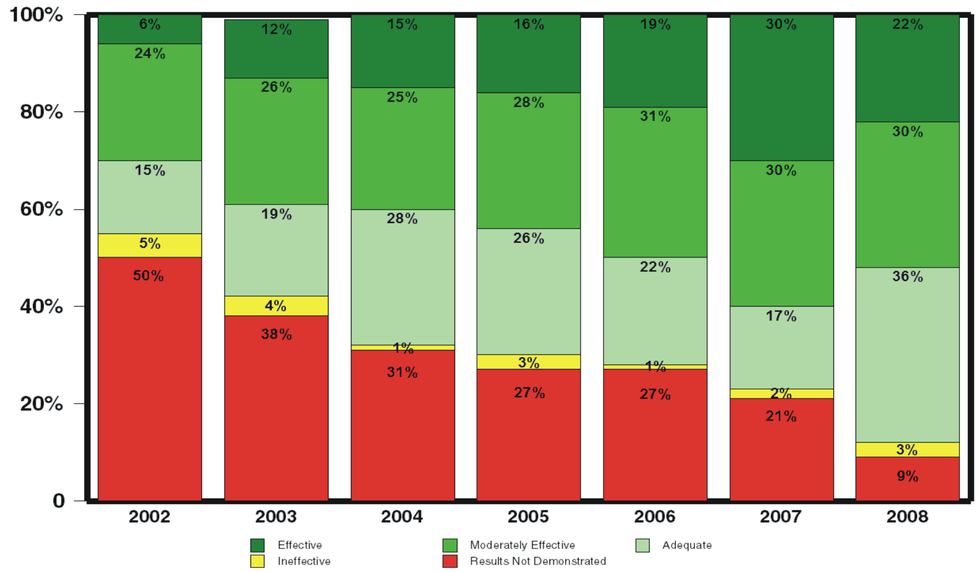
2. Overview of programs` recent ratings

Since the fully implementation of the PART, 98% of all U.S. federal programs have been assessed. Of the more than 1,000 programs assessed by PART, the cumulative distribution of scores is as follows:

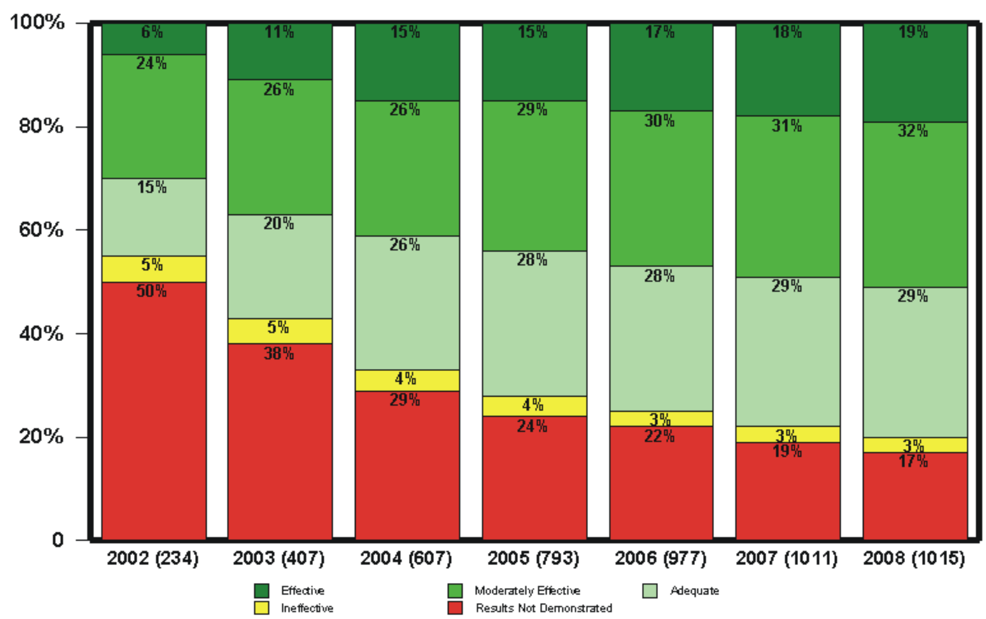
Distribution of Program Ratings		
Effective	193	19%
Moderately Effective	326	32%
Adequate	297	29%
Ineffective	26	3%
Results Not Demonstrated	173	17%

Program Ratings Continue to Improve

Assessment Results when First Assessed (2002-2008)



Cumulative Program Results by Ratings Category (2002-2008)



D. Performance reports

Once each assessment is completed, OMB develops a program improvement plan, so the agency accountable can follow up and improve the program's performance. All improvement plans elaborated so far are available at the ExpectMore.gov with unrestricted access.

E. Challenges and strategies

There is always room for improvement, the reason why each assessed program has a plan to improve performance and management over time. The U.S. Office of Management and Budget – OMB, together with the federal executive agencies, develops these plans collaboratively and then tracks the progress made, providing oversight and assistance as appropriate. As a result, all assessed programs are held accountable for improving their performance and management.

