

## **EXECUTIVE SUMMARY** **Linking Resources to Results**

The idea that the allocation of government resources should be based on demonstrated program effectiveness is attracting new interest in Washington. In response to this renewed interest, this paper offers for discussion a set of recommendations designed to strengthen and help institutionalize the link between governmental resource allocation decisions and high-quality data on program performance.

In August of this year, President Bush promulgated a new Management Agenda, a central goal of which is to allocate scarce federal resources to programs and managers that deliver results. The Agenda reflects his belief, articulated as the Governor of Texas, that:

“Government should be results-oriented – guided not by process but guided by performance. There comes a time when every program must be judged either a success or a failure. Where we find success, we should repeat it, share it, and make it the standard. And where we find failure, we must call it by its name. Government action that fails in its purpose must be reformed or ended.”

Congress has similarly expressed its commitment to this idea in the Government Performance and Results Act of 1993 (GPRA), a central purpose of which is to “improve the confidence of the American people in the capability of the Federal Government, by systematically holding Federal agencies accountable for achieving results.” GPRA was the first major statutory effort to attempt to link resources to results in 80 years.

Currently, however, linking resources to results remains largely an unrealized goal, as Congressional and executive branch decisions about resource allocation are not routinely based on demonstrated effectiveness. A recent General Accounting Office survey of federal agency program managers, for example, found the absence of a strong commitment by the top agency leadership to achieving results, as well as the infrequent use of performance information in key management activities.

### **Recommendations to maximize government’s effective use of results data in allocating resources:**

- **Establishment of an autonomous, nonpartisan Center for Domestic Program Assessment** to produce high-quality analysis of program and policy performance, presented in timely fashion and in ways decision makers can use;
- **Statutory set-asides for program evaluation**, in order to support high-quality evaluation of individual programs by the federal agencies, and to encourage linking evaluations across agencies for programs addressing aspects of the same problem;
- **New Congressional oversight and evaluation procedures** to ensure full use of evidence on program effectiveness; and
- **New executive branch procedures** to maximize the use of evaluation results.

As discussed in the paper, these recommendations together comprise a systematic and effective strategy for expanding the use of evidence on program effectiveness in decision making on resource allocation and program design. They would enable Congress and the federal agencies to overcome many of the barriers that now prevent or slow the generation and effective use of results data. For example, the new Center would provide:

- expertise in designing and implementing high-quality program evaluations, to enable the agencies to surmount many of the practical difficulties in carrying out such evaluations;
- expertise to help Congress and the agencies interpret and reconcile conflicts in evaluation data, and in claims about that data, which might otherwise prevent its effective use by decision makers;
- effective, high-profile advocacy for high-quality evaluations that can serve as a counterweight to political and other pressures that might otherwise block such evaluations; and
- government-wide policy analysis that can transcend the more narrow, jurisdiction-based focus that individual agencies and Congressional committees now bring to policy analysis.

The Center and the other recommended changes could create a climate of public debate on results that should have substantial influence on policy and budget deliberations. For example, if the Center were now in existence, it could help inform the ongoing Congressional deliberations on reauthorizing the Elementary and Secondary Education Act by providing a detailed assessment of the impact of various education initiatives on student achievement. If the Center had existed in the early 1990s, it could have informed Congressional deliberations on the 1994 Omnibus Crime Control and Safe Streets Act by providing high-quality assessments of existing community policing initiatives, school-based crime prevention strategies, initiatives to recruit and deploy additional police officers, and other crime prevention and control activities.

**These recommendations address a new imperative for effective government.**

The notion of allocating federal resources on the basis of program effectiveness is an idea that has resonated with many policymakers over the years. It was the central goal, for instance, of the Carter Administration's Zero-Based Budgeting initiative, the Nixon Administration's Management By Objectives initiative, and the Johnson Administration's Planning-Programming-Budgeting System. None of these earlier, short-lived efforts should obscure the importance of responding effectively to the impulse that inspired them all. The need to better link resources to results grows steadily more urgent as the size and reach of government, and the public's demand for services, continue to expand. In addition, the recent terrorist attacks on our nation may, over the long run, significantly diminish the government resources available for domestic programs, and thus increase the imperative for existing resources to be deployed more effectively.

What this paper offers, as a starting point for discussion, is a specific, effective strategy to link resources to results. Such a strategy, if it succeeds, would fundamentally improve the performance of government in addressing our most pressing social and economic needs, and make a major contribution to the quality of American life.

## Linking Resources to Results

### A Discussion Paper

Co-sponsored by the Council for Excellence in Government and the Committee for Economic Development

November 2001

#### I. Introduction

There is once again in Washington much talk of the need to relate more resource allocation decisions to the demonstrated effectiveness of government programs. This interest, periodically revived over the years, was brought newly to the fore by enactment of the Government Performance and Results Act of 1993 (GPRA), and given particular focus by the Bush Administration's stated commitment to relate budget and management decisions more directly to program results beginning with its fiscal year 2003 budget justifications. It was further reinforced by the President's recent promulgation of a new Management Agenda, a central goal of which is to allocate scarce federal resources to programs and managers that deliver results.<sup>1</sup>

To achieve this goal, it will be necessary for decision makers to undertake the hard, costly, time consuming, and sometimes politically risky steps to produce and use better performance data. This paper proposes ways to help decision makers take those steps, including:<sup>2</sup>

1. First and foremost, the creation of an autonomous, nonpartisan, entity -- a Center for Domestic Program Assessment -- that would produce high-quality analysis of program and policy performance, presented in timely fashion and in ways decision makers can use. The Center would encourage and support evaluation of individual programs by federal agencies and have the capacity to initiate such evaluations where necessary. It would assess the cumulative performance of programs in multiple agencies addressing aspects of the same problem, in ways that individual executive branch agencies or Congressional committees do not normally do.

If the Center were successful in generating higher-caliber analysis and disseminating it at key junctures in the decision process, it would create a climate of public debate on results that should have substantial influence on policy and budget deliberations. For example, if the Center were now in existence, it could help inform the ongoing Congressional deliberations on reauthorizing the Elementary and Secondary Education Act by providing a detailed assessment of the impact of various education initiatives on student achievement. If the Center had existed in the early 1990s, it could have informed Congressional deliberations on the 1994 Omnibus Crime Control and Safe Streets Act by

---

<sup>1</sup> The President's Management Agenda, Executive Office of the President, Office of Management and Budget, August 2001.

<sup>2</sup> We focus on the domestic program arena because that is the area the authors know best, but the generic issues in national security and international affairs are similar and deserve comparable attention by others.

providing high-quality assessments of existing community policing initiatives, school-based crime prevention strategies, initiatives to recruit and deploy additional police officers, and other crime prevention and control activities.

2. Expanded Congressional oversight on demonstrated results as a more direct input into program design and resource allocation decisions, and the establishment of set-asides in law for high-quality evaluations of net program effects.
3. Greater and more active consideration of results in the executive branch's development of budget and legislative proposals and budget justifications.

The notion of establishing a clearer link between results and resource allocation is hardly new to Washington. Most participants in the program design and budget process at all levels firmly believe they are in fact basing their positions on the best information available to support the potential effectiveness of choices. Still, something must be missing if a law must be enacted to attempt to formalize the link. GPRA itself followed a succession of high-profile efforts in the preceding 30 years, most prominently President Johnson's Planning-Programming-Budgeting System, President Nixon's Management By Objectives, and President Carter's Zero-Based Budgeting. The general demise in just a few years of each earlier effort demonstrates the barriers facing GPRA and the Bush Administration in institutionalizing the concept and practice of linking resources to results.

