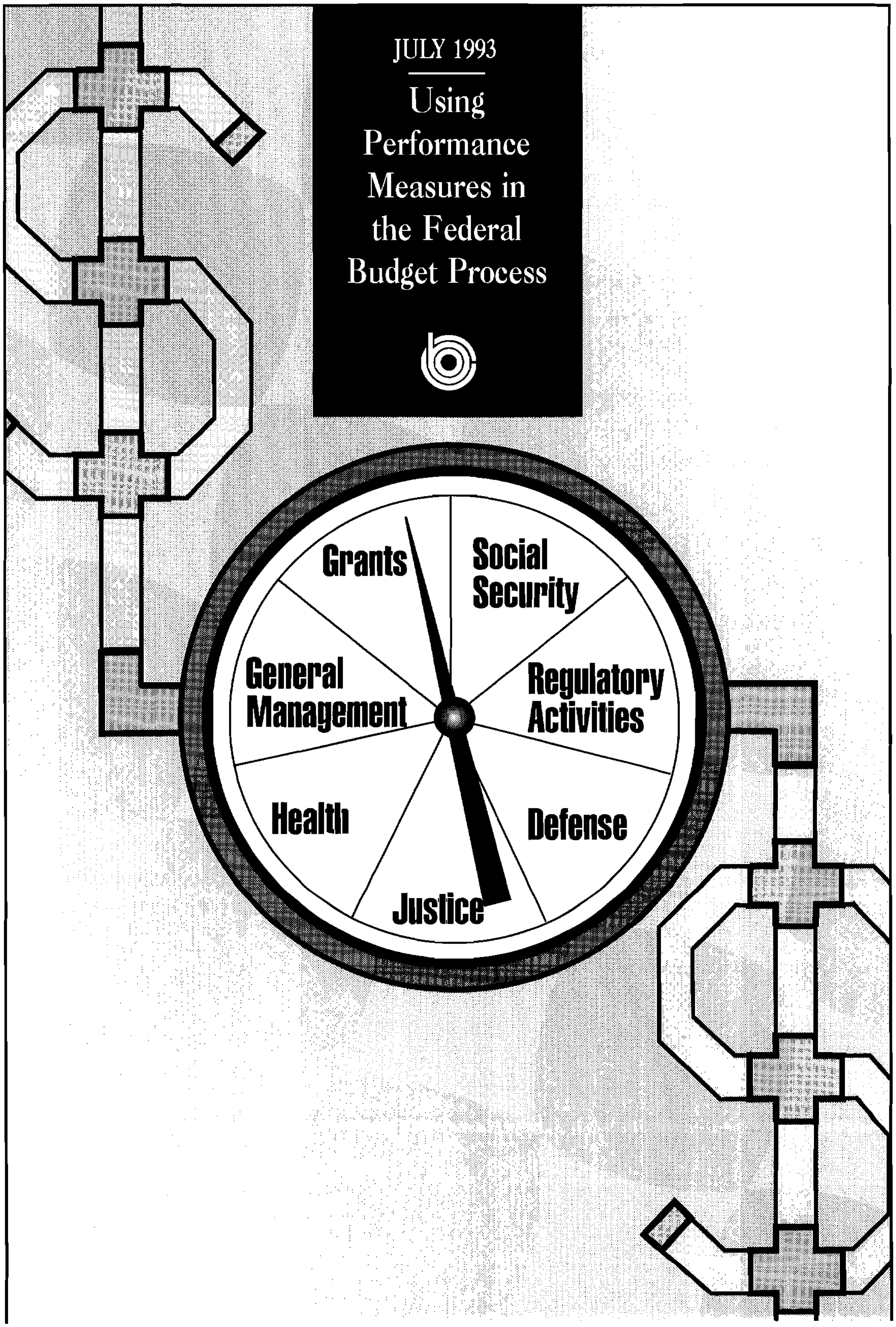
CONGRESS OF THE UNITED STATES
CONGRESSIONAL BUDGET OFFICE

A CBO STUDY

JULY 1993

Using Performance Measures in the Federal Budget Process



USING PERFORMANCE MEASURES IN THE FEDERAL BUDGET PROCESS

The Congress of the United States
Congressional Budget Office

NOTE

Numbers in the text of the study may not add to totals because of rounding.

Preface

This study of the uses of performance measurement in federal budgeting was conducted at the request of Senator John Glenn, Chairman of the Senate Committee on Governmental Affairs. It analyzes the probable effects of S. 20 and similar performance measurement efforts on the federal budget process. S. 20 was passed by the 103rd Congress during its First Session.

Philip Joyce of CBO's Budget Process Unit prepared the study under the supervision of Marvin Phaup and Robert Hartman. Leslie Bruvold provided the case studies of federal agencies included in Chapter 4. Ron Feldman, James Hearn, and David Moore reviewed early drafts of the study. Other CBO contributors included James Blum, Susan Borghard, Mary Braxton, Thomas Cuny, Fritz Maier, Linda Radey, Marty Smith, Ralph Smith, David Torregrosa, Paul Van de Water, and Bruce Vavrichek.

Valuable comments and suggestions were also made by others outside of CBO. Professor Robert Behn of Duke University conducted a particularly thorough review. Others who offered valuable comments included David Plocher and John Mercer of the Senate Committee on Governmental Affairs, Donald Goldberg and Kevin Sabo of the House Committee on Government Operations, Denise Fantone and John Kamensky of the General Accounting Office, Harry Hatry of the Urban Institute, and Professor Roy Meyers of the University of Maryland-Baltimore County. Special thanks are due those individuals in federal agencies and state and local governments who participated in the case studies included in this report and who reviewed sections of the report for thoroughness and accuracy.

Sherwood D. Kohn edited the manuscript, with editorial assistance from Christian Spoor. Martina Wojak-Piotrow prepared the study for publication.

Robert D. Reischauer
Director

July 1993

Contents

BOX

Summary

Recent events have focused attention on the measurement of performance by federal agencies, and the feasibility of applying those measures to the federal budget process. These efforts have one concept in common: the reasonable notion that federal agencies should be able to develop measures of program success, and that these measurements would be useful to managers and other policymakers. At least three separate efforts are proceeding in that direction: the application of federal financial management reforms; passage by the Congress of legislation covering performance measurement; and the inception of an effort to review the operation of the executive branch (embodied primarily in Vice President Gore's review of agencies' performance).

Measuring Government Performance and Applying the Measures to the Budget

Measures of performance are inherently different in the public and private sectors. In the private sector, the primary measure of performance for an organization as a whole is profit. The greater the profitability of a private firm, the better that firm is performing. Public agencies have no such accepted measure of performance. Performance must be judged against the goals of a public program. Goals differ from program to program, and even within a program there may be disagreement about the precise nature of its mission.

It is more difficult, therefore, for public agencies to determine with any certainty what they are trying to accomplish. In addition, once agencies have determined what they should be accomplishing, it is perhaps even more challenging to measure progress toward those goals. In fact, government agencies have usually focused on measures of input and output--resources used and activities performed. An emphasis on performance requires the use of more measures of outcomes--the broader results that are achieved because of the agency's activities. The differences between measuring public- and private-sector performance are less clear at the suborganizational level. Conversely, entities in both private and public sectors share the difficulty of measuring the success of such activities as personnel administration or research and development.

Even more difficult than measuring public-sector performance is finding a way to apply these performance measures to the allocation or management of resources in the public sector. Budgetary linkages can be developed in three different ways. The first is by budgeting based on performance. That is, traditional line-item budgets could be replaced by a system in which program managers have considerable flexibility in managing their resources but are held accountable for program results.

That type of system is the goal of many people who advocate performance measurement as a tool that can transform the nature of the federal budget process. Second, even if decisions about governmentwide allocation of resources are not affected, agencies may find measurement valuable for improving management of a given level of resources, regardless of whether their use results in a significant shift of resources from one program or agency to another. Third, measurement may be used for financial reporting. In this case, the government or its agencies can use performance measures to report on their accomplishments.

Local, State, and International Experiences with Performance-Based Budgeting

Much of the impetus for performance-based budgeting comes from those who believe that it has been successfully applied in local, state, and international governments and that the concept should therefore be used by the U.S. government. The Congressional Budget Office (CBO) reviewed these experiences in general, as well as specifically through selected case studies. In addition to using work done by the General Accounting Office (GAO) and other researchers, CBO made site visits to four local governments (Dayton, Ohio; Charlotte, North Carolina; St. Petersburg, Florida; and Portland, Oregon) and two state governments (Florida and Oregon) that are currently using performance measurement. In addition, CBO examined the experiences of several other nations that are measuring, reporting, or budgeting on the basis of performance.

Although all of the units of government studied use some form of performance measures, most are focused on the activities of agencies rather than on results. In general, the conclusions suggest little evidence of the much-touted advances in performance-based budgeting in local, state, and foreign governments. Performance measures have a limited ability to influence the allocation of resources, but they benefit management and financial reporting. State government success stories in performance budgeting are particularly difficult to find, a fact that may be related to the relative newness of state performance measurement. Neither CBO's visits to Florida and Oregon sites nor in-depth studies by GAO of five other states identified any effect on the allocation of resources by the government in any state currently using performance measures.[1] At all levels of government, however, performance measures are used in individual agencies to influence the use of resources and are a valuable management tool. Further, the Governmental Accounting Standards Board has influenced some governments to expand the use of performance measures for financial reporting.

Past Federal Efforts to Link Performance and Budgeting

The current effort to improve the linkage of performance measurement and budgeting is a logical successor to three similar attempts by the federal government during the past 40 years, namely performance budgeting, the planning-programming-budgeting system (also known as PPBS), and zero-based budgeting (ZBB). At least two of these (PPBS and ZBB) attempted to budget on the basis of program results. In general, these systems have fallen short of their goals. In particular, the substantial effort that went into these systems failed to change the way that federal resources were allocated.

1. General Accounting Office, *Performance Budgeting: State Experiences and Implications for the Federal Government*, GAO/AFMD-93-41 (February 1993).

The designers of each of these reforms tried to mandate a solution to federal budgeting and management by linking the budget to intended outcomes. Several important lessons emerged from their failures. First, such budgeting systems might prove antithetical to traditional incremental budgeting, and might be resisted by those who have some stake in the process that is already in place. Second, it is important, if such efforts are to succeed, to reach agreement on goals and objectives and to involve key staff who will be putting the system into effect. Third, each of the systems requires a great deal of data in order to survive. Much of the data collected in the earlier efforts were never used, which suggests that it is necessary to think through precisely how information will be used and how systems will be put into effect before setting out requirements. If the designers of future budget systems do not take into account the difficulties of tying the measurement of outcomes to budgeting, no system is likely to deliver on its promise.

Current Practices in Federal Agencies

Before assessing the possible effect of changing the practices of federal agencies and tying performance measuring to the budget, it is important to understand what federal agencies are doing. CBO reviewed the results of a GAO survey of federal agencies and studied six federal agencies or programs in detail to determine whether they had developed measures of performance and to what extent the measures were tied to the budget process.

GAO Study Results

GAO surveyed 102 federal agencies, each of which had more than 1000 employees or spent more than $500 million a year. These agencies accounted for 87 percent of all federal employees and 92 percent of total outlays in fiscal year 1990. Agencies were asked to report on what performance measures (if any) they used as a part of their operations and how they used them in agency management. The survey focused on the types of measures used, their users, the impetus for their development, and top managers' use of and satisfaction with them.[2] The responding agencies indicated that they used a wide array of performance measures. The majority of agencies reported that their measures were based on results. Agency managers reported that performance measures help with making budget decisions, managing programs, assuring accountability, and measuring program results or outcomes.

The GAO survey gathered information about a large number of federal programs, but a self-reported survey is limited in its ability to provide detailed and verifiable information. In fact, case studies of several agencies also reported in the GAO study uncovered much less performance measurement activity than the general survey results indicated. That finding suggests that surveys may be poorly suited to generating useful information in a complex, hard-to-define area such as performance measurement.

CBO Case Studies of Federal Agencies

CBO also visited six agencies in order to gather additional information on performance measurement. These agencies were selected on the basis of discussions with staff overseeing the GAO survey and on reports concerning which federal agencies had already been using performance measurement. The six agencies evaluated were the Environmental Protection Agency, the Internal Revenue Service, the Department of Labor's Employment and Training Administration, the Farmers Home Administration, the Department of Defense (internal service activities), and the Public

2. These results are reported in General Accounting Office, *Program Performance Measures: Federal Agency Collection and Use of Performance Data*, GAO/GGD-92-65 (May 4, 1992).

Health Service's Healthy People 2000 program.

In general, the results of these case studies were consistent with the experiences of state, local, and international governments. CBO concluded, as did GAO, that the GAO survey results may have overstated the extent to which outcome measures are being used in the federal government, and the extent to which these measures are tied to agency decisions about resource allocation. In fact, in the current setting, it is extremely difficult for agencies to link their performance measures and the budget process in any meaningful way. None of the agencies studied used performance measurement to make decisions about the level of resources that the program obtained in the budget process, although some intended to do so in the future. Performance measurement is used in some cases to award money to employees or subunits, but the basic aspects of agency budgets are not determined by the relationship between inputs and outcomes.

Even in these six programs, which were regarded by GAO evaluators and other experts as some of the best examples of federal performance measurement, gauges of outcome are relatively rare. The measure of job retention used by the Employment and Training Administration to evaluate the Job Training Partnership Act program may be the most effective, but even this measure is relatively brief (assessing results 13 weeks after the training has ended) and does not provide data on the marginal impact that this program has in the long run.

Further, certain governmental tasks or functions seem less amenable to effective measurement of performance. Programs that involve the provision of direct services that are under the agency's control are easier to measure than other operations, such as regulatory and grant activities. In general, performance systems seem to work best where there is direct accountability or clear cause-and-effect relationships between what the agency does (activities) and what it wants to happen (outcomes).

Prospects for Using Performance Measures in the Budget Process

The road to improving federal performance and tying its measurement to the budget process is studded with obstacles. Legislation cannot always bring about lasting change, but it can assist in some areas, such as motivating agencies to measure results, increasing flexibility for program managers, and mandating that policymakers receive information on program performance. Commitment on the part of the executive branch, however, is crucial in making any real changes in agency management, by providing incentives for agency managers, improving information systems, and tailoring solutions to individual problems.

Even if the legislative and executive branches were committed to improving performance measurement and tying those measurements to the budget process, two chief obstacles would remain. First, developing measurements that accurately reflect the performance of federal agencies is difficult, given that the success of so many programs managed by the federal government is determined in large part by nonfederal factors. Second, there is not enough demand by policymakers to change the way policies are made so that they are more responsive to the measurement of outcomes.