Programs are reassessed to improve its rating when significant changes have been made. For example, programs might be reassessed when new performance measures are implemented, follow-up actions have been completed, new performance data have been compiled, or a program evaluation has been completed. While the Administration hopes that, over time, increasing numbers of programs will fall into the “effective” category, the completion of PART improvement plans is even more important than a program's PART rating. The goal is continuous improvement of program performance.

The “improvement plan” is the set of budget, management, policy, or legislative follow-up actions that are created to respond to assessment findings. In many ways, the improvement plan is the most important aspect of a PART review since it is intended to link budgeting and performance and to demonstrate a commitment to continuous program improvement.

The type and scope of the follow-up actions in improvement plans vary greatly. In some cases, the recommended actions focus specifically on one or two key areas demanding improvement. In other cases, the follow-up actions are much broader. Sometimes, the assessment finds that a program is a duplicative of other or even that the program has already fulfilled its original purposes. In such cases, one of the follow-up actions might be to work with the Congress to end or terminate the program.

Congress is an important partner in the PART process. Members of Congress, Congressional Committees, and their staffs can review assessments of public programs under their jurisdiction and ask of those programs: “are the goals the program has set the right goals?”, “are the goals aggressive enough?”, “what can the program managers do to improve its management?”, “is the program doing everything it can to maximize its performance and get better every year?”.

F. ExpectMore.gov

ExpectMore.gov is a website launched in 2006 to provide the public with information on how federal programs perform. Thus, it includes information about every federal program assessed - what its purpose is, how it performs, and what it is doing to perform better.

The content on ExpectMore.gov is developed by the U.S. Office of Management and Budget – OMB together with the federal executive agencies. While the federal agencies implement and manage programs, the OMB works to ensure that the executive branch makes the best use of its resources to achieve the established goals. By making candid assessments of programs more accessible to the public, ExpectMore.gov raises awareness of what federal programs are doing to improve.

There are over 1,000 assessment summaries and detailed assessments available on ExpectMore.gov. Each summary provides an overview of the program mission and function, some of the key assessment findings, and related follow-up actions (improvement plan) based on the results of the assessment. It also links to the full program assessment details, the program's improvement plan, the program's website (when available), and the search results for similar federal programs. Detailed information about the assessment can be linked to from its summary, including the answer, explanation, and list of evidence for each question as well as the updated funding levels, performance data, improvement plan status, and actions agencies are taking to address those findings. Performance improvement plans are included with the detailed assessments, showing the completed actions a program has accomplished, and which other actions agencies intend to accomplish, so to improve their performance for the coming year.

One of the goals of ExpectMore.gov is to make more public what program managers are doing to improve performance and therefore spur more attention, particularly from Congress, to making programs get better every year. With this additional attention, it is expected that agencies will be most aggressive in taking the necessary steps to improve its performance as not all programs are taking sufficient action in this issue.

ExpectMore.gov is updated twice a year, in February with the publication of the President's Budget and in the spring/summer when the PARTs that are being assessed for that year are published. Additionally, during both dates, the most recent performance information is also updated and set available to the public. Over 5,000 measures are tracked on ExpectMore.gov.

III. Conclusion

The “New Public Management”, due its main concerns about efficiency, efficacy, effectiveness and accountability, clearly demands the adoption of performance measurement systems to evaluate public agencies and programs.

During the 1990s, the measurement and reporting of program performance are management tools that became widely popular in many countries. Demands for data documenting the performance of public programs rose from citizens, executives and legislators, who do have high expectations on the benefits of performance measurement efforts so far. Nevertheless, the analysis showed that unless public managers know what to do with performance measurement data collected, the answers evaluations offer are not enough to actually support the decision making processes.

The Government Performance and Results Act successfully increased the production of goals and measurement within the federal executive agencies. However, it did not, by itself, always motivate public managers in using these goals and measures to support the decision making processes. On the other hand, the interviews conducted for a report on the Program Assessment Rating Toll suggest that the PART reviews were key to motivating greater agency attention to outcome-focused goal-setting and production of outcome-focused measurement, although other changes made in the same time period and evolving familiarity with the purpose and requirements of GPRA might also explain increased availability of measurements (Metzenbaum, 2009).

Yet despite all challenges public managers face, the Program Assessment Rating Tool – PART has proved to be a well-succeed innovation carried out by the U.S. federal government in using performance measurement to improve public management. The PART systematically evaluates federal programs and, based on those evaluations, recommendations are addressed directly to program managers, so to improve the results expected by the beneficiaries of public policies.

IV. References

Ammons, D. N. *Accountability for Performance: Measurement and Monitoring in Local Government*: County Management Association, Washington, D.C., 1995.

Behn, R. D. *Why Measure Performance? Different Purposes Requires Different Measures*: Public Administration Review, September/October 2003.

Berman, E. and Wang, X. *Performance Measurement in U.S. Counties: Capacity for Reform*: Public Administration Review, September/October 2000.