GPRA may hold more promise than its predecessor executive branch approaches because of its statutory status and its attempt to engage the Congress and interest groups in the required performance planning and reporting processes. So far, however, GPRA has not received significant attention from most agency leaders or most in Congress, has not generated any material increase in the funding or use of high-quality evaluation, and has not increased attention to major management challenges.

None of the earlier short-lived efforts nor the currently limited effectiveness of GPRA should obscure the importance of responding effectively to the impulse that inspired them all. The need for better linking of resources to results grows steadily more urgent as the size and reach of government, and the public's demand for services and benefits, continue to expand. Whether the budget is in deficit or surplus, there are never enough resources to meet all demands. In addition, the recent terrorist attacks on our nation may, over the long run, significantly diminish the government resources available for domestic programs, and thus increase the imperative for existing resources to be deployed more effectively. Improving the link between resources and results is crucial to getting better value for resources invested and real improvement in government performance.

While we are not under any illusions that the proposals discussed here could be quickly or easily brought to reality, we do believe that now is the right time to put them forward and get the debate moving. The President's commitment to this issue is reflected in his new Management Agenda, which seeks to allocate federal resources to programs that deliver results, and in his promise to breathe new life into GPRA. The President's commitment is further reflected in his statement, when he was the Governor of Texas, that:

“Government should be results-oriented – guided not by process but guided by performance. There comes a time when every program must be judged either a success for a failure. Where we find success, we should repeat it, share it, and make it the standard. And where we find failure, we must call it by its name. Government action that fails in its purpose must be reformed or ended.”<sup>3</sup>

Other developments in the policy climate in Washington also suggest the time is right for this initiative. For instance, the General Accounting Office (GAO) is increasingly focused on performance issues, beyond its traditional charter to monitor financial integrity. *Government Executive* magazine annually investigates and grades federal agencies on their performance and, with the support of the Council for Excellence in Government and others, hosts the biggest annual conference on excellence in government. The Senate Committee on Governmental Affairs draws persistent attention to performance issues. Conferences on performance management, measurement, and budgeting by the National Academy of Public Administration and others dot the calendar year round. More than a dozen nonprofit organizations have joined in a Government Performance Coalition to serve as a collective voice calling attention to performance improvement.

With the ideas advanced in this paper, therefore, we hope to build on the current favorable climate for increased focus on results and stimulate discussion and action for more effective ways to realize the goal.

Former Senator Alan Simpson said:

“In politics, there are no right answers, only a continuing flow of compromises between groups, resulting in a changing, cloudy and ambiguous series of public decisions, where appetite and ambition compete openly with knowledge and wisdom.”<sup>4</sup>

We believe that these compromises, or decisions, need to and can be better informed, and thus more effective, if public leaders accept the challenges in this report to develop new institutional and procedural improvements that would better support linking resources to results. If that happens, then the balance in Senator Simpson’s vision can shift significantly, away from “cloudy and ambiguous” and toward “knowledge and wisdom” as the more prominent characteristics of domestic policy decision making.

## **II. The current decision-making process falls short in linking resources to results.**

Ideally, government should base its action on the best available data about the problem at hand; set clear, measurable goals; carry out a logical strategy to achieve those goals that is based on the best available data; track performance from the outset; and take timely corrective action when program data indicate difficulties. In parallel, government should carefully evaluate its actions over a period sufficient to show whether the goals are being achieved, what contributed

---

<sup>3</sup> Op. cit., no. 1, p. 27.

<sup>4</sup> Statement of Senator Alan Simpson, quoted in a *Congressional Record* insert, December 19, 1995, p. S18925.

to success or failure, and what alternative approaches might be necessary. It should allocate resources in significant measure in relation to the performance achieved. Its focus should be on the results, or net impact, of its actions, defined as what happens that would not have happened in the absence of the action, at what cost, and with what degree of cost effectiveness. Each of government's periodic creations of results-based approaches has incorporated these principles.

The reality is often quite different. While virtually every law has "purpose" statements that establish the public policy goals it is intended to achieve, most resulting programs actually focus on input, process, and short-term outputs, whose causal relationship to achievement of the goals is tenuous at best. Programs whose purpose is to raise educational achievement focus on numbers of children "served." Welfare resources justified on the basis of helping the poor get off welfare for the long term are measured in terms of immediate reductions in welfare caseload.

GPRA is intended to redirect the system to its principles by establishing measurable long-term goals, annual performance plans, annual measures of progress, and public annual reports. The data it produces is intended to be a major influence on resource allocation and program design decision making. Though all observers acknowledge that progress has been made, here are two recent assessments indicating how far the system has yet to go:

"Progress [on the utilization of information produced by GPRA] has been disappointing. Unless the results information agencies produce is used to make future management and resource allocation decisions, then GPRA becomes a paper exercise.... Congress and the Administration must demonstrate that results information will be a major influence in future decision-making."<sup>5</sup>

"Based on the results of our government-wide survey of managers at 28 federal agencies, many agencies face significant challenges in instilling a results-orientation throughout the agency.... At 11 agencies, less than half of the managers perceived to at least a great extent, that a strong top leadership commitment to achieving results existed.... At no more than 7 of the 28 agencies did 50 percent or more of the managers respond that they used performance information to a great extent for any of the key management activities."<sup>6</sup>

There are a host of barriers that make achieving the ideal resource allocation and program design process difficult, including most prominently: (1) barriers to the generation of high-quality program and policy evaluations, and (2) barriers to the government's use of what evaluations there may be. The following sections highlight these barriers. The recommendations we offer for debate demonstrate how each barrier can be addressed.

---

<sup>5</sup> Testimony by The Honorable Maurice P. McTigue, Distinguished Visiting Scholar, Mercatus Center at George Mason University, before the House Committee on Government Reform, June 19, 2001.

<sup>6</sup> Testimony by J. Christopher Mihm, GAO Director, Strategic Issues, before the House Committee on Government Reform, June 19, 2001.

### **III. Barriers to the generation of high-quality program and policy evaluations.**

#### **A. Agencies often do not set measurable performance goals against which the agency and its programs need to be evaluated.**

In its review of selected agency strategic plans in 1999, the Council for Excellence in Government found few agencies successful in setting measurable goals. Many agency goal statements are high rhetoric more suitable to broad mission statements than to goals intended to guide the specifics of annual performance plans that lead to achievement of measurable results. For example, the Treasury Department, in its strategic plan published in 2000, identified these goals:

- “Apply sound governmental policy on all relevant governmental issues.”
- “Make wise capital investments.”

Setting measurable impact goals well is not easy for most program areas with decades of traditions of short-term process focus, but measurable goals can be defined. In 1989, President George H. W. Bush and Arkansas Governor Bill Clinton convened a National Governors Summit on Education. They issued six clearly measurable national education goals to be achieved by the year 2000. These included: “By the year 2000 all children will start school ready to learn,” and “By the year 2000 the high school graduation rate will increase to at least 90 percent.”