It is particularly important to recognize the limited payoffs that may come from a link between performance measures and the budget process, because using performance measures for allocating government resources implies a change from the current processes, which are heavily focused on inputs, to a system that is oriented toward outcomes. Creating such a system by simply demanding it is impossible, and past attempts to create a demand by fiat have failed. Real changes in the budget process could take place, however, if policymakers demanded information on the relationship between dollars and results. Reporting on the results of government programs in government financial statements can improve the capacity to create such information, in addition to developing a demand for it. But it will not be easy to create a demand or spur greater use of performance measures for allocating resources. Finally, the largest potential for real payoffs may be in the area of agency management of the resources once they have been provided in the budget process.

Before taking further legislative action, the Congress might consider exactly how performance measures could change the process of developing and applying policies. If proponents expect that mandating performance measures will lead to reform of federal budgeting and management, they will almost certainly be disappointed. Commitment on the part of the executive branch is probably more important than legislation, although legislation aimed at reducing impediments to managers might be helpful. In addition, it is important to recognize that several concurrent efforts--both legislative and executive--are under way. The Chief Financial Officers Act, the Vice President's National Performance Review, the move to bring total quality management to federal agencies, and the recently passed legislation in the Congress are all related to performance measurement. Actions in these areas should be coordinated in order to avoid duplication or counterproductive activity. In particular, movements to instill strategic planning and the setting of objectives in federal agencies should predate requirements to develop performance measures and to use those measures for budgeting.

Finally, however, the budget process is not likely to be changed substantially until and unless decisionmakers use information on program performance when making allocation decisions. Having this information--if good measurements of program results can be developed--is a necessary, but not sufficient, prerequisite to changing policymaking. Such a change is likely only to occur after the culture of performance measurement infiltrates the policy process. For this reason, the short-run emphasis should remain on developing performance measures for agency management rather than for use as a tool for allocating resources.

Chapter One

Performance Measurement and Budgeting: Issues and Concepts

Interest in measuring government performance, and in using these measurements for budgeting, is not new. Elected officials, government agencies, and other concerned individuals have long been interested in linking funding to results. But measuring results in the public sector is more difficult than in the private sector. The success of public agencies cannot easily be reduced to one (or even a few) measures. Conversely, the ultimate success of a private corporation is often measured by its profits. The struggle to improve the measurement of public-sector performance, nonetheless, has spurred numerous attempts to establish management and budgeting systems that are aimed at this goal.

Recently, there has been a renewed effort to legislate the measurement of performance in the federal government and to tie the results of federal programs more explicitly to budgeting. Several factors have contributed to this effort, including the use of performance measures by state and local governments and the need to reconsider priorities in an era of scarce public resources. In addition, the broad reception of a recent book about government performance, which included many "success stories" drawn from state and local governments, has heightened interest in the topic.[1] Several federal initiatives relating to performance measurement are under way, including the carrying out of federal financial management reforms (such as the Chief Financial Officers Act), the introduction of legislation mandating an expanded use of performance measures, and an evaluation of executive branch activities that is being coordinated by Vice President Albert Gore.[2]

Performance measures can play several different roles in the budget process, including those of resource allocation, financial management, and financial reporting. Performance measures may have many nonbudgetary uses as well. They may, for example, be used to motivate employees to improve their performance or productivity, or for other general management purposes. These uses are largely beyond the scope of this report, which focuses almost exclusively on the implications of performance measurement for the budget process.

Performance Measurement in the Public Sector

Measures of performance are inherently different in the public and private sectors. In private-sector organizations, there is usually an agreed-upon measure of performance for the organization as a whole--namely profit. The greater the profitability of a private firm, the better it is performing. There is no such

1. David Osborne and Ted Gaebler, *Reinventing Government* (Reading, Mass.: Addison-Wesley, 1992).

2. The Chief Financial Officers Act (P.L. 101-576) was passed in the 101st Congress. It established chief financial officers (CFOs) in 23 federal agencies; these CFOs are to focus on financial accountability.

agreed-upon measure of performance for public agencies.

Many government services are established because the market is felt to be giving the wrong price signals, either because of externalities or because the service has the characteristics of a public good.[3] In the case of public programs, then, the "price" is not set by the market, and "profit" is not a relevant measure. Performance must be judged against the purpose for which a public program exists. This purpose differs from program to program, and there may even be disagreement about the precise nature of a program's mission. It is difficult, therefore, for public agencies to determine with any precision what they are trying to accomplish. In fact, government agencies have usually focused on measures of input and output--on resources used and activities performed. An emphasis on performance requires more measures of outcomes--the broader results that are achieved because of the activities of the agency. The differences between measuring public- and private-sector performance are less clear at the suborganizational level. The difficulty of measuring the success of such activities as personnel administration or research and development, for example, is shared by entities in both the private and public sectors.

On the whole, public-sector performance measures are more subjective than measures in the private sector. This subjectivity has led to the development of more indirect methods to measure the worth and results of federal programs (such as benefit-cost analysis) that attempt to find a measure in the public sector that is analogous to the profit measure. In addition, measures often need to be developed at the suborganizational level; they are often quite subjective for both private and public organizations. In the public sector, officials who seek to measure the effectiveness of a program must establish the purposes of government activity and then develop measurements of how well that purpose is being met. Either is a difficult task.

Setting Objectives

The ability to measure performance is inexorably related to a statement of what the agency or program is trying to accomplish. The task of clarifying those goals is much more difficult for public-sector agencies than for private corporations.

Public-sector agencies operate in an environment in which they are usually asked to respond to many actors, including legislative bodies, elected executives, and the general public. Not all of these actors agree on the objectives of the agency or program. For example, there is often disagreement about whether a program should be managed in a way that promotes efficiency (minimizing costs per unit of output) as a primary objective, or whether it should operate primarily to provide equal access to services to as many citizens as possible, with cost as a secondary concern. Both are legitimate goals, but they often conflict.

For example, a mass transit agency that stresses efficiency as a primary value would operate only those bus routes on which fares pay for the operating cost of the route. If that same agency stressed access to service as a primary goal, routes would be operated to provide bus service to as much of the population as possible, regardless of cost. In order to assess performance, one would need to know which of the two objectives is valued most (efficiency or access). In the private sector, there are also complex multiple outputs of services. But prices that reflect consumer choices and costs often combine to reduce these to a single measurement: profit. This single measure, which stockholders or managers can turn to in order to gauge performance for the entity as a whole, has no analog in the public sector.

3. Both externalities and public goods are essential concepts in welfare economics. Externalities are costs or benefits that accrue to society as a whole but are not accounted for in the market transaction for a given good. Public goods have two characteristics that imply that they cannot be provided efficiently by private markets; benefits are provided to people regardless of whether they pay for the good or not, and it is impossible to determine the cost that should be charged to any individual.

Establishing Measures of Performance

Setting objectives is generally only the first step. The next is to design measurements that provide information to policymakers and line managers that will allow them to evaluate their progress toward agreed-upon objectives. Although many government agencies collect a great deal of data, these data have typically focused on the activities of the agency rather than its results or broader societal consequences. Input and output measures are typically used by agencies, even though outcome and impact measures would be more meaningful in evaluating programs.

Inputs represent the resources consumed by the operation of a government program; they are ultimately used to hire personnel, build facilities, contract for services, and so on. Inputs are easily measured, usually in terms of dollars. They are necessary for the achievement of objectives, but the question of how many inputs are required usually goes unanswered.

Outputs are the results of an agency's immediate activities. Although these outputs may not be sufficient for the agency to achieve its objectives, they are often necessary in order for the agency to do so. These gauges of outputs are sometimes also referred to as activity, work load, or process measures. Unlike inputs, outputs are often impossible to translate into dollars because there are no markets for most government activities.

Outcomes represent the broader goals of an agency. Outcome measurement should concern the extent to which the activities (outputs) of the agency have an intended effect. That is, they focus not only on the work performed, but also on the results of that work. Outcome measurements may cover activities that are largely under the control of program managers, or they may extend to an even broader set of measures (often called measures of impact) representing results that the agency may influence but does not achieve on its own.

The most important distinction stressed in this study is that between measurements of output and outcome. They are often confused, but an understanding of the difference is crucial to understanding the challenges to and the potential benefits of measuring governmental performance. This distinction is often clearest when considering specific measures related to individual government activities (see Box 1).

Obstacles to Measuring the Performance of Federal Agencies

It is certainly desirable to determine a program's objectives and then to devise measurements or other strategies to determine how well they are being accomplished. But it is not necessarily easy. There are obstacles to better measuring performance for all organizations, including public-sector agencies in general and federal agencies in particular. These obstacles include the difficulty of reaching agreement on objectives and priorities, the complexity of developing measures of results for government programs, the nature of the activities engaged in by the federal government, and the incentives for participants in the process to obstruct the reporting of valid performance measures.

Defining Objectives

Because objectives in the public sector are open to debate and interpretation, one of the greatest obstacles to public performance measurement involves agreement between and within the Congress or the executive branch, and with other significant actors, on what the objectives of the agency are. A program may also have multiple objectives, but there may be disagreement about their relative importance. The space station is an example. Some might view its objective as the ability to sustain manned orbits for an extended period of time, leading to breakthroughs in scientific research. Others, however, might view the primary objective of the space station to be mili-

Box 1.
Output or Outcome?
Examples of Performance Measures for Government Programs

One of the most difficult tasks in measuring government performance is that of deciding on terminology. The concept of inputs is fairly easy to understand, and these measures do not differ very much from one program to another. In general, inputs represent the raw materials, such as personnel, building materials, and office equipment, that go into the delivery of government services. There is considerably more confusion between outputs and outcomes. In general, in order to be classified as an outcome measurement, the assessment must present information that enables the policymaker to determine how well a particular program is operating in relation to its goals. Measures of output, however, focus on the amount of work accomplished or on the quality of the processes used to accomplish that work.

Outputs are sometimes treated as ends in themselves. But the question of how much work is done is very often distinct from the question of whether the work is achieving a particular purpose. For this reason, if the purpose of measuring performance is evaluating the results achieved by a program, outcome measures are superior to output measures. The importance of the distinction between output and outcome can best be demonstrated by looking at examples of measures from several different government programs (see table).

Public-sector performance measurement often addresses inputs and outputs but addresses outcomes less frequently. Most of the current support for performance measurement comes from those who want to move away from input and output measures toward outcome measures.

Examples of Output and Outcome Measures for Selected Programs

Output Measure	Outcome Measures
Elementary and Secondary Education	
Student-days	Test score results
Students graduated	Percentage of graduates employed
Dropout rate	
Hospitals	
Patient-days	Mortality rates
Average length of stay	Patient survey results
Admissions	Readmission rates
Mass Transit	
Vehicle miles	Population served (percent)
Number of passengers	Late trips (percent)
Police	
Hours of patrol	Rates at which cases are cleared
Crimes investigated	Response time
Number of arrests	Citizen satisfaction
Public Welfare Programs	
Number of requests	Applications processed in 45 days
Amount of assistance	Payment error rates
Road Maintenance	
Miles resurfaced	Lane-miles improved (percent)

tary rather than scientific. Still others might focus on the creation of jobs in the aerospace industry.

Developing consensus, though difficult, is an essential step in developing a meaningful system for measuring performance. It is impossible to design systems to measure the performance of public-sector programs without clearly understanding what the program is trying to accomplish. If a program's objectives cannot be determined, performance measures will always be ambiguous, if not superfluous. In fact, developing the consensus necessary to pass legislation may impede agreement on goals by promoting multiple goals that are poorly defined.

Problems in Developing Measures of Program Results

Many agencies concentrate on measures of work load or activity, since (even where agreement on agency goals is possible) it is often difficult to develop measures of the ultimate results of a program. Outcome measures are particularly difficult to address because it is hard to find acceptable measures of the achievement of a policy's objectives that are under the control of program managers.

For example, the ultimate policy objective for a job training program may be to reduce long-term unemployment through providing people with job skills. Because many other factors affect unemployment, it is very difficult to determine the policy or program's effect. A measure of some intermediate activity (that is, one that is under the control of program managers) may therefore be necessary. In the case of a training program, such measures have usually focused on how many program participants are placed in jobs. But since this measure is not directly related to the program's effect on long-term unemployment, this kind of measure will necessarily be imperfect.

Many effects that are beyond the control of program designers and agency managers influence outcomes. For example, citizens sometimes assist in delivering government services. This is often referred to as "coproduction" and encompasses such activities as participating in parent-teacher organizations and recycling waste products. To the extent that these factors, which are external to public agencies, can influence outcomes, the direct "production" of the agency will misstate the effect of its policies. Because of the difficulty in determining the cause of outcomes, some people advocate that government should focus only on outputs.[4] Whether that is practical may well depend on how many factors are beyond the control of program managers, and whether they can themselves be identified and controlled for. Even where outcomes are not a measure of program success, they may still be important as indicators of broader societal trends.

A troublesome tension may result between the reliability of output measures and the appropriateness of outcome measures. On the one hand, the most meaningful measures of success for a program are measures that provide answers to the question: How well did the program do in solving a defined societal problem or in meeting a broad social or political goal? On the other hand, the question that is easiest to answer reliably is: What did the agency do? These measures of activity are essential in evaluating a program's performance. But unless the precise linkage between agency activities and program results is clear, measuring activities will never provide a full understanding of the true results of a program.

Even if reasonable measures of program results could be found, how would one know whether the goal being measured is an appropriate one? Consider a case in which a federal subsidy is provided to a large group of people rather than being targeted toward individuals on the basis of need. If we measured how efficient the delivery of the subsidy is, and how

4. See Charles T. Clotfelter, "The Private Life of Public Economics," *Southern Economic Journal*, vol. 59, no. 3 (1993), p. 587.

satisfied the recipients are, we might know how well the program met its goals. It would not tell us whether an alternate design of the program (one in which a targeted group of deserving recipients received a direct cash payment, for example, while others did not) might result in a better use of societal resources. Without giving thought to whether the purposes of public programs are appropriate, performance measures might badly misinform decisionmakers. That may be a problem in the case of public programs, since legislative bodies (including the Congress) may avoid clarifying the specific purposes of a government program in order to obtain enough legislators' votes to gain approval of it.

Federal Activities Are Varied and Often Difficult to Measure

The federal government engages in a wide variety of activities, and each suggests different challenges for the measurement of performance. There are several classifications of government activities, but a useful one is included in the *Budget of the United States Government: Fiscal Year 1994*. This classification compares the outlays of the federal government in five categories: direct benefit payments for individuals (46 percent of estimated outlays for fiscal year 1993); national defense (20 percent); grants to states and localities (15 percent); net interest (14 percent); and other federal operations, including deposit insurance (7 percent). Of these categories, only net interest seems completely inappropriate for performance measurement. The characteristics of programs in each of the other categories, however, indicate that very different strategies would be needed in order to measure their performance.

Direct payments for individuals are used primarily by three programs--Social Security, Medicare, and Medicaid. These programs provide services to clients; therefore, the task of developing performance measures would probably focus on the quality with which these services were provided and the satisfaction of those who receive them. For example, measures of performance in the Social Security program would include the timeliness of check issuance and the error rates of checks (whether they were received by the correct recipients). Surveys of client satisfaction are also useful in determining the quality of direct services.