Dupont-Morales, J. A. and Harris, J. E. *Strengthening Accountability: Incorporating Strategic Planning and Performance Measurement into Budgeting*: Public Productivity and Management Review, 1994, 17 (3), 231–239.

Epstein, P. D. *Using Performance Measurement in Local Government: A Guide to Improving Decisions, Performance, and Accountability*: National Civic League Press, New York, 1988.

Epstein, P. D. and Olsen, R. T. *Managing for Outcomes: Lessons Learned by State and Local Government:* The Public Manager, 1996, 25 (3), 41-44.

European Commission. *Evaluating EU Expenditures Programs: A Guide to Ex-Post and Intermediate Evaluation*. In: An Overview of Performance Measurement: Public Management, 1997, 76 (9), S2– S8.

Forsythe, D. W. *Quicker, Better, Cheaper: Managing Performance in American Government*: Rockefeller Institute Press, Albany, 2001.

Hatry, H. P. *Performance Measurement: Getting Results*: The Urban Institute, Washington, D.C., 1999.

Hatry, H. P., Gerhart, C. and Marshall, M. *Eleven Ways to Make Performance Measurement More Useful to Public Managers*: Public Management, Washington D.C., September 1994.

Hatry, H. P. *Where the Rubber Meets the Road: Performance Measurement for State and Local Public Agencies.* In: Using Performance Measurement to Improve Public and Non-Profit Programs. New Directions for Evaluation: The American Evaluation Association, Number 75. Jossey-Bass, San Francisco, 1997, 31–44.

Joyce, P. G. and Sieg, S. *Using Performance Information for Budgeting: Clarifying the Framework and Investigating Recent State Experience*. In: Meeting the Challenges of Performance-Oriented Government: American Society of Public Administration, Washington, D.C., 2002.

Joyce, P. G. *Using Performance Measures for Budgeting: A New Beat, or Is It the Same Old Tune?* In: Using Performance Measurement to Improve Public and Non-Profit Programs. New Directions for Evaluation: The American Evaluation Association, Number 75. Jossey-Bass, San Francisco, 1997, 45–61.

Julnes, P. L. and Holzer, M. *Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation*: Public Administration Review, November/December, 2001.

Koskinen, J. *Testimony on Government Performance and Results Act by the Deputy Director for Management, Office of Management and Budget*. House Government Reform and Oversight Committee, Washington, D.C., February 12, 1997.

Kravchuk, R. S. and Schack, R. W. *Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993*. Public Administration Review 1996, 56(4), 348–358.

Lynn, L. *Requiring Bureaucracies to Perform: What have we learned from the US Government Performance and Results Act (GPRA)?* University of Chicago, 1998. http://harrisschool.uchicago.edu/About/publications/working-papers/pdf/wp-lynn.pdf.

Metzenbaum, S. H. *Performance Management: Recommendations for the New Administration*. Presidential Transition Series: IBM Center for The Business of Government, 2009.

Newcomer, K. E. and Scheirer, M. A. *Using Evaluation to Support Performance Management: A Guide to Federal Executives*: The PricewaterhouseCoopers Endowment for the Business of Government: Washington, D.C., January 2001.

Newcomer, K. E. *Using Performance Measurement to Improve Programs*. In: Using Performance Measurement to Improve Public and Non-Profit Programs. New Directions for Evaluation: The American Evaluation Association, Number 75. Jossey-Bass, San Francisco, 1997, 5–14.

Newcomer, K. E. and Downey, *A. Performance-Based Management: What Is It and How Do We Get There?* The Public Manager 1997/1998, 26 (4).

Newcomer, K. E. *Measuring Government Performance*: The George Washington University, Washington, D.C., 2003.

Osborne. D. and Gaebler, T. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector:* Addison-Wesley, Reading, 1992.

United States. Government Accountability Office. *Managing for Results - Measuring Program Results That Are Under Limited Federal Control*. GAO/GGD-99-16, 1998.

United States. Government Accountability Office. *The Results Act. An evaluator's guide to assessing agency annual performance plans*. GAO/GGD-10.1.20, 1998.

United States. Government Accountability Office. *Performance Budgeting: Initial Experience Under the Results Act in Linking Plans with Budgets*. GAO/AIMD/GGD-99-67, 1999.

United States. *Public Law 103-62, The Government Performance and Results Act of 1993*.

Williams, D. W. *Measuring Government in the Early Twentieth Century*: Public Administration Review, November/December, 2003.

Wholey, J. S. and Hatry, H. P. *The Case of Performance Monitoring*: Public Administration Review 1992, 52 (6), 604–610.

Wholey, J. S. *Clarifying Goals - Reporting Results*. In: Using Performance Measurement to Improve Public and Non-Profit Programs. New Directions for Evaluation: The American Evaluation Association, Number 75. Jossey-Bass, San Francisco, 1997, 91–98.

World Bank. *Impact Evaluation - The Experience of the Independent Evaluation Group of the World Bank*.