One can argue that these goals seem as unrealistic as the ones cited above, but here, specific, measurable time and achievement goals were included that lend themselves to strategies for achievement and that permit an interested public to hold the government to account if they are not achieved. This bipartisan coalition did not last into the stage of developing program strategies and funding to achieve the goals, so we can’t know how well the authors might have taken the responsibility to devise workable solutions, or to evaluate their efforts. Furthermore, there was little focused public reaction either to the goals or to the obvious failure to achieve them as the years passed. Still, the episode is worth marking as an example of how relatively clear and measurable goals can be defined by political leadership in a major policy area.

The fact that agencies often do not set measurable goals that require corresponding impact evaluations to measure results suggests that revisions in the government’s goal-setting procedures are needed. This paper will recommend ways in which the Congressional authorization process for domestic programs might be adjusted to help achieve this objective.

#### **B. The fragmentation of goal-setting among the federal agencies creates little demand for the government-wide policy evaluation that is needed to address national needs.**

Even if individual program or agency goals are defined well, in many domestic program areas, they do not by themselves capture all the major efforts that go into meeting a national need. As agency GPRA Strategic Plans (where five-year performance goals are codified) generally reflect, most agencies define their goals either in global terms that cannot be met by the agency’s action alone, or in terms of their programs’ reach, without acknowledging the limits of

the programs to achieve the larger goal. GPRA identifies this issue but has not by itself led to effective responses. GPRA calls for recognition of partners at the federal, state, local, interest group, and private sector level, but the goals are rarely described in terms of the actions needed by multiple actors to achieve them.

For example, the labor market policy goals of the Labor Department rarely take into account the expected impact of the education policies of the Education Department on workforce skill levels or the economic development policies of the Commerce or Agriculture Departments on job creation, all of which should clearly contribute to government's efforts to raise earnings and income. Similarly, the national goals for early education and child care are defined independently by Health and Human Services' (HHS) Head Start program, HHS' separate multi-billion dollar child care program, the multi-billion dollar child care tax credits at the Treasury Department, and the Education Department's several related programs.

The GAO says, "Virtually all of the results that the federal government strives to achieve require the concerted and coordinated efforts of two or more agencies.... Crosscutting program areas that are not effectively coordinated waste scarce funds, confuse and frustrate program customers, and undercut the overall effectiveness of the federal effort."<sup>7</sup>

The same overlap applies to Congressional committee jurisdictions. The four major agencies spending more than \$1 billion each on multiple programs for early childhood education and child care (Education, HHS, Agriculture, Treasury) are overseen by four different committees in the House and in the Senate. An Environmental Protection Agency official asserts that 35 subcommittees oversee that agency's activities.

With such fragmentation of the goal-setting process, there exists very little demand for the rigorous evaluation of programs and policies from a government-wide perspective. To address this issue, this paper will recommend revisions to GPRA and the Congressional oversight process to foster greater coordination among the federal agencies and Congressional committees in setting goals and evaluating programs from a government-wide perspective. In addition, as discussed below, one of the central purposes of the new Center for Domestic Program Assessment would be to provide the government-wide policy analysis and evaluation needed to address national needs.

**C. Carrying out high-quality evaluations requires researchers with significant expertise in study design and implementation, as well as sharp political skills.**

Even well-specified goals by program or across programs in the same area, are of limited value to governance if they are not accompanied by effective efforts to discover whether they are being met. That is the role of rigorous evaluation and careful experimentation and demonstration.

---

<sup>7</sup> Ibid.



There are many forms of evaluation for various purposes,<sup>8</sup> but many researchers and policy analysts concur with GAO's assessment that "Impact evaluations are the most definitive method for determining a program's effect on its participants."<sup>9</sup> Program impact evaluations, as discussed in GAO's assessment, include:

- Studies employing experimental designs, in which eligible individuals are randomly assigned either to a group that will receive services from the program being studied or to a control group that will not receive services from the program. The relevant outcomes of these two groups are measured and compared, and any differences found between the two can be attributed to the program.
- Studies employing quasiexperimental designs, in which the outcomes for a group of program participants are compared with the outcomes for a non-randomly selected comparison group with similar characteristics.

Well-designed impact evaluations are not easy to carry out. They require researchers with significant expertise in study design and implementation, to ensure, for example, that the study effectively measures the key outcomes of interest to decision makers, includes a sample size large enough to generate statistically-meaningful results, and uses control or comparison groups that are statistically equivalent to the group of program participants being studied. Judith Gueron, the president of MDRC – an organization that has implemented many of the large-scale, experimental evaluations of state welfare-to-work programs – also emphasizes that researchers need sharp political skills to ensure that the staff of the program being studied, as well as the study participants, cooperate and assist in carrying out the study.<sup>10</sup>

Even when evaluations are carefully designed and carried out, they cannot be expected easily to find broad agreement, acceptance, and use. As the *Washington Post* reported, "Extensive studies on particular programs often are inconclusive, or the results have been disputed by partisans on both sides."<sup>11</sup> However, Charles Schultze – the former Director of the Office of Management and Budget (OMB) and Chairman of the President's Council of Economic Advisors – argues that, without good evaluations, policy is subject to the "naïveté of the decision maker who ignores resource constraints and believes that virtue alone produces efficiency."<sup>12</sup>

One of the primary reasons why this paper proposes a Center for Domestic Program Assessment, as discussed below, is to provide federal agencies and others with an independent, authoritative source of the expertise they need to carry out high-quality impact evaluations.

---

<sup>8</sup> For valuable survey of the variety and uses of alternative evaluations, see Katherine Newcomer, "Using Evaluation to Support Performance Management: A Guide for Federal Executives," Pricewaterhouse Coopers Endowment for the Business of Government, January 2001.

<sup>9</sup> Early Childhood Programs: The Use of Impact Evaluations to Assess Program Effects, General Accounting Office, GAO-01-542, April 2001.

<sup>10</sup> Judith M. Gueron, "The Politics of Random Assignment: Implementing Studies and Impacting Policy," presented at Harvard Faculty Conference, June 27, 2000.

<sup>11</sup> Glen Kessler, "Finding Federal Spending's Dividends," *The Washington Post*, May 1, 2001, p. A11.

<sup>12</sup> Charles L. Schultze, *The Politics and Economics of Public Spending*, Brookings Institution, Washington, DC, 1968.

**D. Government decision makers often do not allocate sufficient funding for high-quality impact evaluations.**

Through much of the 1960s and 1970s, as discussed by Henry Aaron<sup>13</sup> and Richard Nathan,<sup>14</sup> there was a general increase in interest, funding, and implementation of social experiments and other forms of impact analyses. In part, this was attributable to the discovery of new analytical techniques in economics combined with a belief in the efficacy of social interventions. Later, however, funding for evaluation in federal agencies declined, and the anticipated increase in evaluation activity after passage of GPRA has not yet materialized. After the 1994 Congressional election, the new majority repeatedly reduced research and evaluation budgets, arguing that the Administration had used studies more for advocacy than analysis.<sup>15</sup>

Aaron suggests that the decline is attributable in part to a naïve view of government and the political process among analysts, as well as mixed evaluation findings offered by analysts to policy makers.<sup>16</sup> Nathan follows on this logic by noting that although there is existing demand for evaluation findings and other analyses, there is the problem of advocacy research filling gaps and being taken more seriously than policy analysis, which often takes more time to carry out and is more costly to underwrite.<sup>17</sup>

The late Elliot Richardson noted that the evaluation process was on occasion “politicized by the party in power or fought over by competing constituencies – outcomes not likely to have occurred had the program’s objectives been clear from the beginning. When you also take into account evaluation’s lack of political sex appeal and consequent vulnerability to budget cutters, it is perhaps not surprising that its hey-day was cut short.”<sup>18</sup>

More generally, politicians of both parties in the White House, the agencies, or in the Congress are essentially impatient, wanting data and solutions to problems within their terms of office -- which for many appointees average 18 to 24 months. They are usually unwilling to invest the money and staff effort in the long-term strategies that are often needed to determine the net impact of programs and policies, within one program or across the several programs affecting a given issue.

