

PROGRAM PERFORMANCE EVALUATION: JUDGMENT VERSUS MEASUREMENT

Charles F. Bingman , The Johns Hopkins University, Washington Center

When attempting to evaluate the performance of public programs, there is a big difference between what can be measured and what needs to be known. To begin with, measurement cannot deal with the external forces that managers and executives must deal with, and they do so by using their own judgment and experience. In addition, some of the most important considerations that drove program design and policy are sufficiently complex and sophisticated that they defy measurement. Measurement seeks to be rational in an often irrational world. Those who advocate performance measurement can rightly claim that they aid in the formulation of judgment, but they can't claim that measurement is a substitute for judgment. This is especially true in public programs which must function within a political system. The realities of government budgeting and political decision-making may totally refute the "facts" arising out of even the most disciplined performance measurement.

Even within the parameters of a given program, measurement is seldom conclusive. There are a wide variety of programs, organizations, and individual activities including many professions, that simply do not lend themselves to a useful level of measurement. Teachers may hope they have had an impact and that it can be measured, but the more complex and sophisticated the teaching, the less certain is the capacity to measure what (if anything) has really been learned. Doctors certainly use measurement for body conditions and for technical test results, but the essence of medicine still seems to be judgment and, undoubtedly, instinct.

So measurement does not necessarily produce "facts", and it almost never reaches the whole range of what it is necessary for executives to understand. It is easier and more productive when what is measured is output: units of production, actions completed, customers served, questions answered, and so forth. J. Edwards Deming became famous, first in Japan and then in the rest of the world, when he showed how very precise statistical analyses of manufacturing processes could show how product quality and reliability could be improved.¹

But the measurement of outcomes is considerably more difficult and less illuminating, and this is especially true of public programs. Most government programs involve multiple organizations at two or three levels of government, sharing responsibilities, and often in conflict. It is a huge problem to get all of the participants to measure what is useful in ways that are

Charles F. Bingman had a 30 year career as a U. S. Federal government manager and executive. He taught public management for 25 years, first at the George Washington University, and currently as a Fellow at the Center for the Study of Government at the Johns Hopkins University Washington Center. He has done consulting assignments with various government organizations in China and 15 other countries. His book "Why Governments Go Wrong" was published in 2006

¹ Walton, Mary, "The Deming Management Method", Chap. 1, New York, Perigee Books, 1986. Also, Deming, W. Edwards, "Quality, Productivity and Competitive Position", MIT Press, 1982

Program Performance Evaluation: Judgment Versus Measurement

compatible, and coordinate the results. And hardest of all are the attempts to measure the imponderables – the unexpected turn of events, the failures of understanding, the triumph of emotion over rationality, budget cuts, bad politics – or even the wrong people measuring the wrong things.

There is also another form of conflict often present. For many, especially among auditors or political officers, “measurement” has been seen as a form of auditing or investigation. Where measurement is sought as the basis for criticism, it may be used in a negative manner against program managers. It then generates a high degree of guardedness and rejection, as if it were an adversarial audit. The program manager may feel that performance measurement is “just another staff exercise”, and is not really going to be much help. Program managers may ask “I get paid for my judgment. Is your ‘measurement’ better than my judgment?” Or they may feel that performance measures seldom tell them anything they did not already know.

In the end, the arts of performance measurement are not up to the task of providing realistic evaluation of programs. Too much remains not adequately assessed or adequately understood. Measurement does not necessarily produce end facts, and in the last analysis, its value is only realized by the exercise of judgment. But all of this cannot be used to conclude that efforts should not be made to evaluate the performance of public programs in ways outside of the judgments of their managers, and there are more substantial forms of performance evaluation that can be utilized.

But first, what is program evaluation? Program evaluation is a more complex and systematic assessment of how well a program is working. It consists of the following types of activity:

1. First, a “needs” assessment: that is, a determination of what the true need is for the program, who it will benefit, and whether the perceived benefits are worth the cost. For public programs, the question must be asked as to whether the perceived needs can better be met by some means other than a public program.