National defense is a particularly difficult area for which to develop measures of ultimate performance, perhaps because it is almost impossible to determine, without some military conflict, whether the objective of national defense is being achieved. Measurement is further complicated by the concept, as some would argue, that the purpose of defense is not winning wars but preventing them. For this reason, many measures of performance for national defense often focus on the quality and readiness of military forces, assuming that a well-qualified, appropriately trained military can best achieve whatever goals are paramount at the moment.

Grants to states and localities vary in purpose and form: that is, some are distributed by formula and others are provided at the discretion of the administering agency. The ways in which they are administered, however, make it difficult to measure the performance of these programs. Many federal grant programs are administered at the state or local level, meaning that their success is only partially under federal control. Determining performance for these programs, therefore, involves understanding the nature of their administration both in Washington and at the recipient level, as well as their goals and purposes.

Other federal operations encompass a wide range of governmental activities, including law enforcement, central management and administration, and regulatory activities. Determining performance is highly dependent on the specific nature of the activity. For law enforcement, measures of success may include the percentage of crimes solved. For an agency such as the Internal Revenue Service, measurement might focus on the dollars collected or the extent of taxpayer compliance. For a regulatory activity, success depends in large part on the companion regulatory activities of state and local governments and on the response of regulated organizations. Thus, it is difficult to link federal action and ultimate results.

The main point is that the ease with which performance measures can be developed and the ability of these measurements to inform decisionmakers differ from program to program. In general, developing measures of success (or at least satisfaction) may be easier for programs in which there is direct interaction between the federal government and some recipient of government service than for programs in which goals are less clear (such as national defense) or in which achievement is controlled by many different factors (such as grants to state and local governments). As a practical matter, the federal government should not be treated as a monolithic entity, implying that similar solutions exist to all problems. Instead, any search for performance measures should be activity-specific.

Possible Obstruction by Participants

People who would typically produce or use performance data may have incentives to misreport it or fail to report it altogether. If federal employees or agency managers believe that performance data might be used to punish them, they may measure only those things that will reflect favorably on their activities. Further, requesters or recipients of performance data can concentrate on measures that will support predetermined policy positions. An effort to identify appropriate measures and monitor the reporting of performance data may therefore be crucial to the success of any performance measurement. Obstruction of performance measures is less likely to be successful in cases where the accuracy and appropriateness of the measurements are monitored.

How Can Performance Measurement Affect the Budget Process?

Many efforts to mandate improvements in performance measurement focus on the budget process. Reforms such as the planning-programming-budgeting system in the 1960s and zero-based budgeting in the 1970s stressed the ability to base budgetary choices more explicitly on desired results. More recently, through the passage of the Chief Financial Officers (CFO) Act and the passage of legislation on performance measurement, the Congress is once again addressing the issue of using performance measures in the budget process.[5]

There are three possible advantages to linking performance measures to the budget. The first is the potential to use performance-based budgeting. That is, traditional line-item budgets could be replaced by a system granting program managers greater flexibility in managing their resources, but holding them accountable for achieving program results. That is the goal of many who advocate performance measurement as a tool that can transform the federal budget process. Second, even if governmentwide decisions on resource allocation are not affected, agencies may find measurements valuable for improving their management of a given level of resources, regardless of whether their use results in a significant shift of resources from one program or agency to another. Third, the government or particular agencies may use performance measures to report on their accomplishments.

5. S. 20 and H.R. 826 are the Senate and House versions of the Government Performance and Results Act. H.R. 826 passed the House on May 25, 1993. S. 20 passed the Senate on June 23, 1993. On July 15, 1993, the House agreed to the Senate version of S. 20, clearing the bill for the President's signature.

Allocation of Resources

The President and the Congress could use performance information to make decisions on how to allocate scarce resources among competing priorities. For them, performance measures are most useful if they can help in determining how much money should be spent on the various purposes of government: health care programs, national defense, environmental regulation, and so on. Ideally, these choices would be aided by a better expectation of what a new dollar for national defense would buy compared with spending the same dollar on some other activity, such as research at the National Institutes of Health.

Using outcome measures to allocate resources is a substantial deviation from the way that most observers describe the current budget allocation process. Policymakers currently are focused on inputs (usually represented by agency line items) and on making marginal changes in those inputs from year to year. They obviously must consider benefits as well, but these assessments of benefits are less formal and comprehensive than assessment of costs. This input-focused system is substantially different from one that would concentrate on trade-offs between results of agency programs rather than on adjusting traditional line-item budgets. The two methods of budgeting imply substantially different emphases; line-item budgeting focuses almost exclusively on how much money is being spent, and performance-based budgeting would concentrate on varying levels of program results that might accompany varying levels of funding.

Using performance measures to allocate resources, however, is difficult. First, not only is it hard to develop measures based on results for government programs, but it is almost impossible to establish common denominators of performance among the activities of government. For example, the measures of success in the health area are different from those in the defense area. Using performance measures for resource allocation implies knowledge of how to compare the measures for those two activities in a way that would inform trade-offs between the two. But trade-offs between government activities are almost exclusively a function of the perceived need and priority for government action. Performance measures can aid those decisions by establishing the correct level of need, but they cannot tell decisionmakers how much to provide to one activity at the expense of another. Politics--in the best sense of the word--plays an important and legitimate role in budget decisions, even where measures of outcomes exist.

Second, the relationship between performance and the budget is not straightforward. Poor results may be caused by the difficulty of the problem being addressed rather than by inadequacies in the design or management of a program. In fact, inadequate funding of a program may cause poor performance, which would lead to the conclusion that poorly performing programs ought to get more resources, and no increases should be given to agencies that meet their objectives.

Even if performance measures cannot be expected to dictate the level of resources given to one agency or program as opposed to another, they may provide useful information for decisionmakers. For example, if policymakers could obtain information about the connection between a given level of resources and a given level of results for a program, they might choose to provide a larger budget only if the incremental improvement in outcomes was judged to be worth the additional expenditure. In this way, performance measures can inform the budget process without dictating budget outcomes. Performance-based budgeting, then, is about transforming the debate over resource allocation from its current focus on inputs (dollars, personnel, and so forth) to focus on results. This change in focus does not, nor does it intend to, remove politics from the process.

Agency Management of Internal Resources

Governmentwide resource allocation is not the only potential function of performance measurement. It is useful to measure performance better, even without a link between gauging performance and allocating funds in a budget. Agencies may find measurement valuable for improving the use of a given level of resources, even if it does not result in a significant shift of resources from one program to another or from one agency to another. For example, an agency that is organized geographically could use performance measures to target resources toward those regions where the work load is greatest or where the problems are most acute. Agency subunits may improve performance if workers become more highly motivated when their efforts are being scored and assessed by management. Moreover, ties may be developed between the measurement of organizational and individual performance, such as proponents of so-called pay-for-performance schemes have suggested.

Financial Reporting

Performance measures can also be used to report the results of government activities to elected officials and decisionmakers. At the state and local level, the Governmental Accounting Standards Board has called for service efforts and accomplishments (SEA) reporting. The SEA initiative encourages state and local governments to include statements of service results in their annual financial reports. At the federal level, both the CFO act and the Federal Accounting Standards Advisory Board (FASAB) are expected to focus on the use of performance measures for financial reporting. The CFO act, for example, requires agency CFOs to develop "systematic measures of performance" for programs in their agencies. Although the FASAB is charged primarily with developing common standards for financial reporting, the need for performance measures has also been discussed in some of the board's deliberations.

Conclusion

Increasing the emphasis on outcome-oriented performance measurement in government is challenging. Performance-based budgeting is even more challenging. Given the obstacles and problems discussed above, it may be surprising to discover that performance measurement is being used to affect government budgeting. Moreover, the gaps between current practice (line-item budgeting) and performance-based budgeting are substantial. It is more likely, in the short run, that performance measures could have some influence on financial reporting or on the management of internal resources by agencies.

But this impression requires confirmation. There are places to turn for insights into the practice of and the potential for performance measurement. In fact, many of the proponents of the federal government's adoption of performance-based budgeting report that it has been successful in states, local governments, and other nations. Understanding what these efforts have achieved, and what they have not, can assist in evaluating the potential for performance-based budgeting to have a lasting effect on resource allocation in the federal government. Moreover, the federal government has adopted budgeting systems that were designed to connect the effectiveness of federal programs to resource allocation. In addition, federal agencies have widely varying experiences and abilities to measure performance. For this reason, reviewing past and current performance measurement practices can help to discover what federal experience and the experience of other governments suggest about the possibility of performance measurement now in the federal government.

Chapter Two

Linking Performance Measurement and Budgeting: The State, Local, and International Experience

The experiences of state and local governments have provided much of the impetus for a renewed call for performance measurement in the federal government. In addition, governments in other industrialized countries have been trying to improve their ability to measure performance. Proponents of changes in federal policy often point to the reported successes of these other governments in prescribing changes. But knowing whether other governments succeeded requires that two questions be asked about their practices. First, are the measurements that they use gauges of outcome or results, or are they limited to activity or process? Second, to what extent are these performance measurement systems integrated into the budget process? For example, how clearly are performance measures tied to choices about resource allocation, decisions about agency management of resources, or financial reporting?

Experiences of Other Governments with Performance Measurement and Budgeting

The general findings of a review of the performance measurement practices of other governments indicate that although all of the units of government studied use some form of performance measures, these measures continue to be heavily focused on the activities of agencies rather than the results of those activities. Moreover, performance measures are only tenuously tied to governmentwide resource allocation decisions. They do benefit internal agency management and financial reporting, however, and provide important signals about possible pitfalls to avoid in establishing a performance measurement system for the federal government.

Several generalizations may be particularly relevant as the federal government considers embracing performance measurement and switching from a budget system based on inputs to one that is based on performance. These generalizations address the differences between structure at different levels of government, the prevalence of outcome measures in these governments, the usefulness of the systems for budgeting, and the costs incurred by other governments in setting up performance measurement systems.

Relationship Between Government Structure and Agreement on Goals

Performance measurement systems seem to have worked best for two forms of government: on the local level, the council/manager forms of local government, such as those in Dayton, Ohio, and Sunnyvale, California; and on the national level, governments with parliamentary systems, such as New Zealand's and Great Britain's. In both cases, the development of goals and the setting of objectives crucial to the generation of meaningful measures are encouraged by a concentration of political power in only one branch of government. City managers report directly to city councils, and the government in parliamentary systems is the party in power. The result is clear signals from political leaders to agencies. In state governments with both a strong legislature and strong governor, and in the U.S. national government, the political system has not promoted agreement on goals, particularly when the branches are controlled by different political parties.

Prevalence of Outcome Measures

Outcome measurements are relatively scarce in the state, local, and international governments analyzed by the Congressional Budget Office (CBO). Measures of work load, process, or output are most prevalent. Even in jurisdictions such as Dayton; Charlotte, North Carolina; and Florida that have used measurement systems for many years, only limited progress has been made in moving to a system that focuses on outcomes. The reason is simple: most public programs find outcomes very difficult to measure. The difficulty of measuring results led New Zealand's government to explicitly target outputs for measurement, excluding outcomes. This approach can result in significant advances, but uses of the measures are limited. Citizen satisfaction surveys, such as those done in Dayton and in Portland, Oregon, are a notable exception. But even those have pitfalls since they are applicable to a limited range of government programs and it is difficult to establish benchmarks for comparison.

Use of Performance Measures for Budgeting

Performance measures can influence budgeting in three ways: through performance-based budgeting, management of resources at the agency level, or financial reporting. In state, local, and other national governments, performance measures are generally used for the latter two purposes and not to inform or influence substantially the allocation of resources. The links between outcome measurement and the budget process are particularly weak in many of these jurisdictions.

In virtually all cases, budgeting is done according to a traditional incremental model (increases and decreases in agency budgets are mainly uniform), not based on performance. The rationales for spending may be formed on the basis of anecdotal information about the results of programs but not on hard evidence. In particular, governments that are cutting spending tend to employ across-the-board methods for budget reductions, even though performance measurement systems should foster more careful targeting of budget reductions. State success stories in performance budgeting are particularly difficult to find. Local governments and national governments outside the United States are more advanced than state governments in their use of performance measures for budgeting. Although 14 states report outcome measures in budget documents, the ties of those measures to budget allocation decisions are tenuous.

Further, although General Accounting Office (GAO) and CBO site visits have uncovered some states that have potential for developing ties between performance measures and budgeting, the current efforts in these states are not far enough along in order for CBO to judge how well they will perform. The commitment of agencies is strongest when their managers see a benefit that accompanies the move to performance measures. For example, ensuring that performance measurement will lead to justifying an agency's budget in the face of fiscal limits (as in Oregon) often results in acceptance of the system at that level.

Even if elected officials and other policymakers decide that linking performance measures and budgets at a governmentwide level is not advantageous, individual agencies have much to gain from improving the linkage between performance measures and resources. In Florida and Oregon, human-services agencies have used performance measures to influence agency management. In Dayton, the Human and Neighborhood Resources Department uses performance measures to target funds toward individual parks and recreational facilities. In fact, given that elected officials in these jurisdictions have paid scant attention to performance measures, it is likely that the real benefit results from enabling agencies to target their use of a relatively fixed set of resources toward the achievement of stated objectives. In short, performance measurement is much more likely to be effective at the agency level than to influence governmentwide budget allocations.

Finally, performance measures will probably be used much more for financial reporting in the future. The efforts of the Governmental Accounting Standards Board (GASB) have already expanded the number of performance measures that are appended to government financial statements. The future should see not only an expansion of the number of governments that report measurements in their financial reports, but an increased reliance on outcome-oriented measures as a replacement for the more traditional output measures. Indeed, within 10 years, the GASB may require that all financial statements include measures of service efforts and accomplishments (SEA).

Costs and Benefits of Generating Performance Information

Performance measurement systems can generate a great deal of information, but the information must be used in order to justify the cost. Setting up a governmentwide performance measurement system can be expensive. Although some jurisdictions (most notably, Sunnyvale) use this information for management and budgeting purposes, others collect a lot of useless data. In most of the state and local governments that CBO visited, questions were raised concerning the usefulness of the data. Not only can this kind of data gathering be wasteful, but it antagonizes agency personnel whose support is necessary in order for such a program to work. Therefore, it is important to understand exactly how the collected data are going to be used before mandating the reporting of measures.

Results of Performance Measurement in Local Government

The progressive reform movement of the early 20th century was focused, first and foremost, on city governments. Professional city managers moved cities away from purely political administration by involving experts in day-to-day management. Most of the attempts at linking performance and budgeting in the United States have been made at the level of local government.

In general, the responsibilities of local governments in the United States make it easier for them to develop performance measures than for the state or national governments to do so. Because a larger portion of their budgets is devoted to the direct delivery of services (police, fire, utilities, schools, and so on), performance is somewhat easier to measure. Moreover, local governments provide these services to citizens who live in a relatively narrow geographic area, making accountability somewhat easier. Particularly in council/manager forms of local government, there is considerably more agreement on goals, or there is a clear process for adjudicating disputes. The manager works for the council in a way that has no parallel at other levels of U.S. government (that is, governors are not appointed by state legislatures, and the President is not appointed by the Congress).