This paper will recommend steps to increase the resources allocated for high-quality impact evaluations. For example, it will propose modifying Congressional rules to require such evaluations and the needed funding as part of the program authorization and appropriations processes.

---

<sup>13</sup> Henry J. Aaron, *Politics and the Professors*, Brookings Institution, Washington, DC, 1978, p. 151.

<sup>14</sup> Richard Nathan, *Social Science in Government*, Rockefeller Institute Press, 2000.

<sup>15</sup> See, for example, cuts in the research and evaluation budgets of the Food and Nutrition Service and the Education Department.

<sup>16</sup> Henry J. Aaron, “Politics and the Professors Revisited,” Richard T. Ely Lecture, Washington, DC, August 21, 1989.

<sup>17</sup> *Op. cit.*, no. 14, pp. 199 - 205.

<sup>18</sup> Unpublished preface to a report for the Council for Excellence in Government and the Hitachi Foundation.

#### **E. Agencies and grantees sometimes resist rigorous evaluation.**

Fear of unpalatable evaluation results and the resulting consequences if the study is negative (loss of funding or political confidence) can deter enthusiasm for good evaluation within an agency. Negative findings can also forestall the agency's ability to obtain the cooperation of states and other grantees for future evaluations.

Beginning in the early 1970s, the Labor Department took bold steps to carry out longitudinal net impact studies of its job training programs, at a cost of several million dollars per year.<sup>19</sup> After the initial study had run for several years, it was clear that the comparison group approach chosen (matching earnings and employment data from Social Security records) would not suffice. Labor then shifted to a random assignment approach, a far more difficult method, but one that holds greatest promise for clear net impact findings. The new approach produced meaningful results in the early 1990s.

To oversimplify, the study showed that adult programs had small but significant positive net effects (defined as increases in enrollee earnings and employment compared to the comparison group) for some. The findings for youth programs were highly negative. The agency did not have a strong program response to offer to the youth findings (e.g., imposing criteria to fund only successful programs while instituting reforms of others). Both the Administration and Congress cut funding sharply, justifying the action by reference to the study. In subsequent years, local grantees recalled the experience and strongly resisted participating in new studies, blaming the study rather than the poor performance for the loss in funds. To its credit, Labor continues to believe in the necessity for impact evaluations.

The pre-1996 welfare law permitted HHS to waive certain provisions for state welfare reform demonstrations if the demonstrations were accompanied by well-designed impact evaluations. Often a state was more interested in obtaining the flexibility to pursue a model it favored than in the complex evaluation needed to validate the model, studies that might demonstrate that the policy model was not as effective as its supporters hoped. It was thus not uncommon for states to resist the requirement for evaluation and difficult for HHS to hold firm. In some cases, only the requirement for OMB approval assured that some waivers were accompanied by the necessary studies.

Many of this paper's recommendations, discussed below, are designed to surmount such resistance to rigorous evaluation. For example, the independent Center we propose is designed in large part to serve as a counterweight to such resistance, by increasing the public visibility and understanding of evaluation results, highlighting their relevance to program design and funding deliberations, and increasing the incentives for agencies to undertake rigorous program evaluation.

---

<sup>19</sup> The authorizing statute then was the Comprehensive Employment and Training Act, but the later change to Job Training Partnership Act did not materially alter the training program activity (versus the much more controversial subsidized employment programs).

**F. Political pressures on evaluators can be a barrier to high-quality program evaluations.**

Program evaluators are usually agency staff and contractors. They work in a political context and are subject to pressures that can influence their analysis.

The authors of the 1983-1987 *National Assessment of Chapter 1* (the largest federal elementary and secondary education program) cited the greatest challenge in all aspects of the study as the continual need to find the balance between the desires of the Democratic House of Representatives (sponsor of the study) and the Republican executive branch. In the end, they opted for structuring the study design and analytical approach to satisfy the desires of the Congressional staff who had been in place longest and whom they had reason to believe would still be there long after the administration of that day had departed.<sup>20</sup>

One key virtue of the independent evaluation Center we advocate is its ability to be an effective buffer between evaluators and political pressures from inside and outside the program agency. Even the knowledge that such an entity will be providing timely critical comment on an agency's major study would be an effective deterrent to some forms of pressure to slant results.

**G. Existing government analytical agencies, such as the GAO, Congressional Budget Office (CBO), and OMB do not now undertake independent, complex impact evaluations.**

The analytical agencies of the Congress and Executive Office of the President do not actually conduct complex evaluations, though GAO does do some independent research. In the main, they collate, assess, and apply what others have done. For reasons of charter, tradition, and placement, the work of the Congressional Research Service and the CBO (though not the GAO) focuses on analysis, not recommendations. CBO has great power to affect resource level policy through Congressional rules that require consideration of its cost estimates as against Congressional budget resolution levels before passing bills, but CBO does not generally make program or funding-level policy recommendations that might improve the effectiveness of the spending it estimates.

For more than a decade, from the early 1980s to the early 1990s, Congress had the benefit of a remarkable agency, the Office of Technology Assessment (OTA), to provide it with independent analysis of major science and technology issues. OTA was designed to provide analysis free of control of either house of Congress or political party (though there was ample opportunity for member input in its choice of topics and in judging the quality of its products). The story of OTA's demise is muddled somewhat by partisan arguments on both sides, but few seem to deny the value of most of its work to the deliberations of Congress.<sup>21</sup>

---

<sup>20</sup> Beatrice F. Birman and Mary M. Kennedy, "The Politics of the National Assessment of Chapter 1," *Journal of Policy Analysis and Management*, vol. 8, no. 4, 1989, pp. 613-632. See especially the comments by the leading Congressional staffer and the assistant secretary to whom the researchers nominally reported.

<sup>21</sup> See, for example, Peter D. Blair, *Technology Assessment: Current Trends and the Myth of a Formula*, 1994. Also, Warren E. Leary, "Congress's Science Agency Prepares to Close Its Doors," *New York Times*, September 24, 1995.

GAO has traditionally focused most heavily on the essential issues of financial integrity and the old triptych of “fraud, waste, and abuse.” In recent years, especially under the leadership of Comptroller General David Walker, GAO’s emphasis is shifting to assessment of how well programs work and how well government is performing in the context of national need. Still, GAO remains bound by its requirement to respond to the requests of the Congress, which, while subject to some negotiation with GAO on scope and focus, reportedly drive some 95 percent of its work.

In addition, Congressional control of GAO appropriations is a tool that Congress can use to try to influence GAO’s work. For example, in the mid-1990s, the new Congressional Republican leadership cut GAO’s appropriation by 25 percent, based on their perception that it had become a tool of Congressional Democrats. Defenders of GAO countered that its agenda is determined primarily by the party in power, which in preceding years had been the Democratic party.<sup>22</sup> Regardless of the merits of the argument, the incident represents a case in which Congressional control of GAO appropriations was apparently used to penalize GAO for pursuing a particular agenda.