2. Since almost all public programs are framed by some law or formal policy, it is necessary to evaluate whether the major public policies that define the program are current, realistic, feasible and “rational”, recognizing that some of the answers to these questions may be by political determination. Because the justification for a program or its objectives may be political, it does not mean that the program is irrational. What it does mean is that the interpretation of rationality and feasibility must be made in political terms.

3. Evaluations of critical processes and means for program implementation must be made in order to serve program objectives and to identify ways to simplify them and to determine if they can be improved.

4. Evaluations must deal with the real results of the program in terms of both outputs and outcomes; i. e. whether anything has changed as a result of the program activity. These are not just measures; such evaluations are expected to diagnose problems beyond just measuring them, and to yield some ideas about how to cure problems or maximize results.

5. The results of these program evaluations must then be presented to program managers and their superiors, and defended.

6. It is usually useful to attempt to mandate responses from these officials, and there should be the capacity to undertake revisions or improvements in the evaluation. Remember, these responses may be managerial or political, and both views require response.

It should be noted that program/project management is a distinct management discipline, with a large population of professional practitioners, a rich literature, and a number of very strong and active professional associations and institutes active in the field. It is strong especially in engineering, manufacturing, construction, facilities management, and information technology. A **program** can be defined as a long term, coherent set of related activities or projects designed to implement major policy initiatives, under the direction of a program office and manager. Programs are usually multi-faceted (for example, development of a national highway system) and of long term duration. A **project** is a finite activity dedicated to the achievement of specific defined goals, with a definite beginning and ending. Projects may either be separate ventures such as a space science probe or one of several projects undertaken within the framework of a continuing program such as a series of highway construction segments. Projects are typically organized by task under the control of a project office having largely autonomous authority and funding.²

There is a wide variety of government activities designated as “programs”:

1. Provision of customer related activities such as a social security program, health insurance, housing, national parks and forests, veterans hospitals, tax collection, civil and criminal investigation.
2. Facilities operation such as hospitals, parks, embassies, ports, airports, military facilities and research centers.
3. Systems development for such things as weapons systems, scientific experiments, aircraft design and development, transit and highway networks, or health care delivery systems.
4. Public regulation, including labor relations and protection, consumer product safety, nuclear safety regulation, environmental protections, equal opportunity and civil rights protections, and public utility regulation.
5. Research programs in such fields as agriculture, medicine, space sciences, information technology, transport, or international affairs.
6. Financial subsidy programs such as student loans, tax subsidies or penalties, small business promotion, environmental protection, import and export subsidies, mortgage insurance, rent subsidies for the poor.

² See for example Kerzner, Harold, “Project Management: A Systems Approach to Planning, Scheduling, and Controlling”, Hoboken, N. J., John Wiley and Sons, 2003.

Program Performance Evaluation: Judgment Versus Measurement

7. National security: military capabilities, foreign relations and assistance, and homeland security.
8. Disaster assistance and recovery, both domestic and international.

There are literally hundreds of types of public programs that can be undertaken in any given country, and a whole range of different organization structures have been invented to manage them. There are standard government agencies, staffed by civil servants, and operating only under enabling statutes, and approved government procedures. Another widely used organization, particularly in the People's Republic of China, is the state owned enterprises (SOE), which are a blend of partial operational independence with tight government policy control. In many cases such as the Public Service Units in China or the government sponsored enterprises in the U. S. loosely controlled organizations such as universities rely on governments primarily for policy guidance and some funding but are financially largely on their own. In China and elsewhere, there are mixed ownership enterprises, partly owned by the government and partly by private interests, usually with careful ground rules for the exercise of the government's involvement in enterprise decisions.³

The type of organization may vary substantially given the nature of the program. The following delivery systems are commonly used in all governments. In the design of a public program, lawmakers should recognize that they have a number of sophisticated options to choose from. Program evaluation must also recognize the importance of the organizational design in appraising what it can and cannot do.