One impetus for performance measurement in local government has been the move toward common accounting standards. This effort, spearheaded by the GASB's service efforts and accomplishments reporting, emphasizes the results of government activity rather than the activities themselves.[1] The GASB encourages local and state governments subscribing to its standards to improve their reporting of performance.

The GASB has issued a series of SEA reports covering a wide range of state and local services, including education, transportation, public safety, public health, and economic development. These reports provide prototype indicators for each of these service areas and recommend ways in which these measures can be used to evaluate performance for government programs. In particular, the reports advocate giving managers the ability to report explanatory information when actual performance deviates from planned performance; that ability is particularly important when events beyond the control of program managers cause the deviation. Further, the reports discuss the purposes for which SEA data can be used, including external reporting, internal management, and budget decisionmaking. Although SEA reporting is catching on (particularly in local government), one recent estimate suggests that it may take as long as 10 years before standards actually take effect.[2]

CBO visited several local jurisdictions--Dayton, Ohio.; Charlotte, North Carolina; St. Petersburg, Florida; and Portland, Oregon--that use performance measurement systems. In addition, the General Accounting Office and the Office of Management and Budget visited Sunnyvale, California, which has a system that many people consider to be among the most sophisticated in the country. All of these jurisdictions have council/manager forms of government. Their experiences suggest that it is generally difficult to link performance measurement and budgeting, and that it is difficult to sustain such systems once they are established. Nonetheless, evidence suggests that individual managers find the measures useful for managing their agencies or programs.

Dayton, Ohio

Dayton first inaugurated a performance measurement system two decades ago. The anchor of the current system is a strategic plan that lists objectives and measures considered most important to the city over the next five years. Each department is required to identify measures that address the key objectives of the strategic plan. Dayton's Office of Management and Budget reviews both agency budgets and agency objectives. The measures themselves are a combination of work load, efficiency, and outcome. Dayton was also one of the first jurisdictions to use citizen satisfaction (survey) data, which are still collected and used today.

The system does not seem to have had any substantial effect on the citywide allocation of resources. Staff in Dayton's Office of Management and Budget and the heads of two city departments (the Human and Neighborhood Resources Department and the Planning Department) agreed that the system was not used to affect resource allocation at the departmental level. Agency budget levels are fairly predictable from year to year; the council tends to make relatively minor budget adjustments at the margins. The system is used by the agencies, however, to manage their internal resources (for example, to decide whether to target resources toward parks or recreation facilities concentrating on particular types of activities).

Dayton does not plan to change the system. A recent evaluation by Wright State University, however, suggested substantial dissatisfaction with the system among rank-and-file city employees. Their dissatisfaction stems from a belief that the system is applied inconsistently (that is, it is not taken seriously at all levels of government), that it has not been

1. Harry P. Hatry and others, eds., *Service Efforts and Accomplishments Reporting: Its Time Has Come, An Overview* (Governmental Accounting Standards Board, 1990).

2. See Paul Epstein, "Get Ready: The Time for Performance Measurement is Really Coming!" *Public Administration Review*, vol. 52, no. 5 (September/October 1992), pp. 513-519.

tied directly to the city budget, that objectives do not tie well to job responsibilities, and that there are too few measures that assist in evaluating the results of city programs.[3]

In short, the system has not substantially affected the budget allocation process in Dayton. The system is not tied to budgeting (except at the agency level), but agencies such as the Human and Neighborhood Resources Department indicate that they use the system to manage their own resources. Budget office employees report that the system (particularly the citizen satisfaction surveys) is used as a management tool by the city manager. Moreover, Dayton reports the results of these surveys and other performance measures to the general public and to elected officials as a part of its strategic planning process. With the exception of citizen survey data, outcome measures do not seem to be prevalent in Dayton.

Charlotte, North Carolina

Charlotte's system is also 20 years old. City officials say that the first performance measurement effort was unsuccessful, largely because agency managers thought that the data collected were of limited use. Although the initial system focused almost exclusively on output data, the use of more qualitative measures has increased over time. In addition, within the last three or four years, the city has tried to expand the use of efficiency measures. The movement toward measuring program outcomes has not progressed as far to date.

Performance data are collected as a part of agency budget requests, and the budget office plays a major role in suggesting and refining the measures. Objectives are not tied directly to resource allocation. Major changes in the budget are not driven by performance measures. In practice, except for the police department (which tends to be exempt from budget cuts) agencies usually receive inflationary increases (in good economic times) or across-the-board cuts (in bad times).

Agencies submit three reports each year to the budget office, city manager, and council. These reports focus on the actual achievement of objectives compared with the planned level and are issued at five months, eight months, and the end of the year. Most agencies, however, do not use this information for internal management purposes. Some, such as the city's Department of Transportation and the Solid Waste Services Department, do use measures, but it is not clear whether the requirement to collect measures is a motivating factor. The city council has not used the system for budgeting or any other purpose.

Although some people have recommended abolishing the system, the city is reluctant to scrap it without a replacement. The budget office has initiated such an effort that would pay more attention to a hierarchy of objectives--a movement that would stress developing mission statements, strategic plans, objectives, and outcome measures.

St. Petersburg, Florida

St. Petersburg has experimented with performance measurement since 1973. The initial effort was discontinued in the late 1970s. St. Petersburg's second stage of performance measurement began in the mid-1980s, when a management study recommended that a system be developed to measure performance more effectively. The system has three components: efforts to develop mission statements, goals, objectives, and measures; surveys to measure citizens' satisfaction with specific internal services; and management studies of particular problem areas.

Agencies are required to report measures of performance when they send their annual budget request to the city manager and the council. The measurements are primarily assessments of work load and have not been used

3. Wright State University, Center for Urban and Public Affairs, *MBO Evaluation Report* (Dayton, Ohio: Wright State University, September 3, 1991).

much by the manager and the council in making budget decisions. Further, of three agencies interviewed--recreation, police, and planning--only the police department seemed to be using its measures for internal management. All three agencies agreed that the manager and council did not use them. In fact, although there is some recognition that data overload killed the old system, the agencies still perceive that they are producing measures for no good reason. The budget office is working to interest the council and city manager in the system by including the measurements in the briefing book that accompanies the budget to the council. In fact, during the budget process for fiscal year 1993, the interest in performance data by both the public and the council increased in response to these efforts.

Portland, Oregon

Through the office of the city auditor, the city of Portland prepares an annual report of city service results, modeled after the GASB call for service efforts and accomplishments reporting. Portland's SEA report focuses on the largest and most visible city services, such as police, fire, parks, and street maintenance. The report covers approximately 70 percent of city spending and staff.

The auditor's office used three general approaches in developing the report. First, a trend analysis compared data on Portland services to past service levels and standards of performance. Second, the results of Portland city services were compared with those of six other cities. Third, the auditor's office administered a citizen satisfaction survey. All large city agencies participated, even though their participation was voluntary.

CBO interviewed three officials from agencies participating in the SEA program. Each of these officials had parallel efforts to measure outcomes in place even before the SEA initiative. The SEA measures, then, were largely a recasting of data that the agencies were already collecting. The managers did not feel that the results of the citizen surveys were particularly useful. In one case, the managers felt that the survey was inadequate. The other two respondents could not use the survey results to evaluate agency performance because they lacked baseline data. In late 1992, however, information reported to the city council began to focus attention on problems that the measures suggested. All respondents agreed that outcome measures should be tied to the budget process, but only the parks and recreation department has any specific near-term plans to do so. The auditor's office changed the timing of the issuance of the 1992 report in an attempt to promote the use of the information in the budget process.

Although the report summarizes the results of city activities for Portland citizens and other interested parties, agency practices seem to have changed only slightly in response to the survey. In part, the slight change can be attributed to the emphasis on external reporting, in addition to the fact that city agencies are already initiating performance measurement systems. Particular progress seems to have been made in the area of citizen surveys, but using the surveys to inform management or budgeting will depend on developing several years of data for comparison purposes. In addition, the city has had some difficulty in collecting comparable data from other jurisdictions.

Sunnyvale, California

Sunnyvale has perhaps the best-known system for measuring the performance of local government. The city of Sunnyvale began to develop a system in the early 1970s in response to an effort spearheaded by the General Accounting Office to assist state and local governments in developing their budgeting and accounting systems. The city developed a more elaborate system in the late 1970s. It combines strategic planning, performance-based budgeting, and pay for performance for management employees. Each department identifies measurable performance goals; budgets and goals are closely tied.

In 1991, the Office of Management and Budget and GAO sent a joint delegation to visit Sunnyvale. The delegation reported that the city's budget system integrates performance measures, agency management, resource allocation decisions, and long-range planning. Several factors seem to contribute to the success of the Sunnyvale system. First, the city operates on a full cost-accounting basis, which allows officials to identify with some precision how much it costs to provide different levels of service. Second, the budget focuses on the outputs that will be produced rather than on spending. In fact, the city council does not even vote on the budget, in a traditional line-item sense. Rather, the council approves goals for city programs, and the level of resources necessary to meet those goals is implicit. Third, a detailed set of objectives and performance measurements govern all municipal functions. Many of these measures are oriented toward outcomes. Fourth, early-warning and long-range planning are stressed. The city sets overall goals and objectives 20 years in advance and projects the revenues and expenditures necessary to achieve these objectives for 10 years in the future. Fifth, under a detailed pay-for-performance system, there are explicit ties between individual and organizational performance.[4]

The OMB/GAO site visit verified the general impression that the system works well for Sunnyvale, but some skepticism remains about whether such a system could be adapted to the federal government. The council/manager system in Sunnyvale incorporates a degree of agreement on goals that may be difficult to sustain in other forms of government. In addition, the expansion of Sunnyvale's tax base in recent years means that the city has operated without significant resource constraints, which has allowed some certainty in long-term resource planning. Moreover, the city's managers are held fully accountable for results; salary bonuses are tied to the achievement of planned outputs or outcomes.

4. Statement of Thomas F. Lewcock, City Manager of Sunnyvale, California, before the Senate Committee on Governmental Affairs, May 23, 1991.

State Governments' Experiences with Performance Measurement

Many states have generated performance data in the past, and these measurements have generally not influenced resource allocation. Several states are undertaking a renewed effort to carry out statewide management or budgeting reforms. Although state governments do not focus as intensely as local governments on direct service delivery, they nonetheless administer many programs that should be good candidates for outcome measurement. Moreover, the GASB service efforts and accomplishments requirements have affected state governments as well, although SEA reporting has not progressed as far in state government as it has in some local governments.

The majority of states use performance measures in their budget processes, but these measures usually do not concentrate on program effectiveness. The National Association of State Budget Officers recently surveyed performance management practices in the states. The survey found that in 1991 34 states reported that they used performance measures as part of their budget requests. Most of these states required that agencies submit performance measures with their requests. Only 14 of these states reported that they used measures of program effectiveness (outcomes or results). The measurements are typically reported in the executive budget document. They usually cover three years or more and report these data at least at the program level.

State uses of performance measures have increased. Another recent study that reported the results of a series of surveys of state budget practices between 1970 and 1990 remarked on the use of program performance measures in budgeting.[5] The survey found that almost all states required agencies to submit productivity and effectiveness (an approximate surrogate for output and outcome) measures when requesting funds for new programs in 1990. Only about one-quarter of the states had required such measures in 1970. Moreover, the percentage of states reporting measures of effectiveness in the budget document increased from 29 percent in 1970 to 65 percent in 1990. The use of productivity measures and projections has also increased markedly over the 20 years.

There is a difference, of course, between collecting and reporting measurements and using them in any significant way to affect budget decisions. To review current practice and future prospects for linking performance measurement and budgeting, CBO visited two states--Florida and Oregon--that are considered to be among the leaders in developing performance measurement systems. In addition, CBO reviewed the results of a recent GAO study that included data from several other states, including Hawaii, Iowa, North Carolina, Louisiana, and Connecticut. Both the states visited by CBO and those in the GAO study were chosen specifically on the basis of reports that they had substantial performance measurement efforts in process.

Florida

The state of Florida has been attempting to tie performance measurement to the budget for 25 years. Over that time, each agency has been required to report measurements with the budget request that goes concurrently to the governor and the legislature. State officials interviewed by CBO uniformly viewed this system as creating a lot of paper without significantly influencing budget policy. This is mainly because the measurements have been focused on output in the past and have not been very useful for setting policy.

Recently, however, there has been renewed interest in performance measurement, which will be included in the current statewide planning process. At the highest level, a series of Florida benchmarks will include a series of 100 quality-of-life indicators, which are intended to serve as guidelines for policymaking. Part of this process involves the development of agency plans related to state planning objectives.

Although the linkage of performance measures to resource allocation statewide is not likely to occur in the immediate future, it could happen earlier at the individual agency level. For example, the Department of Health and Rehabilitative Services (HRS), which is the umbrella agency for human services, is attempting to develop agency plans that feed into the statewide plan and to include outcome measures in this plan.

Florida planners hope to tie the statewide planning process and the budget process together. They would develop agency objectives that tie into the statewide plan, and the legislature would give lump-sum appropriations to those agencies and hold them accountable for results. These efforts would change the budget process from a system in which controlling expenditures is paramount to one in which planning for the future is perceived as the most important goal.

Florida is also trying to refine the process of developing agency plans (similar to the HRS plan) and induce agencies to develop better measurements. Within one year, Florida would require all agencies to measure outcomes. Eventually, the state hopes to tie individual agency indicators to the state plan and to report those indicators in its budget.

Over the years, Florida's budget process has required agencies to produce a lot of perfor-

5. Robert D. Lee, Jr., "Use of Program Information and Analysis in State Budgeting: Trends of Two Decades," in Thomas Lynch and Lawrence Martin, eds., *Handbook of Comparative Public Budgeting and Financial Management* (New York: Marcel Dekker, 1991).

mance data, but the state's failure to tie performance measures to the budget has antagonized many agencies. Furthermore, a discussion with the state auditor's office indicated some problems with the reliability of data. In some cases, audits were unable to verify the existence or accuracy of the performance data being reported. Planning and budgeting processes are not linked, although this is an explicit goal for the future. Even if the system is never used for budgeting or gubernatorial management, however, the efforts of individual agencies produce some benefits. For example, if HRS manages its resources better and more in line with stated goals, this effect is positive regardless of whether it influences the state legislature's or governor's decisions about resource allocation.

Oregon

Several events have taken place concurrently to improve the environment for performance measurement in Oregon. First, the state is engaged in an ambitious effort to increase the use of performance measures in the budget process. Second, in November 1990, Oregon voters passed a local property tax limitation, called Ballot Measure 5, which will ultimately create a state budget shortfall of more than 15 percent.[6] The state also has a long-range planning process, called Oregon Benchmarks, which was developed in advance of the performance measurement effort.