OMB serves in an advisory role to the President, and in that role seeks to learn and apply evaluation work that others have done in advising the President on policy and resource allocation options. But OMB does not undertake its own impact evaluations nor have extensive expertise in high-quality impact evaluations. Furthermore, OMB is constrained, by virtue of its institutional position, to adhere to the President’s current policy stances in its public statements.

The Center that we recommend below is designed, in part, to fill the need for an independent institution that has expertise in high-quality impact evaluations and, where necessary, the ability to carry out such evaluations. It would thereby serve a unique function that the existing government analytical agencies such as GAO, CBO, and OMB do not now serve.

#### **IV. Barriers to government’s effective use of high-quality program evaluations.**

##### **A. Decision makers are often unable to sort out conflicting evaluation data, and conflicting interpretations of the data.**

Virtually every policy proposal is accompanied by the sponsor’s statement that its direction is the one indicated by current research and evaluation. Opponents often simultaneously offer alternative assessments of the data or cite studies supporting different views. Advocacy groups on all sides aggressively promote their particular view of the problem and how the data support their position. The data reviews of each participant in the debates are, of course, selective, intended to support the proposed position, but the fact they are present at all indicates the widespread recognition of the value of demonstrating a research and evaluation base for policy proposals. Often missing from the debate is a balanced appreciation of the whole body of applicable information from a reliable, disinterested source.

---

<sup>22</sup> John Heilemann, “Proof That Gingrich & Co. Don’t Know What They’re Doing,” *The Washington Monthly*, vol. 27, January 1995, p. 14.

For example, Success for All (SFA), a reading program developed by Robert Slavin at Johns Hopkins University,<sup>23</sup> has generated widely differing reports on its effectiveness. There is research that supports the effectiveness of SFA and research that challenges it. One writer argues that SFA is mostly based on research done by the developers of the program, not objective analysts.<sup>24</sup> On the SFA website, the program developers reject this contention.<sup>25</sup>

The Job Corps works with highly disadvantaged youth normally in a residential setting. From its beginnings in the 1960s it has been attacked for high costs, low overall benefits in relation to costs, and limited benefits to enrollees. It survived initially because of its status as the major program trying to deal with the most difficult youth, anecdotal success stories, and a strong external constituency. A major net impact evaluation using a comparison group methodology was launched in the 1970s, which by the early 1980s had produced data indicating significant positive net benefits.<sup>26</sup> The methodology was quite complex and like most such studies, open to question and differing interpretations. The Reagan administration did question the methodology and results as it attempted to cut program funding. The program survived, in part because of the positive evaluation findings.

A *Washington Post* article in May 2001, subtitled “Tax Cutters, Backers of Social Programs Debate Long-Term Economic Benefits,”<sup>27</sup> notes that “two programs under debate -- federal funding to reduce class size and Head Start, a preschool program for the poor -- demonstrate the difficulty of determining the impact of federal spending. In both cases, advocates and foes disagree sharply over whether the programs are effective, sometimes pointing to the same data to make their case.” The debate over Head Start, with conflicting studies, ambiguous findings, and starkly different interpretations of the same or similar data, has been ongoing since the 1960s.

Sorting out the meaning of a body of data on any really important issue is not simple. Few studies are definitive; all require interpretation. Decision makers must somehow assess the validity of the assertions and claims of all sides. Nothing can or should prevent decision makers from using data as they wish to support their positions, but something more is needed to give them and the public a more balanced view of the data.

The Center for Domestic Program Assessment we propose would offer such balanced, non-partisan assessments. For example, it could help Congress and the federal agencies interpret the meaning of the whole body of evaluation research that bears on a particular policy issue, by reviewing the methodological strengths of the various studies, and reconciling and synthesizing their findings in a fair and meaningful way.

---

<sup>23</sup> Success For All, <http://www.successforall.net/about/whoweare.htm>

<sup>24</sup> Debra Viadero. “Miami Study Critiques Success for All,” *Education Week*, January 27, 1999.

<sup>25</sup> Success For All, <http://www.successforall.net/resource/faqresearch.htm>

<sup>26</sup> Conducted by Mathematica Policy Research for the Labor Department

<sup>27</sup> *Op. cit.*, no. 11.

## **B. Often evaluation results are not available in time to influence decision making.**

Large scale, multi-year program demonstrations and evaluations – the types most likely to yield reliable data -- are often not completed until after the associated policy had been decided on, a new program implemented, or the original problem diminished in importance (in fact or as a political matter). Evaluators and data analysts do not necessarily time their reports to when decision makers need their output, such as upcoming reauthorization dates or funding cycles. Decision makers themselves do not often authorize and fund analyses early enough to give them what they will need several years in the future.

Richard Nathan notes, “The negative income tax demonstrations [of the 1960s and early 1970s] were moderately successful as research projects, although much less successful as an aid to policy-making. Their results came very late in the policy process and were at best ambiguous from the point of view of advocates of the negative income tax concept.”<sup>28</sup>

As noted earlier, the landmark 1996 welfare reform legislation that replaced Aid to Families with Dependent Children (AFDC) with Temporary Assistance for Needy Families (TANF) was preceded by years of small-scale demonstration programs testing different approaches to helping poor people get off and stay off welfare. Even the most dramatic provision of the 1996 law, a lifetime limit on receipt of benefits, was being tested under the AFDC waiver authority, but there were no findings ready in 1996. This dramatic policy shift was enacted without waiting for indications of its likely impact on families.

The reauthorization process for TANF, which receives \$16 billion in federal funding annually, is underway. It will be informed by better but still not sufficient data on the time limits and other substantial issues. While the law continued substantial funding for welfare evaluation and research, including continuing studies of on-going AFDC waiver experiments, HHS decided not to attempt an overall impact evaluation of TANF.

As discussed below, a critical function of the Center that we propose would be to develop a long-term analytical agenda that is timed to coincide with critical points in the decision-making process, so that decision makers will have timely analysis to inform their program decisions.

## **C. No mechanism currently exists to ensure that programs are modified over time in response to evaluation results.**

New programs are regularly launched in a spate of enthusiasm generated by public attention to a problem and the desire of decision makers to do something about it. After a time, tension between the initial intuition and data on actual program performance can increase, calling into question the program’s original rationale or, more often, its specific design. If they ignore or downplay such data in the name of consistency with policy, decision makers can miss critical opportunities to restructure the program to make it effective. And they can miss the right time to cut losses and end the program to avoid wasting taxpayer dollars. The lack of a mechanism requiring response to data by lawmakers sacrifices a great deal of potential accountability to the public.

---

<sup>28</sup> Op. cit., no. 14, p. 51.

An example of strong intuition confronting weak supporting data is the drug abuse prevention program called Drug Abuse Resistance Education (DARE), which sends police spokesmen into schools to teach children the bad effects of drug abuse.<sup>29</sup> DARE is extremely popular with police forces and parent groups.