1. There is direct service provision by government civil servants in standard public agencies operating on detailed legal and procedural constraints.
2. There are programs which primarily involve direct transfer payments from the government to individuals.
3. The government may purchase from private sector organizations it supplies, equipment and some services. The needs for purchase are a government decision, and the private sector organization has no policy responsibility.
4. There is the more complex contracting out of major program or project activities to the private sector. Under this form of procurement, the contractor is given responsibility under the contract for many of the responsibilities of the program/project office itself. In NASA and DOD especially, contractors have been given extensive responsibility for the design of whole systems (an aircraft, ship or space vehicle) including the estimated costs, schedules and technical performance parameters, all within the overall program goals of the supervising agency. In other cases, the program office may define its requirements explicitly in the form of technical, cost and performance objectives, with contractors undertaking the more limited role of fabrication and testing. In any event, however much responsibility is placed in the hands of contractors, the program officer is never relieved of the ultimate responsibility and accountability.

³ Yang, Dali L, "Remaking the Chinese Leviathan", Stanford U. Press, 2004.

5. The government may make grants of funds to another government unit or to a non-profit organization. Government officials are responsible for determining the necessity for the grant, the purpose of the grant, the amounts to be paid, and for supervising compliance by the grantee with the terms of the grant. Grantees are given considerable latitude and control over the manner of their own performance. In addition, for more complex projects (e. g. construction of a subway segment), the grant instrument may be more like a contract which spells out the nature of the work, rules governing performance, and the right of the government to mandate specific actions where considered necessary. Under this form of grant, funds may be withheld by the government if it is judged that grantee performance is not adequate or compliant. In China, such grants are largely discretionary from the top down, are often awarded on a patronage basis to friends and allies, and have an unsavory history of being diverted from their intended purposes, as for example, from environmental protection to economic development or managerial perquisites.
6. Vouchers may be given to eligible citizens for specific purposes such as school costs or housing subsidies. Government officers are responsible for determining who will be eligible, the amounts to be paid, and for auditing to assure that the recipients use the money for the defined purpose.
7. The government may make simple direct subsidy payments to public or private organizations.
8. In other cases, the government may make loans to an organization, or it may guarantee loans to certain organizations when they borrow from banks or other lenders. A good example of this kind of program is guaranteed student loans.
9. Some public programs are designed to create forms of insurance for certain transactions, such as home mortgages.
10. There are literally hundreds of programs that are based on government power to regulate – both organizations and the actions of individuals. Regulation may be either economic (e. g. public utilities or securities markets), or they may deal with the health and safety of people by regulating foods, drugs, air quality, water pollution, or hundreds of other potential threats.
11. Finally, all governments have elaborate programs for the administration of justice, from courts, police and public prosecutors, to anti-terrorist protection and anti-crime enforcement.

In a similar vein, governments have evolved many ways to finance their public programs, which also dictates who is responsible for evaluating a program, because each participant wants to guard the use of its own money. The normal pattern is to finance each program out of the general revenues of the government, which may be obtained in a variety of ways, including user fees, service charges, private sector financing, or matching fund agreements between levels of government. In some instances such as social security or highway programs, special trust funds are established to receive certain revenues, and these trust funds are supposed to be expended only for their defined purposes – but of course, they seldom are.

There are also a variety of ways in which a government can expend its money. In addition to the normal forms of public programs, governments may create “tax expenditures” in which it may have legal authority to collect some revenue but chooses not to do so. An example

Program Performance Evaluation: Judgment Versus Measurement

is making mortgage payments tax deductible. Grants of funds which are outright gifts may be made for certain purposes such as research, or cost sharing formulas may be arranged with grantees or private organizations. Both the annual budget and the tax system are famous vehicles by which the government can offer special advantages to some parties. Many of these arrangements might be evaluated as “bad management” or poor business practice or not cost effective but they may be a perfectly valid response to the mandates of enabling statutes. But it is reasonable to expect that governments at all levels will consider many options for the most cost-effective means for financing its programs, and there will be both political strategies and managerial tactics in selecting the optimum financing approach. In China, a major consideration is always which option permits the greatest retention of political authority for the minimum cost.

These options are also important in determining the most effective program design. One may argue that the intelligent selection of financing approach can save the government (or the tax payer) money, and thus the choice of financing is an important element of program performance evaluation.