Current initiatives primarily focus on training agencies to use performance measures that allow for the weighting of both efficiency and effectiveness. Because the budgetary environment after Ballot Measure 5 stands to be much more competitive, agencies have incentives to develop measurements, and agencies that fail to do so may find themselves at a disadvantage in the appropriation process. More broadly, one of the governor's explicit goals is to educate Oregon citizens about the services provided by state government and how the measurement system feeds into that effort. All agencies have now participated in the training sessions.

Oregon is attempting to expand the reporting of performance measures in the budget and to link together performance measures and the benchmarks in the budget process. For the last biennial budget, each agency was required to develop eight to ten indicators that will be expanded and refined for the following two years. In addition, each agency has been given incentives to develop new initiatives that tie in with the Oregon benchmarks. For the most recent budget process, the state used a modified zero-based budget process in which agencies were required to forecast the impact of reducing their budgets by 10 percent to 20 percent from the previous year's level. Agencies were then given opportunities to recover a portion of these lost funds, in part by adding programs that are explicitly related to the Oregon benchmarks.

It is still too early to tell whether the performance measurement effort has created any lasting change in Oregon's managerial or budgetary environment, but there are some impressive signs of progress. Some agencies (particularly the Adult and Family Services Division) seem to be using the reports generated by the measurement system as their primary management reports. In addition, some of the measurements themselves are quite ambitious. For example, the state's vocational rehabilitation department, which had previously collected little but work load data, now has some very good outcome measures, including (most notably) the percentage of clients that remain employed 18 months after they have left one of the department's programs. Other agencies do not seem to be quite as committed to the system.

Although there has clearly been progress in Oregon, the picture is not all positive. First, because agencies have agreed to use the measurement process primarily as a tool for selling their programs, there are no incentives for

6. Ballot Measure 5 is expected to have this effect on state finance because it would require the state to replace a portion of property tax revenue lost by school districts.

reporting negative performance. Second, the use of the system varies widely from agency to agency; the most important single variable seems to be the commitment of the agency head. Third, though the legislature seems supportive, it remains to be seen whether performance measures can be used to affect resource allocation in any meaningful way.

General Accounting Office Study of Five States

The General Accounting Office has reported the results of a recent study of five other states--Connecticut, Hawaii, Iowa, Louisiana, and North Carolina--that researchers visited to determine how extensively performance measures were tied to budgeting in these states.[7] GAO chose these states because surveys and other reports had listed them as among national leaders in developing performance measures. GAO researchers interviewed both executive branch and legislative branch officials. They reported that:

> Despite long-standing efforts in states regarded as leaders in performance budgeting, performance measures have not attained sufficient credibility to influence resource allocation decisions. Instead, according to most of the state legislative and executive branch officials we interviewed, resource allocation decisions continue to be driven, for the most part, by traditional budgeting practices.... Outside the budget process, state officials say that performance measures have aided managers in (1) establishing program priorities, (2) strengthening management improvement efforts, (3) dealing with the results of budgetary reductions, and (4) gaining more flexibility in allocating appropriated funds.[8]

The results of these site visits by GAO are generally consistent with CBO's conclusions about Florida and Oregon. Performance measures do not influence governmentwide decisions about resource allocation in these five states; rather, more traditional budget practices predominate. Although budgeting was not tied to performance measurement in any of the five states, state officials reported using performance measures to support management improvement efforts within departments. In short, if performance measures make a difference in these states, it is in agency use of resources rather than in the allocation of resources by governors or state legislators.

Experiences of Other Nations

At least five industrialized countries--Great Britain, Australia, New Zealand, Canada, and Sweden--have put into practice management or budgeting reforms that have included an element of performance measurement. Because these countries have parliamentary forms of government, it is undoubtedly easier to adopt and carry out these reforms than it would be in the United States. In such a system, the party (or coalition) in power is, by definition, the government. This contrasts with the U.S. system, in which the Congress and the Presidency, even when they are controlled by the same party, can become deadlocked over policies. Still, the lessons of these countries can inform the development of a measurement system in the United States.[9] These lessons generally focus on two areas: the progress that has been made in improving the measurement of results, and ties that have been developed between measurement and the

7. General Accounting Office, *Performance Budgeting: State Experiences and Implications for the Federal Government,* GAO/AFMD-93-41 (February 1993).

8. Ibid., p. 1.

9. Unless otherwise noted, the conclusions concerning the developments in these countries are based on Allen Schick, "Budgeting for Results: Recent Developments in Five Industrialized Countries," *Public Administration Review*, vol. 50, no. 1 (1990), pp. 26-34.

budgeting and financial management processes.

Measuring Results

Each of these five countries continues to develop better measures, with differing levels of success. In many cases, the performance measurement efforts in these countries have accompanied a transfer of responsibility from the central government to the line agencies, which commit themselves to concrete performance targets in exchange for control of a specified level of resources. In Great Britain, experience thus far suggests that the objectives have not been expressed precisely enough to allow assessment of their achievement. Australia's progress in developing indicators has been confined mainly to the more easily measurable efficiency and work load targets. Improving measurement continues to be an important objective in Australia.

The movement toward performance measurement in New Zealand is part of a broader effort to reform the public budgeting and accounting systems.[10] First, the government of New Zealand switched the budgeting system from an emphasis on inputs to an emphasis on outputs. Second, the need to assess the performance of all departments gave rise to a move from cash to accrual accounting.[11] These changes were accomplished through the passage of the Public Finance Act 1989.[12] Under the system established by this act, the government specifies its outcomes (objectives), and individual agencies are responsible for designing programs around outputs to achieve these goals.

Budget Reform

In each of these jurisdictions, the relationship between performance measures and the budget is tenuous, although in many cases (particularly Great Britain and Australia) the focus of budgeting shows some signs of shifting from measuring work load to measuring results. In Great Britain, the linkage of performance measures and budgets has been promoted by the publication of these measures in the annual Public Expenditure White Paper; the paper for 1988/1989-1990/1991 included more than 1,800 measures.

Canada and New Zealand are exceptions to this trend, for different reasons. In Canada, the performance measurement system has made no inroads into the budget process. New Zealand has made progress, but only in the area of measuring activities or outputs; the budget process ignores outcomes at the agency level. Agencies in New Zealand are only held accountable for outputs; it is the responsibility of the government (the party that controls the Parliament) to understand how outputs are related to outcomes. Unlike Britain and Australia, there is an explicit tie between measures and resources in New Zealand, but only at the level of outputs, not outcomes.

Finally, the Swedish performance measurement effort is linked to the budget through a three-year budgeting system in which agencies are given greater flexibility to use funds in exchange for more information on performance. Agencies have been given the opportunity to receive three-year budgets in exchange for a more detailed budget justification process. Each agency that participates in the system must submit an annual outcomes report in the two intervening years of the three-year period. At the end of the three-year cycle, the agency undergoes a new, thorough justification process.

10. For a description of the New Zealand system, see Frances Goldman and Edith Brashares, "Performance and Accountability: Budget Reform in New Zealand," *Public Budgeting and Finance*, vol. 11, no. 4 (1991), pp. 75-85.

11. Accrual accounting recognizes costs of decisions when spending is committed to, rather than when cash flows out of the Treasury.

12. For an example of financial reporting under the accrual accounting requirements of the Public Finance Act 1989, see *Financial Statements of the Government of New Zealand for the Year Ended 30 June 1992* (October 1992).

Chapter Three

Past Federal Efforts to Link Performance Measures and Budgeting

The current effort to link performance measurement to budgeting more effectively is a logical successor to three other efforts by the federal government in the past 40 years to tie performance to budgeting on a governmentwide scale: performance budgeting, program budgeting (including primarily the planning-programming-budgeting system, also known as PPBS), and zero-based budgeting (ZBB). These systems had one basic characteristic in common. They represented, in part, an effort to increase the level of information available for federal budgeting and management. At least two of them (PPBS and ZBB) attempted to budget on the basis of program effectiveness. In general, these systems achieved less than their designers envisioned. In particular, the effort that has gone into these systems did not change the way that federal resources are allocated.

Performance Budgeting

The first Hoover Commission (1949) proposed that more data be integrated into the budget process.[1] This proposal, which was referred to as "performance budgeting," was designed to allow managers to develop measures of work load and cost-effectiveness. Performance budgeting was clearly aimed at increasing efficiency. The objectives of programs were taken as a given, and the purpose of performance budgeting was to determine the least costly method of achieving them.

Performance budgeting indicated a shift from budgeting based on expenditure "control" (as was the case with the early line-item budgets, for example) to budgeting based increasingly on "management" concerns. The emphasis was not on making governmentwide budgetary trade-offs; rather, it was on measuring work load on an agency basis. The focus was on the work to be done, not on the usefulness of the objectives themselves.[2] Although performance budgeting was never adopted as a governmentwide budgetary process, it is significant because it emphasized the integration of program information and budgeting. This emphasis was to be continued in future reform efforts.

Program Budgeting

The second type of broad-based reform, program budgeting, had its federal heyday dur-

1. Commission on Organization of the Executive Branch of the Government, *Budgeting and Accounting* (1949).

2. Allen Schick, "The Road to PPB: The Stages of Budget Reform," *Public Administration Review*, vol. 26, no. 4 (1966), pp. 29-30.

ing the administration of President Lyndon Johnson. By contrast with performance budgeting, program budgeting was explicitly focused on budgetary choices among competing activities. Although performance budgeting is designed to discover which of several ways is the most efficient method of accomplishing a given objective, program budgeting treated the objectives themselves as variable. As such, program budgeting is not a management system, but a resource allocation system. It is a specific alternative to the traditional government budgeting system, which focuses on marginal adjustments to the status quo.

Although program budgeting has been attempted at various levels of government, the federal effort is embodied primarily in the mid-1960s effort by President Johnson to put into effect a planning-programming-budgeting system throughout the federal government. PPBS was based on a successful effort by Secretary of Defense Robert McNamara in the 1960s to put PPBS into practice in the Defense Department.

President Johnson ordered the Defense Department's system to be employed throughout the federal government in 1965. By 1971, however, requirements that agencies include the PPBS system data were suspended, and PPBS ceased to exist, at least explicitly, in the federal government.

Reports and evaluations indicate that the PPBS was much less successful in civilian agencies than in the Defense Department, in part because it lacked administrative commitment to integrating it fully into civilian agency management systems. Although Secretary McNamara clearly was committed to putting PPBS into effect in the Defense Department, in most other cases the system was treated as an annual reporting requirement rather than as a management or budgeting opportunity. Several other arguments have been made for the failure of PPBS to catch on:[3]

- It was introduced across the board without much preparation.

- It flew in the face of existing budgetary traditions and relationships; in particular, many people strongly objected to the suggestion that the budget process, which is inherently political, could be made "rational."

- It was not given adequate resources, and top managers were not entirely committed to it.

- Good analysts and data were in short supply, and they were necessary to produce the kind of information crucial to the success of PPBS.

- Because PPBS required a review of all activities in each year, the reform caused so much conflict that the political system was incapable of handling it.

- PPBS made a clear assumption that efficiency was the primary value to be considered in evaluating the usefulness of programs, which was not generally agreed to and is a difficult assumption to make in the case of a public program.

- PPBS was an executive budget system, and largely ignored the role of the Congress in the budget process. The result was that the Congress ignored the system in favor of its established procedures.

PPBS was unable to fulfill its promise as a budgetary reform. In particular, it did not change the budget process from one that focused primarily on making small incremental changes to one that could make comprehensive alterations. Like performance budgeting, however, PPBS increased the role of analysis in budgeting and led to better-informed budget decisions. This increase has continued long after PPBS has faded from view.

3. Unless otherwise noted, each of these arguments is taken from Allen Schick, "A Death in the Bureaucracy: The Demise of Federal PPB," *Public Administration Review*, vol. 33, no. 2 (March/April 1973), pp. 146-156.

Zero-Based Budgeting

The Carter Administration brought its own model of budget reform to Washington in 1976 in the form of zero-based budgeting. ZBB, which President Carter had used as governor in Georgia, was designed to permit comprehensive review of the federal budget, as opposed to the traditional, more incremental approach.

In a "pure" ZBB system, instead of concentrating on budgetary changes at the margin, all programs are evaluated each year. The process of arriving at a budget is literally to start from scratch. At the federal level, that would require answering such questions as, "What if we didn't have an army and navy?" or "What if Social Security did not exist?"

In practice, no government ZBB system, including the federal one, went this far. President Carter's Office of Management and Budget used a variant of ZBB in which agencies were asked to rank their programs within some funding limits. The question thus became, "What if the Department of Defense only received 90 percent (for example) of the current year's funding?" This approach, in theory, forces the agency to assign priorities to its programs in order to decide what is in the "base" (the 90 percent, in this case) and what is to be identified as a possible reduction.

In fact, the federal version of ZBB that was practiced in the late 1970s was very similar to the latter kind of system. Approximately 10,000 "decision packages" were prepared each year. Each package included three different funding levels--a reduction level, a current-services (baseline) level, and an enhancement level that offered upgraded services. Decision packages were to be ranked by importance. In practice, the reduction level assumed a minimum funding level of 80 percent to 85 percent, and budgetary decisions concentrated on the proposed additions to that level. This meant that, rather than starting from a zero base, ZBB concentrated on marginal additions to an established (albeit lower) base, which is also how traditional budgeting works.

In addition to not being zero based, federal ZBB was criticized as being excessively time consuming and unrealistic. For example, many programs were bound by statute to provide a given level of service; administrators of these programs were hard pressed to identify realistic ways to reduce funding levels. ZBB, in practice, eliminated or reduced very few programs. President Reagan abandoned the experiment in 1981.[4]

Common Themes of Budgets and Management Reform

Each of these reforms represents an attempt by a President to institute a "top-down" solution to federal budgeting and management. The reforms can be lumped together into "effectiveness budgeting," which is characterized by an effort to define important objectives and tie their achievement to the means of attaining them. While each of the reforms discussed in this chapter traveled a different road to get there, each (with the exception of the earliest reform, performance budgeting) was concerned with linking the use of resources and the outcomes that result from their use.[5] In addition, with the exception of performance budgeting, each was focused specifically on agency program or budget outcomes. Analyses of the various reforms and comparisons among them suggest some common themes.

---

4. Robert D. Lee and Ronald Johnson, *Public Budgeting Systems* (Rockville, Md.: Aspen Publishers, 1989), p. 86.

5. The term "effectiveness budgeting" is taken from James M. Harkin, "Effectiveness Budgeting: The Limits of Budget Reform," *Policy Studies Review*, vol. 2, no. 3 (August 1982), pp. 112-126.

These "rational" budgeting systems generally proved antithetical to those who had some stake in the status quo process. Some critics of rational budgeting systems argue that they fly in the face of political norms and therefore cannot be made to work in a pluralist society. According to this argument, since budgeting is about allocating resources, it is inherently a political process. If that is true, the notion that it is possible to arrive at a rational allocation of resources that is somehow "above" politics is unrealistic. Furthermore, the political system deals best with budgetary schemes that reduce, not emphasize, conflict.