A 1988 University of Illinois report on DARE evaluations<sup>30</sup> found little difference in the level of drug use based on whether students participated in DARE, and no positive outcomes that lasted into high school. A study by the University of Kentucky of more than a thousand students who participated in a sixth-grade DARE program found “no difference between those who received DARE and those who did not in their use of cigarettes, alcohol, marijuana or other drugs.” The only difference was that those who received the training had lower levels of self-esteem.”<sup>31</sup> The Justice Department itself funded a study by the Research Triangle Institute which found that DARE has a “limited to essentially nonexistent effect on drug use.”<sup>32</sup>

Despite an abundance of negative findings like these, even in studies by the funding agency, DARE continues to receive high funding and its popularity is growing. The program can be found in all 50 states and in more than half of U.S. school districts, and has been adopted in 44 foreign countries.

Institutional mechanisms are needed to ensure that programs such as DARE respond and adjust to evaluation results over time. The recommendations that we discuss below are designed to create such mechanisms.

**D. Neither Congressional nor executive branch processes for allocating resources are structured such that evaluation results are a meaningful input into resource allocation decisions.**

Virtually all appropriators and appropriations committee staffers would assert that they use the best performance data available in making their decisions. Committee reports occasionally bemoan the lack of good data from agencies. On examination, however, the majority of appropriations bills and reports focus on process or output factors, such as numbers of individuals served or grants made, not on evidence of program impact.

The President’s annual budget usually contains a chapter with selective data on performance that supports significant allocation decisions. The treatment, however, has been highly selective, as befits an essentially policy-driven document. Serious program management problems are discussed, but normally in a separate part of the budget books, without direct

---

<sup>29</sup> Drug Abuse Resistance Education Program, <http://www.drcnet.org/DARE/section1.html>.

<sup>30</sup> Dennis P. Rosenbaum and Gordon S. Hanson, “Assessing the Effects of School-Based Drug Education: A Six-Year Multi-level Analysis of Project D.A.R.E.,” *Journal of Research in Crime and Delinquency*, vol. 35, no. 4, November 1998, pp. 381-412. This report also found that “most of these studies are of limited scientific value because of their weak research designs, poor sampling and data collection procedures, inadequate measurement, and analysis problems.”

<sup>31</sup> Diane Barnes, “DARE Indoctrination Fails to Work and Ends Up Endangering Families,” *The Detroit News*, April 2, 2000.

<sup>32</sup> *Ibid.*



reference to the spending choices described elsewhere. (The Bush Administration has promised to improve this situation.)

Agency GPRA strategic and annual plans and performance reports are nearly always written without reference to resource needs. Arguably, this reflects the proviso that such documents not be used to appeal for greater or fewer resources than are proposed in the pending President's Budget. But without reference to resources, GPRA documents are not a serious reflection of the choices made in the agency to deliver performance. In many agencies, GPRA processes have been set apart from actual agency management processes and decision making.

As noted in the testimony referred to above from J. Christopher Mihm, GAO Director of Strategic Issues, there is a further layer of problem where more than one agency supports major programs aimed at the same domestic policy purpose, as happens frequently. It is extremely rare that multiple agencies dealing with the same issue collaborate on evaluation or on policy setting and virtually unheard of for them to work together to devise GPRA strategic goals or annual operating plans and success measures. Thus the disconnect between evaluation and resource allocation is further accentuated when multiple agencies and programs are addressing a common national need.

The recommendations that we outline in this paper are designed to foster the effective use of evaluation results in resource allocation decisions in all three of these areas – the Congressional appropriations process, the President's budget process, and the GPRA-related planning processes that support program management in the federal agencies.

**E. Policy decisions that ignore evaluation results are not held up to public scrutiny.**

Maurice McTigue of the Mercatus Center argues that “transparency” or disclosure are the missing ingredients, noting, “In my opinion, the limited use of performance information is due to the fact that there has been no consequence for either using or not using the information.”<sup>33</sup> His views carry weight because as a Member of Parliament and Cabinet Minister in New Zealand, he was deeply involved in that nation's movement toward performance-based decision making and has seen the positive influence results data can have on elected and appointed officials.

In sum, there is much rhetoric but there are few practical incentives in the current system to generate and use high-quality impact evaluation to inform resource allocation and program design decisions.

\* \* \*

**V. Recommendations**

In this section, we outline for debate characteristics of the institutional and procedural innovations that we believe would lead to a quantum improvement in the quality, quantity, timeliness, utility, and use of program evaluations to inform program design and management and resource allocation decisions.

---

<sup>33</sup> Op. cit., no. 5.

The principal recommendation is the establishment of a strong new capability to assess and report publicly on the results of programs and policies, that is linked to Congressional and executive branch decision making. This is followed by recommendations to change current institutions and procedures to make the best use of those assessments and to upgrade generally the incentives for linking resources to results.

### **Recommendation 1. The Center for Domestic Program Assessment**

*Overview.* We recommend the establishment of a Congressionally-chartered, nonpartisan Center for Domestic Program Assessment, to strengthen and help institutionalize the link between government resource allocation decisions and demonstrated program effectiveness. The Center would lead the movement to upgrade the visibility and use of high-quality program evaluation. It would fill the many gaps in evaluation coverage of major programs and polices by systematically encouraging the funding of evaluation, monitoring the quality of evaluations, conducting missing evaluations where necessary, reporting regularly and publicly to decision makers on what works and what does not, recommending program redesign and funding changes indicated by the data, and continuously monitoring performance.

The Center is by no means intended to replace existing evaluation and analysis capabilities. It would build on that work, help improve its design by its overt use of the products, and knit together others' findings on program results with its own. As noted in recommendation 2 below, we particularly support regularizing the funding of agency evaluations through statutory funding set-asides.

What follows is a preliminary outline of the proposed Center's functions, agenda, board of directors, and funding, which is intended to serve as a starting point for future discussion. We recognize that significant additional deliberation is needed in such areas as the Center's governance structure, reporting mechanisms, and internal processes (e.g., peer review), to ensure that the Center's activities strike the right balance of independence from government and partisan politics on one hand, and accountability to the needs of Congress, the executive branch, and the public on the other.

*Functions of the Center.* The Center would assess and report regularly and publicly on the status of progress in resolving the country's most significant domestic issues, as evidenced by evaluation of the results of programs and policies. Its assessments and reports would take a government-wide perspective on policy issues, which may at times diverge from the more narrow, jurisdiction-based focus of individual agencies and Congressional committees. Of necessity, its analyses would put it in the position of commenting on how legislative goals and agency goals (e.g., in GPRA plans) are expressed, to help drafters specify goals in measurable ways. In addition,

- Based on this analysis, it would recommend changes in program design and resource levels. The proposed changes may, in some cases, cut across the jurisdictions of individual Congressional committees and federal agencies, when the evidence supports such an approach.

- It would assess the quality of available information on programs and management performance, support funding of new evaluations and data collections, and generate new evaluations with its own funds where needed.
- The Center would be a resource to agency evaluators as they design evaluations, focusing on selection of the most appropriate methodologies, helping ensure that the design adequately addresses the factors likely to affect results, and relating evaluations of individual programs to evaluations of programs in other agencies that are intended to have an impact on significant aspects of the domestic problem at issue.
- The Center would solicit, interact with, and record in its reports public views on programs. It would provide opportunities through electronic communications, public hearings, and the like for public interaction with its analysts.
- The Center would testify before Congress and the Administration on its views.