Key Program Management Roles and Responsibilities

To be able to exercise real control over a project, the project office must be responsible for the following:⁴

1. Project goals, objectives and definition of outcomes. This includes the definition of products or services to be achieved including money and other resources, performance requirements, safety and reliability standards and sequencing of activity.
2. Overall project control: Even where parts of the project are let out through contract or delegated to others such as local governments, the central project management office remains fully responsible for all elements of the project including evaluation of the performance of contractors or local governments. In order to do this, the project office must be staffed with people of high technical and managerial skill, and must act coherently in the supervision of all elements of the project.
3. Huge projects/programs such as the Three Gorges Dam project in China involve an extraordinary number and range of government agencies, local governments, contractors, and subcontractors. The project office is responsible for the allocation of tasks to the various participants, and the coordination of their efforts, usually against a fixed time frame. The project office must track performance, expenditure patterns, problems, threats, shortages and cost overruns. The project office must formulate the overall budget proposal, and cope with the changes that might occur during the approval process. The Three Gorges program is not only an extraordinary development in itself, but it stands as a major test of the ability of the Chinese government to master the skills of program/project management.
4. The project office will determine the “make or buy” decisions where parts of the project will be contracted or assigned to other organizations, usually some other government ministry or

⁴ Frame, J. Davidson, “The New Project Management”, San Francisco, Cal., Jossey-Bass, 2002.

local government. In each case, the project office remains responsible for all work, and thus it must create to capacity to evaluate the performance of all other organizations who are partners in the project. This evaluation has three characteristics. First, it must determine whether the partner is properly meeting its responsibilities. Second, the evaluation should go beyond mere compliance and judge the adequacy or excellence of their performance. And third, and of serious importance in China is a careful scrutiny to prevent all forms of corruption, especially in the flows of money. It is in this arena especially that the Chinese never seemed to have achieved control.

5. Most projects are, of necessity held to a definite schedule or in fact a series of interlocking schedules. Failure to do so almost always ends up in expensive project “drift” both in the total time to completion, and in the thousands of sub schedules for subsidiary elements of the project. If key scheduled dates are missed, the project office is responsible for determining what to do to get back on a hard schedule. Nothing is more expensive than for the project to stretch out the time to completion. One of the most important elements of schedule control is called “change control”. While some change in a project is inevitable, and often desirable (i. e. technical improvements), the introduction of changes can be so disruptive that it they threaten the integrity of the project. Therefore, at some point in time, the design or plan for the project must be “frozen” and no further major changes accepted.

6. Finally, every project office must report to a higher level of management to report on progress, account for funds, and justify the current program of activity. In China, this upward reporting is especially complex, since it involves not only some supervising ministry for each program, but also a separate, and powerful, office of the Chinese Communist Party. The Chinese government practices a particularly tight form of “vertical administration” in which lower governments or organizations must report to the next higher level of government, right up to the top.⁵ Given the two parallel paths of bureaucratic reporting and political reporting, and the excruciating level of coordination required among a multitude of players, it may be extremely difficult to obtain timely decisions, and even know who gets to make each decision.

The Necessity for Program Evaluation

It is perhaps human nature that managers do not like to have other people looking over their shoulders or questioning their judgments, but it also true that organizations can enhance their results if more than one mind is at work. It is perhaps also true that, for most politicians, program evaluation is not popular, since it may reveal problems that they think will be politically embarrassing, and often, there is a notable lack of political will or ability to address shortcomings. As a result, there program evaluators may be made gun shy about challenging political sensitivities. This is especially true in China, where certain laws remain that makes any criticism of public activities highly sensitive and therefore dangerous. It is almost impossible to understand the line between acceptable criticism and “counterrevolutionary” activities which remain illegal.

⁵ Tsai, Lily L., “Accountability Without Democracy”, Chap. 2, Cambridge U. Press, 2007.