Traditional budgeting has the virtue of forcing fewer choices than rational budgeting. It does this by concentrating on a minimum number of decisions, rather than opening the entire budget to reexamination, and by requiring few explicit decisions to be made about how much to give to one group and take away from another. A large portion of the budget is not subject to detailed annual review.[6] Conversely, proponents of these reform systems argue that it may be desirable to make politically expedient decisions more difficult. These politically expedient decisions are usually exemplified by the political system's paying excessive attention to local concerns in allocating funds.

Each of these systems, though ostensibly either "top-down" or consensual in design, suffered from the lack of clear agreement between superiors and subordinates over goals and objectives. Objectives are very hard to determine and measure in the public sector. In order to budget on the basis of performance, one must have an objective. Different actors in an agency's environment, such as the President, the Congress, interest groups, and agency managers, may have different perceptions of the agency's goals or how they should be accomplished.

Each of the budgeting systems reviewed suffered from an inability to achieve agreement between these actors on goals and objectives. Support by top managers and the White House must be continuous--it is not enough to announce the reform and then turn it over to agencies for carrying out. Lacking that agreement, it is difficult to measure performance, much less budget, on the basis of it.

"Top-down" systems run the risk of running into a hostile bureaucracy. It is necessary to involve the agency staff, whose support will be needed in order to carry out the program. In addition to the difficulties of achieving consensus on goals and objectives, "top-down" systems often prove threatening to agency management. If staff members believe that a system is being established to "punish" nonperformance (by cutting budgets, for example), it may be difficult to get their cooperation. That may lead to the development of self-serving indicators, falsification of results, or other counterproductive actions on the part of staff administering the program.

Each of these systems had an almost overwhelming need for data in order to survive. With these systems, particularly PPBS and ZBB, the necessity of producing yearly the data and paper justifying or analyzing a full range of government programs was more than the system could handle. Indeed, many analysts would argue that PPBS and ZBB died of their own weight. A process that produces mounds of paper that is neither read nor used for decisionmaking is inherently inefficient and breeds cynicism among participants.

Each of these systems may have had an unintended consequence of increasing the demand for analysis, the capability to do it, and its use in devising solutions to public policy problems. Indeed, this may be viewed as the most lasting legacy of the "rational budgeting" movement. Agency planning offices suddenly found that their products were in demand. A new demand also developed for personnel with policy analysis skills. As a result, even after these rational budgeting systems had faded from view, their impact could be seen in the

---

6. Aaron Wildavsky, "A Budget for All Seasons? Why the Traditional Budget Lasts," *Public Administration Review*, vol. 38, no. 6 (November/December 1978), pp. 501-509.

staff that were hired to do policy analysis. Not only have these past efforts increased the ability of the bureaucracy to evaluate program success, they may make it easier to put future systems into effect.

Mandating effectiveness budgeting has pitfalls that must be explicitly recognized in designing any system that is intended to link measurement and budgeting. Experience yields three important lessons. First, commitment to the measurement system by agency personnel at all levels is necessary if it is to be successful. Second, the data that are collected must be used. Third, the design of budget systems must take into account the difficulties in linking the measurement of outcomes to budgeting. Time must be devoted to clarifying these linkages.

Chapter Four

The Experience of Federal Agencies with Performance Measurement

Many federal agencies already use measurements to assist with their day-to-day management, their budget justifications, and their reporting to external constituencies. Important issues for the current debate are the extent to which measurements are already used in federal agencies and the insights that current practice can provide about the potential for using performance measurement for budgeting. In examining these issues, two questions arise that will be addressed by reviewing the results of a General Accounting Office study of federal practices and case studies of individual agencies. First, to what extent are federal agencies focusing attention on the results of programs rather than only on activities? Second, to what extent are performance measures used for budgeting? Knowing about the state of performance measures within these agencies should reveal the implications of using measurement throughout the agencies of the federal government.

GAO Study of Performance Measurement

In response to a request from the Chairman (Senator John Glenn) and Ranking Minority Member (Senator William Roth) of the Senate Committee on Governmental Affairs, GAO surveyed federal agencies to evaluate the state of performance measurement in the federal government. Surveys were completed by 102 agencies, each of which had more than 1,000 employees or over $500 million in annual outlays. These agencies comprised 87 percent of all federal employees and 92 percent of total outlays in fiscal year 1990.

GAO asked agencies to report on what performance measures (if any) they use as a part of their operations and how they use them in management. The survey concentrated on the types of measurements used, their users, the impetus for their development, and top managers' use of and satisfaction with the measurements.[1]

Agencies reported to GAO that they are using a wide array of performance measures. Although a majority of the agencies said they had a strategic plan that sets out goals or objectives, only about 50 percent use their measures to gauge their progress in achieving those goals. The majority of agencies reported that they collect results-oriented (outcome) measurements.

Seventy percent of agency managers who responded said that performance measures help them make budget decisions, manage programs, assure accountability, and measure program results or outcomes. There was also general agreement (over 80 percent) among respondents that their organization's perfor-

1. These results are reported in General Accounting Office, *Program Performance Measures: Federal Agency Collection and Use of Performance Data*, GAO/GGD-92-65 (May 4, 1992).

mance measuring methods are timely, reliable, or detailed in content.

Agency reports to GAO, however, may have overstated the use of performance data by federal agencies. There was, for example, some inconsistency in agency responses concerning the use of outcome measures, which raises questions about whether agencies understood the survey queries or whether their responses truly reflect their use of performance measures.

In fact, GAO questioned the validity of the survey results and attempted to verify responses by visiting several agencies and trying to determine just how much these agencies were using performance measures. The results of this follow-up by GAO confirmed investigators' impression that the survey data generally overstated the use of performance measures by federal agencies. That is probably the result of the limitations of survey research to adequately assess practices in complex, hard-to-define areas.

CBO's Detailed Case Studies

The Congressional Budget Office conducted several detailed studies of agencies and their use of performance measurement. The agencies were selected on the basis of discussions with staff overseeing the GAO survey, and on other reports concerning which federal agencies were already using performance measurement. Furthermore, CBO attempted to choose a range of agency types that would represent federal agencies as a whole. The six agencies evaluated were the Environmental Protection Agency, the Internal Revenue Service, the Department of Labor's Employment and Training Administration, the Farmers Home Administration, the Department of Defense (business operations), and the Public Health Service's Healthy People 2000 program.

The case studies addressed three general issues:

- The extent to which agencies have developed clear objectives or goals for their programs, since performance measures have limited usefulness without such plans in place;

- The types of measures that the agency is using and how they are calculated, and how much the measures relate to the agency's mission or goals; and

- Whether these agencies are using performance measurement data as a tool to facilitate their budget decisionmaking processes.

Environmental Protection Agency

The Environmental Protection Agency (EPA) has used various measures of performance for the past decade. The agency also uses an annual planning process that outlines goals in the measured areas. In the past, the measures have focused on activities (for example, handing out permits) rather than outcomes (improving the environment).

The goal of the EPA planning process is to reshape priorities so that the agency can concentrate on problems that pose the greatest environmental risk. The agency defines performance targets and goals covering several years and is developing measurements that can adequately gauge agency progress.

This effort has tried to change the focus of performance measurement from one of regulatory inputs to one of environmental changes. Some of the previous emphasis on measuring regulatory activity arose from EPA's need to conform to statutory requirements. By focusing on environmental outcomes instead of regulatory activities, EPA employees may now use statutes as a tool to achieve desired environmental outcomes rather than view adherence to the law as an end unto itself.

According to agency staff, four kinds of indicators are best suited to measure EPA performance;

- o Activities undertaken by the programs themselves;
- o Releases of pollutants into the environment;
- o Ambient pollution levels;[2] and,
- o Information about environmental effects.[3]

Although the first type clearly measures output, the other three measures are part of a conscious effort by EPA to expand its use of outcome measures. EPA believes, however, that output measures can still play an important part in gauging agency progress toward measurable environmental goals. Further, EPA staff noted that this planning process has made agency employees more willing to have their performance measured against environmental outcomes. They noted that the relatively long tenure of past EPA administrator William Reilly made using the outcome-oriented approach much easier.

One problem that plagues EPA in the design of its outcome measures is the question of causality. In many instances, outcome measures are difficult to design because it is hard to ascertain whether EPA "causes" marginal changes in the environment. For example, it is difficult to take into account all of the variables that contribute to air quality.

EPA budgeting has been tied very closely to its system of output measurements. Agency managers have been able to tie a given level of regulation to the work levels needed to carry it out. EPA has not used outcome measures for budgeting, which would imply a very different focus. EPA hopes to target its resources toward areas that are assessed as "highest risk" and divert precious money away from relatively high-priced but "low-risk" problems such as Superfund cleanups. EPA plans to design much of its 1994 budget around these risk reduction goals instead of the old output-based funding targets.

Internal Revenue Service

Although the Internal Revenue Service (IRS) has used various performance measures since the 1960s, it started a strategic planning process in earnest in 1986. Three major objectives have been defined as keys to the success of the IRS:

- o Increasing voluntary taxpayer compliance;
- o Reducing taxpayer burden; and
- o Improving productivity and customer satisfaction.

During its strategic planning process, the IRS discovered that most of its measurements are of outputs, such as the number of audits carried out or the number of returns processed in a year. These measures enabled the IRS to compare its performance in these areas with that in previous years, and it often used them to set targets for future years. But these data were not adequate for informing the IRS about how its actions were affecting taxpayers.

The IRS has developed measures of results on only a limited basis. For example, it developed a system of measuring rates of accuracy for toll-free telephone tax assistance, refunds, adjustments, correspondence, and the processing of returns. In addition, the IRS hopes to design methods of measuring citizen satisfaction (via surveys) and voluntary compliance. Although the current set of measurements emphasize process and taxpayer service, a few focus on the results of IRS enforcement activ-

2. This indicator records conditions in the environment caused by the release of pollutants that may contribute to adverse environmental effects.

3. This indicator records effects on human health and ecological damage.

ity. In addition to measuring the IRS's ability to encourage taxpayer compliance through assistance and education, measurements of enforcement would help in finding out whether the IRS's actions are actually increasing it.

Although performance indicators are reported in the President's budget and to the Congress with budget justifications, the IRS has not used them directly to influence the allocation of resources. The IRS is moving to develop more and better measures of performance for management.

Employment and Training Administration

Through its Employment and Training Administration, the Department of Labor manages the Job Training Partnership Act (JTPA). The purpose of this 1982 act is to give disadvantaged youths and adults important job skills and help place them in jobs.[4]

To assist in carrying out JTPA's purpose, the act calls for mandatory performance standards, monetary incentives to reward local programs that exceed their standards, and sanctions for those that fail to meet their standards. Thus, in addition to its basic allocation, a local program can receive additional funds to expand or improve its program.

Performance standards for programs serving adults measure the employment status and earnings of those who complete the program. These data are collected approximately three months after individuals leave the program. Previously, JTPA measured success based on the ability of programs to place participants immediately after they complete their programs. But that created counterproductive incentives. Many programs concentrated on enrolling participants who were the most prepared for employment, ignoring those who were more disadvantaged and in need of training. This practice is commonly referred to as "creaming." JTPA revised its performance measures in 1990 to emphasize job retention, as measured by the proportion of those who complete the program and are employed 13 weeks later. This measure provides for a longer-term assessment of the program's effect. The creaming problem remains, although amendments to JTPA have attempted to address it by targeting funds toward various disadvantaged groups.

The annual planning process at JTPA is decentralized. Each state combines the plans of its individual local or service delivery areas, or SDAs, into a single statewide plan, which is sent to the national office for approval. (SDAs are the geographic subunits responsible for carrying out JTPA activities.) The national office evaluates these plans on the basis of their compliance with statutes, regulations, and consistency with the program's overall goals; it does not use them to monitor performance or ability to meet planning targets.

To encourage SDAs and states to perform at or in excess of their yearly performance targets, JTPA has a built-in system of monetary incentives. Each state must set aside 6 percent of its Title II-A allocation for rewarding local programs and providing technical assistance to those that need to improve their performance. Each governor controls the formula for distributing the 6 percent set-aside among SDAs within a state. Within certain limits, governors may give "bonuses" to programs that perform well. Governors are authorized to impose sanctions, such as reorganization, if an SDA falls below performance standards for two years. Most governors would rather give SDAs technical assistance than impose sanctions. Performance measures are not linked to JTPA's resource allocation process. The vast majority of SDA funding is based on need.

Furthermore, JTPA managers do not believe that the resource allocation process should be linked to performance standards. They believe that performance measures are more of a tool for agency management and less of a tool for budgeting. Their experience suggests that a performance-based budgeting ap-

4. P.L. 97-300, Sec. 2. (Job Training Partnership Act).

proach can produce unintended effects. Moreover, comparisons among human resource programs that are based solely on performance outcomes do not necessarily identify the most effective programs. Program evaluations, conversely, can track longer-term program effects for a given cohort of participants. These evaluations represent more in-depth attempts by researchers to determine the true effects of a program, often by tracking the experiences of a small group of participants. Thus, they can more easily demonstrate the marginal effect that a given program may have on clients. The data from such studies, however, are often not timely; program design and operation often change before results become available. For this reason, the results of evaluations are also limited in their ability to influence resource allocation.

Farmers Home Administration

The Farmers Home Administration (FmHA), founded in 1935, originally assisted family farmers in regaining self-sufficiency after the Depression. More recently, the Congress has extended the responsibility of FmHA to cover nonfarm programs in rural areas as well. FmHA makes direct loans, guarantees loans made by private lenders, and makes a limited number of grants.[5]

In its June 1992 FmHA Strategic Business Plan, FmHA set out a list of goals establishing agency priorities for the next five years. The report reflects FmHA's most important goals in the short run--improving its credit quality, better managing its loan portfolio, and making sure that it serves as the lender of last resort for farmers.

On the basis of these national goals, FmHA sets performance targets for each of its 50 state offices. If the state offices meet their goals, FmHA will have achieved its national performance targets. FmHA's state and national offices negotiate the targets for state performance. The national office notifies the states when they must take corrective action to achieve performance target levels and gives state offices flexibility to design appropriate strategies to achieve these targets.

For the most part, FmHA uses output measures. These measures include the number of annual performance reports received by state FmHA offices, the number of loan guarantees as a percentage of total obligations, or the number of properties sold. One measure that could be called an outcome measure, given the stated goals of FmHA, is the number of borrowers who change from FmHA lending to other forms of credit. FmHA calls that credit "graduation." In other words, borrowers have improved their credit enough to be able to borrow from private sources instead of having to use FmHA (the "lender of last resort").

FmHA has developed a scoring system to evaluate the performance of each state. The national office links the state's performance rankings directly with its bonus system for upper management. The better a state does in achieving all of the FmHA goals, the larger the size of its bonus pool.