***The Center's agenda and products.*** The Center would initially propose for public comment and input a long-term analytical agenda and schedule of issues to examine. Sequencing would tie in large measure to likely dates when government needs the best information before making its decisions (as in the examples below). "Out of cycle" analyses and agenda revisions would occur as indicated by the Center's on-going monitoring of data on domestic issue areas and the performance of programs. The agenda would be reviewed annually by the Center's Board of Directors (whose structure is discussed below), with input from decision makers and the public, to assure that its focus and timelines were still appropriate.

Examples of potential Center products:

- If the Center had existed in 1999, it would have produced a detailed assessment of the impact of elementary and secondary education programs on raising student achievement to inform debates beginning in that year (and still underway in 2001) on reauthorization of the Elementary and Secondary Education Act.
- Today, the Center might be producing similar assessments of the collective impact on early childhood education and childcare of the programs in Treasury, Education, HHS, and Agriculture to guide efforts to raise educational achievement through effective early intervention.
- If the Center had existed in the early 1990s, it would have informed Congressional deliberations on the 1994 Omnibus Crime Control and Safe Streets Act by providing high-quality assessments of existing community policing initiatives, school-based crime prevention strategies, initiatives to recruit and deploy additional police officers, and other crime prevention and control activities.

In each instance, the Center would look at state, local, non-profit and private sector efforts as well as federal government programs, so that it could present a full picture of the problem, efforts to address it, and new directions indicated, including but not limited to federal initiatives.

***The Center's charter and institutional position.*** Although one can conceive of locating the Center within either the executive or legislative branches, or as a hybrid government entity (e.g., a "government-owned-contractor-operated" facility, as exist for certain research functions),

we do not believe that a government entity could have the needed independence and do not believe that creating new government agencies is a politically viable option in any case. We therefore suggest that the optimum structure for this Center is as an independent, non-partisan, well-funded, non-government entity.<sup>34</sup>

Much like the National Academy of Sciences, the Center would be chartered by Congress in a statute approved by the President, so that its connection to the government decision making processes would be clear.<sup>35</sup> The charter would specify the non-partisan, independent nature of the Center, and would also specify the Center's main functions, discussed above. In addition, the charter would establish incorporators to select the non-government members of the Board of Directors. If desired, the incorporators could be the sponsors of this study, The Council for Excellence in Government and the Committee for Economic Development.

Charter legislation could originate in the Senate Committee on Governmental Affairs and the House Committee on Government Reform, the committees that originated GPRA and other important government management reforms of the 1990s. The Center would represent for these committees a natural next step in their efforts to raise the quality of government performance. The charter legislation would charge these committees and the executive branch to collaborate on the funding of an independent evaluation of the Center and report findings to both branches and to the public.

**Board of Directors.** The Board would be composed of two types of members. The first type – non-voting, federal government members -- would be selected by the House, Senate, and the President from among elected and appointed officials. These directors would participate in all deliberations on the Center's agenda, judge and opine on the quality of its work and its reports, and represent the views of the federal government. The government directors would not, however, vote on board decisions on agenda or analyses. In this way, the Board would be insulated from direct government pressures but still responsive to government needs and views.

The government directors could also help ensure the accountability of the Center by conveying to the Congressional committees that originated the charter, and others, their views on the Center's success in: living up to its charter, maintaining the necessary level of excellence in its work, maintaining nonpartisan objectivity, and investigating high-priority issue areas.

The second type of Board member -- the voting directors – would be selected by the incorporators from a broad cross section of evaluation experts, policy analysts, state and local government officials, interest groups, and lay people. They would serve staggered terms. The voting directors would appoint the Executive Director of the Center and establish that

---

<sup>34</sup> Should there be strong government interest in a Center with the other attributes we propose, but a political necessity to place its functions within government, GAO is the one current entity with the best capability of carrying out these functions. Unlike CRS and CBO, it routinely makes policy recommendations, and unlike OMB, it has relative freedom to present its analyses without needing to adhere to the President's (or the Congress's) current policy stances. We do not find this approach as likely to satisfy the necessity for independence as the Center proposal, but it could come as close to that as any inside government placement can.

<sup>35</sup> In fact, if the National Academy of Sciences were willing to serve as a base for the new Center and to adopt the operating approach outlined here, Congress may wish to consider locating the Center within the Academy.

individual's performance contract with them, on which compensation and tenure would depend. The voting directors would have to approve the Center's strategic (or business) plan and issue agenda, and review the Center's products for quality and adherence to the purposes of the charter.

***The Center's budget and Funding.*** We believe the government's initial investment in the Center needs to be quite substantial, both to indicate government's seriousness of purpose and to support development of the high-quality capability needed to carry out its functions across a reasonable number of policy areas. At the same time, the Center must not be beholden to government for annual funding or spending authority if it is to have the essential characteristic of independence.

To help size the Center, here are the annual budgets of entities that in some respects might be comparable as staffs of public policy analysts with a report generation capability:

- The former Office of Technology Assessment: approximately \$22 million
- The Brookings Institution: approximately \$28 million
- The Heritage Foundation: approximately \$28 million

We suggest that the Center receive a one-time appropriation of \$1 billion. A part of this sum, perhaps \$50 to \$100 million, would cover initial start-up and operating costs (operation of the board, hiring, space and related costs, design conferences, the process of gathering public input on the agenda, development of an interactive website).

The remainder would be invested as an endowment whose income would supply operating funds. The Center would also be able to receive contributions from foundations and individuals, and grants and contracts, to support its work.

***The Center's Executive Director.*** The Executive Director chosen by the Board would be a widely respected individual with these characteristics:

- A thorough appreciation of the need and uses of evaluation data by government and the public.
- Intimate knowledge of the budget and policy development processes and the opportunities for enhanced use of impact data.
- The ability to oversee the assessment (and where necessary, the design and execution) of high-quality evaluation.
- The ability to recruit and manage a staff of domestic policy and evaluation experts.
- The ability to deal effectively and on a non-partisan basis with officials of government at all levels, and with interest groups across the political spectrum.
- Knowledge of creative uses of technology to enhance interaction with the public and to publicize and disseminate the products of the Center.

As noted, the Executive Director would serve under a negotiated performance agreement with objective, measurable goals.

\* \* \*

We wish that demand for the products of the Center could be assumed to be self-generating, or at minimum had already been made manifest by Congressional and executive branch use of GPRA authorities. We have shown that this is not the case. We believe that maximum benefit would be obtained by the acceptance of the recommendations in the next section, which speak to generating the incentives and demand for such analyses.

**Recommendation 2. Related improvements in Congressional and executive branch processes.**

*Evaluation funding set-asides.* The design and funding of high-quality evaluations of net program effects should be made routine. While subject to the normal Congressional and executive branch review and final decisions through appropriation acts, OMB apportionments, and other devices, agencies should know that every major program will be evaluated.

1. House and Senate rules should require that any domestic program authorization, including mandatory spending and significant tax expenditures, can only be considered on the floor of either body if it contains a requirement for an impact evaluation and a funding set-aside to finance it.
2. OMB should promulgate guidance to agencies (e.g., as part of the Circular A-19 legislative program procedures) that requires any bill drafted in the executive branch to have an evaluation provision with a funding set-aside before the Administration will transmit it to Congress.

*Congressional oversight.* All committees of Congress carry out oversight of programs and policies under their purview. Some of this oversight is carefully structured to illuminate the issues, features a range of expert opinion, and is focused on surfacing findings and recommendations for program improvement. Much of it, however, is also (often primarily) political theater, necessary in our system but not likely to advance understanding of issues.