Program Performance Evaluation: Judgment Versus Measurement

But examining the nature of program management described above, it is very clear that simple processes of performance measurement are of only marginal value, and that even an occasional effort at more sophisticated performance evaluation is not enough. What emerges is that performance evaluation must be thought of as a constant and ubiquitous process of judging and evaluating program performance in many dimensions. In order to be effective, it must consider all elements of the project from the political and policy dimensions down through the technical, cost, schedule and human relationship imperatives. Every program/project manager should be conducting his/her own internal evaluation, and various forms of evaluation can be “built in” to each management system. The larger and more complex the program, the more sophisticated evaluation tends to become. It should be recognized that every competitive contractor selection is an evaluation; that every decision about re-competing a contract is an evaluation; every determination about a contractor’s performance and compliance with the terms of the contract is an evaluation, every modification in plans or budgets must be based on evaluation. In fact, it must be recognized that the judgment of managers is the greatest asset of all.

NASA Program Management Principles

One of the most interesting and exciting programs of the last 50 years have been those pursued by the U. S. National Aeronautics and Space Administration (NASA), especially in its early formative years. Both its manned space flight program and its extended series of unmanned scientific probes in the solar system and beyond and classic examples of program/project management, often at its finest as successful managerial design, often invented as programs were defined. The following are brief discussions of some of the philosophy of management that NASA so successfully developed.

1. Marshalling national resources

NASA rejected the option of building up a huge staff of government civil servants in favor of programs that would be largely executed through contracting. Further, as articulated by James E. Webb, NASA’s second administrator, he recognized that the journey into space would require not just a large government agency, but “the marshalling of all elements of American society including scientists, universities, industry and the American public.” Further, he spoke of the need to marshal the best resources in the United States because he perceived that NASA’s programs would be of extraordinary technical and managerial complexity.⁶

2. Securing the highest levels of government talent

NASA recognized that the direction of this massive effort of directing several manned space projects and scientific exploration projects along with a continuing research program would require the highest level of employee skills – not just adequate people, but outstanding people. It took excellent advantage of the lure of its new missions to attract that kind of top talent, despite government salaries that were not even competitive. NASA forced the best out of a government personnel system designed more for constraint than for innovation.

⁶ Webb, James E., “Space Age Management: The Large-Scale Approach”, New York, McGraw-Hill Co., 1969.

3. Flexibility in assignment/movement of key leadership

NASA, under the pressure of its fast moving program, had the skill to evaluate its key leadership, and its administrators and program office directors did not hesitate to remove people who were not adequate for their jobs. It could move swiftly to shift top people within the organization. It sought and got the flexibility to hire top people from outside of government and did so often to obtain the best possible talent.

4. The primacy of the project management approach

NASA was built on the stable base of the National Advisory Committee on Aeronautics, and the field centers of that organization became the R & D capability of the new NASA. But NASA, borrowing from the project management experience of the Defense Department, became an expert user of the program management style of management and an expert user of specific project offices housed in the field centers for institutional support. But each project office had explicit authority of its own that could not be constrained by the centers. Projects were finite in purpose, of specific limited time duration, possessed of independent authority and resources, and all powerful in their own sphere of activity. The permanent NASA field center base helped to set up and staff each project, support it during its lifetime, and absorb its people as the project phased down.

In a similar vein, the complex of contractors was developed along the same lines. That is, some contractors such as for data management or technical services were semi-permanent, but most contractors were hired and directed by the project offices, and were expected to expand or contract as needed during the finite life of the project. Thus, both the civil service workforce and the contractor workforce were directly tied to the current needs of the project and could be efficiently trimmed back as those needs declined. This was seen as far more efficient and cost-effective than the alternative of a large standing federal workforce which could be reduced or shifted only with great difficulty.

5. Program management vs. project management

NASA was, at any given time, responsible for a whole series of related manned space flight or space science activities. Groups of related activities were placed under the direction of a headquarters program manager in the agency headquarters, and then specific projects within the program were assigned to one or more field centers for implementation. While these respective roles can and did differ in detail, it is usually felt that NASA at its best typified a highly successful balancing of central direction with strong but decentralized project management in the field centers. This distinction was very important in the NASA culture. The management in the field centers viewed their management latitude as essential to performance optimization, and this is one of the intangibles that no performance measurement scheme would ever deal with. Conversely, even the field center leaders recognized that performance would also be weakened if the headquarters failed to live up to its role for planning, coordination, resource allocation, and conflict resolution. This was often referred to within the agency as the “mutual success/failure relationship.”