Even though the bulk of FmHA funding is awarded to states on the basis of legal formulas (the variables that contribute to the calculation of appropriate funding levels include the economic and demographic characteristics of a state's rural areas), FmHA believes that the achievement of state objectives will save the government money by reducing loan losses. In other words, by pushing states to reduce delinquencies and encourage borrowers to "graduate" to private lenders, FmHA is trying to reduce the costs of its programs as well as the loss that failed loans impose on the U.S. Treasury.

5. President Clinton's budget for fiscal year 1994 recommends that the name of the FmHA be changed to the Farm Service Agency. This new agency would combine the FmHA's current functions with various other activities not now under its jurisdiction.

Department of Defense--Business Operations

The Department of Defense (DoD) is the largest federal agency, employing more people than any other department. The department uses various surrogate measures to assess progress toward its outcomes. For example, while the ultimate outcomes of defense are concerned with national security, many of the measurements that the department uses gauge military readiness. Many of these measurements are classified, and there is no explicit linkage between these macro-level measures and the budget.

The department has, however, been defining measures of both efficiency and effectiveness for its internal service, or business, activities. This process is tied to the development of a unit cost structure for DoD and is also associated with the development of the Defense Business Operations Fund (DBOF) and the application of the Chief Financial Officers Act.

The DBOF was established on October 1, 1991. The purpose of the fund is to consolidate all revolving funds within DoD, with the goal of giving department managers the management tools and information necessary to provide services at the lowest cost. Clients of the fund (the military services) are to be charged rates that recover the full cost of providing a given level of service. Given that DoD's budget is being reduced, this effort is particularly timely. For fiscal year 1992, the fund had $69 billion in collections and $72 billion in disbursements.

DoD has recently published guidelines for the development of performance measures for the DBOF. These measures will focus on both efficiency and effectiveness. The department is planning to use fairly standard measures of efficiency (cost per unit of output) for the DBOF. Effectiveness measures will focus on timeliness (how quickly the item is delivered), accuracy (assuring that the right item is delivered), and customer satisfaction. The specific measures are still under development.

The measures are being developed in response to legislated requirements. Not only does the CFO act require chief financial officers of agencies to develop performance measures, but the fiscal year 1993 Defense Authorization Act specifically required performance measures for the Defense Business Operations Fund by March 1, 1993.

The key objective of the DoD "performance budgeting" initiative is to create incentives for managers to reduce costs but still achieve operational performance goals. The initiative is also aimed at focusing strategies for improving the processes for considering capital investments and other policies, and for improving efficiency and effectiveness. In practice, this means that targets will be set for operational performance, and reducing costs is a desired outcome only insofar as service does not suffer. As the DoD effort is still embryonic, it is too early to tell whether the agency will be able to achieve this accommodation of efficiency and effectiveness.

The Public Health Service's Healthy People 2000 Program

In 1976, Title XVII of the Public Health Service Act (PHS) directed the Department of Health and Human Services (HHS) to establish goals for the prevention of disease and the promotion of strategies to achieve those goals. The Office of Disease Prevention and Health Promotion, also created under the act, was given the task of coordinating prevention policies and programs within HHS that help achieve these goals. *Healthy People: The Surgeon General's Report on Health Promotion and Disease Prevention*, published in 1979, established broad national goals for reducing mortality rates in four age groups from birth to age 65. By 1990, which was the target year for the achievement of these goals, three out of the four had been achieved. Thirty-two percent of those national objectives were achieved by 1990, and another 34 percent of the objectives showed progress.

The Public Health Service repeated this process in 1987, designing a program called Healthy People 2000, which established an extensive set of national health objectives to be achieved by the turn of the century. Healthy People 2000 has three broad goals:

- To increase the span of healthy life for people in the United States;

- To reduce (and finally eliminate) health disparities among subpopulations of the United States; and,

- To give all people in the United States access to preventive services.

Healthy People 2000 work groups developed 300 measurable objectives that can be separated into 22 priority areas. These measures can be categorized into three groups:

- **Health outcome measures**, such as the reduction of deaths caused by heart disease. The quality of such data is relatively accurate, but the direct effect of PHS actions on such statistics is hard to assess.

- **Risk reduction measures**, such as immunization rates or the number of people who quit smoking. The data for this group of measurements is less exact than that for outcome measurements because much of the information is from self-reported health surveys.

- **Service and health protection measures**, such as the percentage of women who receive prenatal care in the first trimester of pregnancy.

Although the data on the quality of service and health protection is not as good as in the other two measurement groups, PHS actions in these areas are most likely to have the greatest direct effect. This points up a recurring paradox for PHS: areas in which its actions can have the greatest effect on health may be difficult to measure, and areas that may be easiest to measure are those on which PHS's actions have the least demonstrable marginal impact.

It is difficult for PHS to link performance and overall health results directly, and thus it is doubly difficult to bestow rewards or punishments on the basis of performance measures. The task is further complicated by the fact that the national results in these health areas are affected by the actions of the private sector as well as by those of all three levels of government.

The linkage between these measurements and budgeting or resource allocation decisions is still in the developmental stage. Internally, PHS prepared its 1994 budget in terms of support for Healthy People 2000 objectives. The current link between measurements and budgets, however, is not refined.

Areas that are easier to measure, however, are better suited to the linkage between measurement and resource allocation. In immunization programs, cost-effectiveness can be analyzed to determine the amount of money needed to improve the immunization rate by a certain amount and thus limit the infection rate for that disease. Such linkages between spending and health effects, however, are much more difficult for incurable diseases and those for which the causes are varied and hard to control. Because of such limitations, the link between budget decisionmaking processes and performance measurement is still under development at PHS.

Lessons Learned from Current Practices of Agencies

Several themes emerge from the experience of federal agencies, including the importance of having a clearly defined mission, the difficulty of developing outcome measures for many government functions, and the often tenuous ties between budgeting and measures of performance.

The Difficulty of Defining Agency Missions or Goals at the Federal Level. Managers of federal agencies can be influenced by several authorities, including the White House staff, the Office of Management and Budget, multiple Congressional committees, political appointees within agencies, and other agency managers. Because clearly defined goals and missions are the foundation on which performance measurement is built, performance measurement initiatives are severely hampered. Without clear goals, it is difficult for an agency to measure how well it is performing its task. Before one can talk about performance measurement, agreement must be reached on agency goals and priorities.

The Difficulty of Linking Performance Measures and the Budget Process. None of the agencies studied used performance measurement to make decisions about the level of resources that the program received in the budget process, although some intended to do so. Performance measurement is used in some cases to determine the allocation of monetary awards to employees or subunits (for example, in JTPA and FmHA), but such reallocations are at the margins. Basic aspects of agencies' budgets are not determined by the relationship between inputs and outcomes.

Measurements of Outcomes Are Few and Far Between. Even in agencies that are regarded as the best examples of federal performance measurement, outcome measures are rare. JTPA may come the closest to the mark with its measurement of job retention, but even it is relatively short-term (measured 13 weeks after the training has ended) and does not give a sense of the marginal impact that the program has on its trainees in the long run. According to JTPA staff, a true analysis of the program's effect would necessitate a program evaluation that addressed this factor. In fact, program evaluations may even contradict the "results" based on performance measures. Moreover, even program evaluations generally cannot separate out alternative explanations for program outcomes.[6] Many agencies (such as EPA) find it difficult to make a direct connection between what they do and what happens as a result.

Function Matters. Certain governmental tasks or functional areas are simply less amenable to effective performance measurement. Programs that provide direct services that are under the control of an agency are easier to measure than others, such as regulatory and grant activities. Within the IRS, for example, certain tasks, such as the accuracy of processing tax returns, are easy to quantify and measure; others, such as the accuracy of taxpayer assistance, are more difficult to quantify or measure in a meaningful way. EPA and PHS have trouble discerning the marginal impacts of many of their programs. In fact, it is difficult to determine what constitutes "good" or "bad" performance by agency subunits in these hard-to-measure areas. In general, performance systems apparently work best where there is direct accountability and where cause-and-effect relationships can be clearly drawn between what the agency does (outputs) and what it wants to happen (outcomes). That does not mean that bad performance should be expected in hard-to-measure areas, but rather that the indicators of success may be more intangible in some cases than in others.

Commitment of Top Managers Directly Affects the Success of Performance Measurement Systems. The cases of FmHA and EPA illustrate the general truism that the success of performance measurement programs (as in any others) depends directly on upper management's commitment to them. In the case of EPA, performance measurement initiatives in the risk reduction area have been aided considerably, according to EPA managers, by the presence of an EPA administrator who has held office for a relatively long time. At FmHA, the presence of a committed, hands-on management team has made the agencywide application of its performance measurement system more

6. For an expanded discussion, see Robert Behn, *Leadership Counts* (Cambridge: Harvard University Press, 1991).

complete and effective. If upper management is not committed to performance measurement, or is transitory, the chances that performance measurement might gain a foothold in federal agencies are greatly reduced.

Conclusion

Performance measurement within federal agencies may not be as prevalent as agencies' responses to the GAO survey would indicate. Some impressive strides, however, have been made in the agencies visited by CBO. Even within those agencies that are good examples of performance measurement, it has been difficult to develop outcome measures and to connect performance measurement and budgeting. Any performance measurement system--whether it is governmentwide or designed on an agency-by-agency basis--should incorporate the lessons learned by these agencies and others that have put performance measurement systems into effect.

Chapter Five

Evaluating Proposals for Measuring and Budgeting for Federal Performance

The measurement of performance by federal agencies and the conversion of the federal government to a system of performance-based budgeting have recently drawn serious scrutiny. Principal among the reasons for the attention are three separate efforts: the reforms of federal financial management; the passage of legislation in the 103rd Congress covering performance measurement; and a review of the operation of the executive branch.

Few people would disagree that an increased focus on the results of federal programs might produce benefits. Translating these hoped-for benefits into reality, however, is difficult. This study concludes that obstacles remain to improving the measurement of federal performance and adopting performance-based budgeting. Even if these obstacles can be overcome, however, no performance measurement effort will succeed without a combination of legislative and executive efforts to improve financial and program management. Legislation is limited in its ability to bring about lasting change but can assist in some areas, such as motivating agencies to measure results, increasing the flexibility of program managers, and mandating that information on program performance be provided to policymakers. Executive branch commitment, however, is crucial in bringing about any real changes in agency management, through providing incentives for agency managers, improving information systems, and tailoring solutions to individual problems.

Given a mix of legislation and executive branch commitment, there are still two main obstacles to improving performance measurement and tying those measures to the budget process. First, it is difficult to develop measures that accurately reflect the performance of federal programs, especially since nonfederal actors largely determine the success of so many of these programs. Second, changing the way that policies are made in order to make them more responsive to the measurement of outcomes would require a stronger demand for them by policymakers than currently exists.

The payoffs that may come about from performance-based budgeting are limited. Using performance measures for allocating government resources requires a transformation of current processes, which are heavily focused on inputs, into an outcome-oriented system. It is impossible to effectuate such a system by simply demanding it; attempts to create one by fiat have failed in the past. Although it may be possible to foster such a demand by improving the information available to policymakers, it is unlikely that simply mandating performance measurement will have any real effect.

If policymakers demand information on the relationship between dollars and results, however, they could bring about real changes in the budget process. Reporting information on the results of government programs in government financial statements can improve the capacity to create such information as well as develop a demand for it. But the largest potential for real payoffs may be in the area of agency management of resources after they are budgeted. This approach is much more fruitful than performance measures for governmentwide resource allocation, and it requires, first and foremost, the commitment of political and career management to supporting the efforts of agency employees.

Efforts Aimed at Improving Federal Performance Measurement

The renewed attention paid to performance measurement in the federal government manifests itself in three main (and related) ways: the application of laws and administrative efforts designed to improve the quality of federal accounting and financial reporting, the introduction of legislation that would increase requirements for reporting performance, and President Clinton's recently announced performance review of federal agencies, which focuses on improving the management of the executive branch.

Current Federal Efforts to Improve Accounting and Financial Reporting

The Chief Financial Officers Act (P.L. 101-576) became law in 1990. It established the Deputy Director for Management at the Office of Management and Budget as the chief financial officer of the federal government and installed chief financial officers in 23 of the largest federal agencies. The act's primary purpose was to improve federal financial management. But the bill also included a provision that explicitly addressed performance measures. This provision requires agency CFOs to develop "systematic measures of performance" for programs in their agencies. It also instructs CFOs "to prepare and submit to the agency head timely reports" and requires that financial statements "shall reflect results of operations." In none of these cases, however, does the act or the legislative history elaborate on how the Congress expects agencies to respond to this provision.

Subsequently, the Office of Management and Budget (OMB) instructed agency CFOs to prepare performance measures for fiscal year 1992 financial reports, which were submitted to OMB on March 31, 1993. OMB initiated an interagency effort, designed to develop indicators that are common to agencies as well as those that are agency-specific. These indicators are to be compared against the agency's mission and to concentrate on both outputs and outcomes. They are to be used both for internal management and to inform the public and elected officials about the accomplishments of federal programs.

OMB reviewed the preliminary results of this effort in a July 1992 memorandum to agency CFOs. With the exception of research and development, the teams for each function were able to develop common measurements. Still, most other agencies had trouble developing good measures of outcomes. The current challenge is to expand the number of agencies that report such measurements and to improve the quality of those that they report (for example, by including more outcome measures).

A related effort by the Federal Accounting Standards Advisory Board is under way. The board is an interagency group charged with developing common accounting standards for the federal government. It has shown interest in using performance measures for financial reporting purposes. In the past, financial reporting has been concerned primarily with the consistent and accurate reporting of costs. In addition to reporting accurately how many dollars are spent, it would be useful to have information on what government funds purchase for the public--the results of public programs. This focus on measurement of results is consistent with emphasis that the Governmental Accounting Standards Board places on service efforts and accomplishments reporting at the state and local level.

Legislative Efforts to Expand the Use of Performance Measures in the Budget

The Congress has also turned its attention to performance measurement. Several Congressional committees have held hearings or issued reports on performance measurement or federal management in general, or on problems existing in particular agencies. In addition, many authorizing committees include performance targets in reauthorization bills. This concern for measuring performance ultimately led to the introduction of several pieces of legislation to mandate the development of performance measures and their use in the budget process. One of these bills, S. 20, passed the Senate on October 1, 1992; it was reintroduced (as S. 20 in the Senate and H.R. 826 in the House) in the 103rd Congress. On May 25, 1993, H.R. 826 passed the full House by unanimous consent. S. 20 passed the full Senate by unanimous consent on June 23.[1] On July 15, the House passed S. 20, clearing the measure for the President's signature.

S. 20, as introduced, concentrated more on the budget process than does the current version. The bill that ultimately was signed into law in 1993 includes requirements for strategic planning, annual performance planning and reporting, and managerial flexibility. It also defines a series of pilot projects as tests for changes such as performance budgeting and managerial flexibility.