Periodically, Congressional leaders attempt to restructure the oversight process to direct more attention to substantive analyses. Senator Peter Domenici, former Chairman of the Senate Budget Committee, is a champion of bi-annual budgeting, because, as he notes, “The authorizing committees in my humble opinion have atrophied in the United States Congress, at least in the United States Senate, because it's very hard for them to find two weeks to debate a very important bill on the floor, when we're using about 75 percent of the time for budget resolutions and for appropriation bills and supplemental appropriation bills.”<sup>36</sup>

1. Congressional rules should require oversight hearings timed to inform upcoming debates and, upon creation of the Center, to require presentation of the Center's views (among others) of the issues at such hearings.
2. The rules should further provide for joint hearings of the major committees engaged in an issue area, so that all the major efforts to address it can be considered in the same context.

---

<sup>36</sup> Senator Pete Domenici, speech before the Council for Excellence in Government, April 26, 2000.

(Recall the early education and childcare example above, which would call for four committees to deliberate together to make the best informed choices). Ideally, such hearings would bring the cognizant House and Senate committees together as well.

3. Except in emergency situations, bills not preceded by such hearings and not informed by available Center analyses should not be allowed floor consideration.

***Sunset and performance goal provisions.*** No government program should escape periodic assessment of the how well it is working and what changes might be needed.

1. House and Senate rules (and parallel executive branch procedures) should require each major domestic program authorization, mandatory spending program, and tax expenditure to have a sunset provision of no more than five years in duration, that requires renewed Congressional approval of its structure and funding authorization, based in significant part on the results of the impact evaluations called for above.
2. Rules should require that program authorizations be enacted prior to passage of discretionary appropriations (such current rules are often routinely waived), so that disagreements over program structure are argued out in the authorizing committee framework.
3. All such bills should have to include measurable goals and performance standards on which evaluations and agency GPRA plans would, in part, be based. Progress toward these goals and standards would be reported annually in the GPRA performance reports. By requiring goals and performance measures in statute, the Administration and Congress will have to negotiate these seriously, a practice not generally realized by GPRA's requirement for "consultation" on goals and measures.

***Enforcement in Congress.*** Some rules similar to those above have been tried before; some are still in the rule books. None are routinely enforced or effective. To emphasize Congress's commitment to raising the quality of government performance, an unusual but not unheard of standard should be set for a waiver of any of the above rules:

Waiver of the above rules should require a 60-vote majority in the Senate and a similar high hurdle in the House (perhaps in the Rules Committee procedures). Such requirements to supercede "majority rule" are quite properly reserved for extraordinary situations, such as waiving budget enforcement rules. We believe this approach is warranted by the importance of heightening the performance focus of government.

***Strengthening the Government Performance and Results Act.*** GPRA is the first major statutory effort to attempt to link resources to results since the Budget and Accounting Act of 1921. Progress has been made in many agencies in establishing measurable performance goals, focusing on management improvements needed, and putting in place the infrastructure to support regular performance reporting. More needs to be done, especially on the dimension addressed by this paper -- obtaining and using more high-quality evaluation data on results to inform decision making.

The following changes would build on this record and would enhance the current Administration's proposal to increase reliance on performance data in its management and budgeting.

1. Amend GPRA to require reporting of current evaluation findings under each performance objective, including specifying how recently evaluations have been conducted, the status of on-going studies, and plans for new studies. Some agencies do this now (e.g., the Department of Labor); others relegate the discussion of evaluations to an appendix and make no clear connection between evaluation findings and resource allocation or management's strategy to improve performance.
2. Require a summary of funding for evaluation and research on results, indicating trends and the pending budget request and justification for such funds. This summary is not an opportunity for agencies to appeal for additional funds, but rather the way the agency demonstrates its commitment to learning the results of its effort and an explanation of what that focus costs.
3. If an agency does not integrate its annual budget with its annual GPRA plan and performance report, require that the budget justifications use the same language on performance as is found in the GPRA documents.

OMB rules do not now require such integration, but it seems only logical that the best way to link resource allocation to program results is to integrate performance reporting with budget justifications. Because of the inherent complexities in doing this well, we do not now recommend it for all agencies, but we do believe OMB should work with each agency to encourage it and in time, to require it.

4. Require joint GPRA plans and reports across agencies for issue areas in which more than one agency has a major role, so that Congress and the public can see clearly how programs and program management address common problems.

\* \* \*

***Note on the role of technology.*** The power of electronic communications technology to gather information and to obtain and respond to public comment efficiently has been amply demonstrated in recent years, if not yet fully realized. The Council for Excellence in Government has proposed a comprehensive "blueprint" for how government can more rapidly expand and structure the use of electronic communications.<sup>37</sup> The Bush Administration has indicated its support for greater use of new technology, adopting in its budget an e-government fund to finance development of cross-agency initiatives and avoid inefficient duplication of effort. Senators Thompson and Lieberman have established a website to gather views on how to make Congress' participation in the development of web technology more effective.

---

<sup>37</sup> E-Government: The Next American Revolution, Council for Excellence in Government, February 21, 2001.



We believe that this technology can lower the cost and speed the completion of impact evaluations, for example by helping automate the random assignment of participants in an experimental evaluation, by facilitating the creation of statistically-equivalent treatment and comparison groups in a quasiexperimental evaluation, by speeding the collection and analysis of data on relevant outcomes, and by facilitating the rapid dissemination of evaluation results. We would charge the Center with making a special effort to work with the evaluation community to develop and implement these techniques.

We urge the Congress and the Administration, as well as the Center, to use this technology to solicit input to plans and programs during the development stages, and to respond to public input on a timely basis. This will add considerably to the interested public's confidence in the governmental process.

## **VI. CONCLUSION**

These recommended changes form a necessary complement to the growing momentum in Washington, in state and local government, and indeed internationally,<sup>38</sup> for defining program goals more practically and for taking into greater account the performance of current programs when designing new activities and allocating resources. They also strongly respond to the public's demand for high-quality information on society's problems that could go a long way toward restoring public confidence in government. As tracked by Hart-Teeter polls conducted for the Council for Excellence in Government,<sup>39</sup> the current level of public confidence has been alarmingly low for a long time (though the public's response to government's handling of the September 11 terrorist attacks has for the moment raised public confidence to an unprecedented height).

In 1921, OMB Director H.M. Lord said on the occasion of implementation of the Budget and Accounting Act, "Habits, customs, regulations, laws that the passage of more than a hundred years has built into the very machinery of government cannot be eradicated over night... it must be a continuing process that will require years of patient, persistent and courageous endeavor."

The same patience and persistence is called for here, if these recommendations are to have maximum effect, and this time, with the added benefit of the intervening 80 years of experience, we may finally see a firm link established between resources and results.

In the end, no planning or reporting requirement, no injunction to manage well, and no external criticism regardless of source, can alone lead to higher quality performance if decision makers and the public that influences them do not demand it. Nevertheless it is now clear that the focus on results is at a new high, which should make possible an informed debate on these recommendations.

---

<sup>38</sup> See, for example, Government of the Future, the report of the OECD symposium on government management reform held in Paris in September, 1999.

<sup>39</sup> See <http://www.excelgov.org/publication/excel/default.htm>