6. The line vs. staff relationship

Program Performance Evaluation: Judgment Versus Measurement

There was a strong pattern from the beginning that the line managers – i. e. the project managers – had to have the ultimate authority over their own project activities, and they should not be blocked or frustrated by the authority or actions of staff officials (e. g. personnel, accounting, finance, procurement, contract administration). These staff offices thus had two kinds of roles: to some extent, they were auditors of line manager performance and could insist on conformance with certain laws or regulations; but more importantly, they were service organizations as well, intended to support and assist the line managers. In most conflicts between these authorities, the line manager was expected to prevail. If a line manager could, however, be shown to have done something wrong, the fault and the consequences were clearly his. In fact, most of the people in these staff offices accommodated well to the “service” nature of their work, and did not like the “overseer” role as much. Staff organizations in NASA had a very good reputation, and produced many outstanding people.

7. NASA felt it to be absolutely vital to employ very disciplined and sophisticated forms of planning, program definition, objective setting, schedule controls, technical program evaluation and cost controls. The timing and sequencing of thousands of events was estimated and techniques devised for tracking each and reporting their status back into a central control center. This sophisticated planning became an essential form of control, since the failure to meet any one of a series of critical time related events could be determined in advance and actions taken to correct looming problems. In addition, funding could be related to program performance and progress. The cost of achieving each critical event could be priced out and the cost-benefit of each calculated where needed. The financial consequences of delays or failures could be estimated at an early point, and the program cost consequences calculate. NASA became known for its very high understanding of the links and interrelationships between cost, schedule and technical performance. The maintenance of large contractor staffs and facilities was extremely costly and any delay in a program could have drastic cost consequences. Thus, NASA sought to push each program at the best possible speed, and this fast paced performance saved the government money in the long run – a lesson often lost in other public programs.

8. The demands of operating in the strange and demanding new environment of space, far from help, pushed NASA into an extraordinary level of concentration on reliability, safety and quality of performance – both for the hardware and for the people. While many of the technical elements of such quality control are not transferable to other types of programs, NASA did demonstrate that it is possible to create strong institutional motivations for reliability and safety, embed such motivations into any organization, and find ways to enforce these disciplines.⁷

9. Constant program review

NASA developed an approach where the progress of its projects were subject to constant structured and disciplined review and evaluation at the highest levels of the agency. Most of the reviews were done monthly in front of the whole top management team, and there at their best when they were frank, open and powerful. Knowing that each project official would be exposed to this intense scrutiny and critique of their peers and superiors gave real muscle to these reviews. This is in marked contrast to the insipid staff meetings that tend to characterize other agencies. NASA people in turn subjected contractor organizations to the same kind of program

⁷ See Kraft, Chris, “Flight: My Life in Mission Control”, New York, Plume Books, 2001.

reviews, as well as special “tiger team” reviews dealing with serious problems. Most contract managers seemed to feel the same kind of professional motivation to do well in these reviews that drove NASA’s own managers. As part of this managerial style, NASA exercised its oversight of contractors in ways that gave them the same wide latitude and independence that NASA’s field offices wanted from NASA headquarters. Performance of contractors was defined mostly in outcomes rather than performance specifics. In many cases, NASA devised forms of incentive contracts which induced contractors to optimize certain kinds of outcomes such as innovation, quality, timeliness, or speed.

From this discussion, it should be clear that program performance evaluation is, or should be, a built-in element of the management of any activity. Evaluation must be applied to every facet of any program, and it must be constant and demanding. It will always be internal; from time to time, it may also be worth while to bring some form of external evaluation to bear. If such an external performance evaluation is undertaken, it cannot be useful if it is a simplistic assessment by people who do not understand the technical and managerial elements of the program, or is merely a performance measurement technique that will not recognize most of the truly critical elements that make a program work.