Strategic Planning. S. 20 requires each agency to develop a strategic plan for activities under its jurisdiction. The first of these plans would be submitted to the Office of Management and Budget by September 30, 1997. Plans would cover at least a five-year period. Strategic plans would include a comprehensive mission statement, a set of general goals or objectives for the program(s), and a list of any factors outside the agency that may affect achievement of those goals and objectives. Virtually all agencies would be covered by this requirement.

Annual Performance Planning and Reporting. Beginning with fiscal year 1999, the budget for the U.S. government would, under this legislation, include a performance plan. These plans would include performance goals and indicators (quantitative, where possible) enabling the Congress and the public to gauge whether agencies have complied with the goals. Each agency would be required to submit a specific performance plan, on a schedule to be determined by OMB, that would cover the major activities for which it is responsible.

In addition to this information reported in the budget, beginning in fiscal year 1999 each agency would be required to submit program performance reports to the Congress. The first of them would be submitted no later than March 31, 2000. These reports would include information comparing actual with planned performance, a discussion of the success in meeting goals, and remedial action if the goals were not met.

Managerial Accountability and Flexibility Waivers. Agencies would be authorized under the bill to ask OMB for waivers of nonstatuto-

1. The legislative history is as follows. On October 3, 1990, Senator William Roth introduced S. 3154, the Federal Program Performance Standards and Goals Act of 1990. No action was taken on this bill in the 101st Congress, and it was reintroduced as S. 20 at the beginning of the 102nd Congress. Two separate hearings were held by the Senate Committee on Governmental Affairs, and a committee-reported substitute bill was passed by the Senate on October 1, 1992. Several other bills aimed at mandating performance measures were introduced in the 102nd Congress as well. During the 103rd Congress, S. 20 was reintroduced; an almost identical bill (H.R. 826) was introduced in the House by John Conyers, Chairman of the Government Operations Committee. Both the Senate Committee on Governmental Affairs and the House Committee on Government Operations held hearings on performance measurement legislation in the 103rd Congress, and each committee reported bills (to the full House and full Senate, respectively) in that Congress. On May 25, 1993, H.R. 826 passed the full House by unanimous consent. On June 23, 1993, S. 20 passed the full Senate by unanimous consent. The House approved S. 20 on July 15.

ry procedural requirements. The purpose of these waivers would be to remove impediments to agency managers, who would be held more accountable for results in exchange for the removal of administrative regulations constraining their actions. These waivers might cover nonstatutory personnel policies or spending restrictions.

Pilot Projects. The legislation would establish several sets of pilot projects. The first set would include pilot tests for developing performance goals, which would operate in at least 10 agencies for three years, beginning in fiscal year 1994. The OMB Director would be required to report on the results of these pilots by May 1, 1997.

A second set of pilot projects would test the concept of managerial flexibility. These projects would operate for two years beginning in fiscal year 1995; the May 1, 1997 OMB report would include a discussion of them as well.

Third, the bill would mandate pilot projects for performance budgeting. Performance budgets would present varying levels of performance resulting from different budgeted levels. At least five agencies would be required to participate in these pilot projects, which would run for fiscal years 1998 and 1999. The Director of OMB would report on the results of the performance budgeting pilots no later than March 31, 2001.

Executive Branch Proposals

In addition to the Chief Financial Officers Act, several efforts aim to change fundamentally the way that federal services are provided; some of these are targeted to alter federal budgeting as well. One effort, which focuses almost exclusively on management systems, is the move toward total quality management in the federal government. The reform related most closely to the development of performance measures, however, is represented by the "Reinventing Government" movement (after a book of the same title).[2] The book relied primarily on examples from state and local government in calling for changes in the way government agencies are managed.

David Osborne, one of the authors of *Reinventing Government,* subsequently wrote a chapter for the Progressive Policy Institute's book *Mandate for Change* that attempted to advise the incoming Clinton Administration about the reforms necessary to "reinvent" the federal government.[3] Osborne's chapter called for establishing a performance-based budgeting system for the federal government that would replace an emphasis on line-item control with an emphasis on holding managers accountable for results; developing performance measures for federal programs; and budgeting on the basis on performance targets, rewarding agencies that exceed those targets. In short, the Congress would loosen its control over inputs and would be guided by program outcomes. This shift to performance-based budgeting would be financed by setting 1 percent of the funds for each program aside to finance the development of performance measures (the strategy could represent billions of dollars a year). The chapter also called for an overhaul of the civil service system, sunset provisions for federal programs, and more "future-oriented" budgeting.

Building on the theme of reinventing government, President Clinton announced on March 3, 1993, that Vice President Gore will coordinate a review of all federal agencies. This six-month review will identify current federal activities that could be discontinued and will recommend changes in the way particular federal programs are managed. Part of the charge of this review is to develop a recommendation for carrying out "mission-driven, results oriented budgeting." The goals of this effort closely mirror those included in Osborne's chapter in *Mandate for Change.* In

2. David Osborne and Ted Gaebler, *Reinventing Government* (Reading, Mass.: Addison-Wesley, 1992).

3. David Osborne, "Reinventing Government: Creating an Entrepreneurial Federal Establishment," in Will Marshall and Martin Schram, eds., *Mandate for Change* (Washington, D.C.: Progressive Policy Institute, 1993).

general, the focus is on shifting the emphasis of budgeting away from line items and spending money toward a concentration on results and saving money. Because this process is in its early stages, it is unclear what its outcome will be. A final report, however, is due in September 1993.

In a related effort, OMB has during the past year been staging "quality improvement demonstrations" at three federal agencies--the Social Security Administration, the Internal Revenue Service, and the Department of Veterans Affairs. The demonstrations include specific projects in each agency and are intended, in part, to develop more and better measures of program performance in these agencies. OMB expects to obtain the results of these demonstrations in one to three years.

Strategies for Improving Performance Measurement and Performance Budgeting

Experience and current practice provide insights that may be useful in understanding the prospects for performance measurement and performance budgeting in the federal government. First, any effort to measure performance must confront the issue of the appropriate combination of executive and legislative branch action. Second, the complexity of the endeavor suggests that a deliberate approach may be most fruitful. Third, it is important to understand how performance measures might influence the budget process, which requires understanding the limitations of this type of change in immediately influencing policy.

Legislative and Executive Activity

If performance measurement is to gain renewed currency for either management or budgeting, the appropriate combination of legislative and executive activity must be applied. This entails understanding how each is limited in bringing about lasting changes. As Chapter 2 indicated, states have had more difficulty than local or foreign governments in interesting legislative bodies in performance measurement. The difficulty may derive in part from the nature of state government organization. Like the federal government, state legislative and executive branches are both more separated and more fragmented than those of local or foreign governments. Second--and related--passing laws is not enough; without commitment from an executive branch, any "management" effort is doomed to fail.

Nonetheless, passing legislation could yield several potential benefits. First, federal legislation can motivate the executive branch or particular agencies to pursue changes that they might not pursue unilaterally. Second, if increased managerial flexibility is desirable, and if changes in law are necessary to provide this flexibility, the Congress can enact them. Third, laws can suggest particular kinds of information that the Congress would consider desirable for decisionmaking. Fourth, the Congress can suggest areas that the executive branch might study or test, in anticipation of more sweeping legislation in the future.

Legislation has its limitations, however. Ultimately, without executive branch commitment to performance measures for agency management, the practice will not gain widespread acceptance. Further improving the use of performance measures for financial management, financial reporting, or budgeting implies commitment from executive branch budget offices (in agencies and OMB) and from the committees in the Congress that must change their behavior. An important behavioral change would involve a switch from emphasizing front-end, micro-level inputs to emphasizing the big picture--the results obtained from public programs.

A Deliberate Approach May Be Most Desirable

Performance measurement is limited in its ability to bring about substantial change. Some of these factors have nothing to do with either the limits of legislation or executive branch commitment, but simply with the difficulty of measuring government performance itself. Designing systems that appropriately link the goals of programs with their results, and that link results to budgeting and financial reporting, is an extraordinarily complex task. The case studies included in this study illustrate that the greatest obstacle to using performance measures lies in first identifying the measures. That is true at the local and state level, but is particularly true for the federal government, where the success of so many programs is influenced by other actors, including state and local governments, private businesses, and individuals. It is hard to disagree with the goal of improving the measurement of government performance. There is a danger, however, in acting precipitously. Experience with the planning-programming-budgeting system and other reforms suggests that it can be counterproductive to install far-reaching reforms without a fairly complete understanding of their effects.

The pilot projects included in S. 20 represent examples of the type of deliberate approach needed for deciding whether to carry out a far-reaching performance measurement system. One approach to decision-making in this instance would be to use pilot projects to collect information on costs and potential benefits before adopting such a system throughout the government. These pilots should include agencies that represent the full spectrum of federal activities, since failure to do so may result in invalid conclusions. Furthermore, pilots should not be limited to those agencies that already have systems in place, or even those with easily quantifiable measures. In fact, it is most important to include those agencies in which it is harder to identify measurable objectives. National defense and research and development are two examples. Regulatory or intergovernmental activities--in which it would be difficult to sort out the effects of federal efforts, as opposed to state, local, or private initiatives--are another.

A final comment about the difficulty of measuring performance concerns the relationship between performance measurement and program evaluation. Performance measures should not be viewed as ends in themselves. In most cases, they can only offer clues about how well a program is achieving its results. In many cases, a comprehensive evaluation of a program is necessary in order to determine whether it is operating successfully. As noted in Chapter 4, determining the extent to which graduates of a job training program have been placed or have retained jobs does not offer conclusive evidence of the success or failure of the program. For that reason, it is necessary to understand the limitations of performance measures and interpret them accordingly.

Limits on Using Performance Measurement for Budgeting

It is difficult to disagree with the concept of budgeting based on performance. But the experience of other levels of government suggests limited success. Neither the case studies presented in this analysis nor those discussed in the report issued by the General Accounting Office earlier this year uncovered examples in which performance measures had significantly influenced the allocation of resources. There is little evidence that policymakers use performance measures to help them make large changes in the allocation of resources, or that they receive detailed information on the relationship between resources and outcomes. Where performance measurement has taken hold in the budget process--in some state and local governments, for example--agency managers tend to use measurements to manage their budgets. Thus, performance measures are used much more extensively in carrying out the budget than in preparing it.

The fact that performance-based budgeting has not gained widespread acceptance is not reason enough to discount its potential. Even if both legislative and executive branches were committed to it, and a genuine effort was undertaken to transform government budgeting, there would still be obstacles. It is not clear how performance measures should be used to allocate resources. For example, one cannot simply reward those agencies whose measures indicate "good" performance and take resources away from those whose measures indicate "bad" performance. Reality is certainly much more complicated than that. A thorough understanding of all of the factors (including the level of funding) that contribute to performance is necessary in order to understand how performance measures can be used to allocate resources.

Furthermore, performance measures seldom make the task of choosing between different uses of public resources easier. Ultimately, if every program had performance measures, policymakers could understand the trade-offs inherent in spending money on two competing programs. For example, if the choice was between a job training and an air pollution program, we might know that adding $100 million more to the Environmental Protection Agency budget would make the air cleaner by X amount but would cost Y amount of lost wages from workers who had not been trained. If all of these data were available (and believable), that would make decisions more informed. But it still would not necessarily make the choices less difficult.

Budgeting based on performance flies in the face of existing budgeting practice. A system that affords less control over individual line items in order to hold agencies solely accountable for results would be a fundamental change from the current system. Such a system could not resolve the issue of how much money goes to the defense budget and how much goes to domestic spending solely by using performance measures. Budgeting, in a democratic society, is inherently political. No set of budget techniques can substitute for political decisions about "who wins" and "who loses." In fact, the failures of past efforts, such as PPBS, are largely the result of an inability to account for this shortcoming.

The limited potential of performance measures to influence budget outcomes directly, however, does not mean that they have no place in the budget process. There are at least two other ways that performance measures can be tied to the management of resources. Performance measures can be used to assist agencies in the management of a relatively fixed level of resources. For example, an agency's total funding level may be fairly stable, but it may use a performance measurement system to allocate funds among geographic or functional subunits.

Performance measures can also be used to present information on the relationship between inputs and outcomes. That would define for policymakers the relationship between given levels of inputs and the results expected from them. For example, the Internal Revenue Service might present information on how much faster it could process income tax refunds for each of several different levels of funding. But it is one thing to present such information to decisionmakers; it is quite another for them to use it.

If performance measures are to be used to influence the allocation of resources, the change is not likely to happen suddenly. Rather, it may be the result of a change in culture that starts with the development of better, valid performance information at the agency level and with the reporting of that information for nonbudgetary purposes. Once this information is in the public domain, it might be more accepted and eventually used by decisionmakers as they allocate resources. If this change has the effect of forcing some spending to face greater scrutiny, it certainly is desirable.

As more attempts are made to tie performance measures to the budget, using the right measurements and collecting accurate information become more important. But the higher the stakes are, the greater the incentives are for people to identify self-serving measures and report misleading data. Without a process of ensuring that the right measures are chosen and reported accurately, performance measurement will never deliver on its promise. When performance measures become only a more sophisticated means of agency justification, they cease to be a useful policymaking tool. The obvious implication is that verifying the accuracy of reported data is an essential part of the measurement process.

The limited potential, in the short run, for performance measures to influence the allocation of resources should not discourage the Congress and the executive branch from continuing to concentrate more resolutely on the results of public programs. In fact, the greatest reward to be gained from the use of performance measures may have less to do with governmentwide budgeting than with the task of using existing resources to improve performance. For example, performance measures can be useful as motivational tools; that is, they can encourage people to achieve performance targets. Ultimately, repeated use and exposure will result in the development of a culture of performance. Performance targets may not be precisely correct at first, and the measurements may not be either. But encouraging federal managers and employees to think in terms of outcomes rather than inputs or outputs will produce desirable results. Because the measurements will not be right at first, one must be realistic about how much they can be used to influence budgeting in the near term. In short, improving agency performance measurement (because agencies should increase their ability to get the measures right over time) should precede using those measures for budgeting.

Conclusion

Before taking further legislative action, it is important for the Congress to consider exactly how it expects performance measures to change the process of developing and carrying out policy. If mandating performance measures is expected by itself to reform federal budgeting and management, disappointment will almost certainly follow. Executive branch commitment is probably a more important factor than legislation, although legislation aimed at reducing impediments to managers might be helpful as well. Further, several continuing efforts touch on performance measurement: total quality management, carrying out the CFO act, deliberations of the Federal Accounting Standards Advisory Board, legislative efforts such as S. 20, and the Vice President's performance review. These efforts should be viewed as related to each other and should be coordinated. In particular, the effort to identify agency missions and goals should precede the effort to arrive at measures of performance, and the development of performance measures should precede requiring their use for budgeting.

In the final analysis, the budget process is unlikely to be changed substantially until and unless decisionmakers demand and use information on program performance when making decisions about allocating resources. Having this information--if good measures of program success can be developed--is a necessary, but not sufficient, prerequisite to changing the policy process.

587

CONGRESSIONAL BUDGET OFFICE
Second and D Streets, S.W.
Washington, D.C. 20515