Presidential Address: "*Complexifying*" Performance Oversight in America's Governments

Richard P. Nathan

In keeping with my focus as President of APPAM on the "M" in our name, this talk deals with the performance management movement in American government.

Senator Daniel Patrick Moynihan, in a moment of frustration with me at a 1989 Senate hearing on welfare reform, said I am a "complexifier." In this spirit, the way I see it is that the performance management movement in American government is on the right track, but that it *oversimplifies*. I want to serve today in the role of a "complexifier" for the performance management movement. My aim is to be constructive—to suggest ways in which efforts to improve government performance can be reconciled with the pluralistic setting of U.S. public management.

Leaders in the federal government over the past 40 years have oversold simplistic systems for fulfilling public policy goals as expressed in the alphabet soup of systems like PPBS, MBO, ZBB, NPR, and GPRA.

The Alphabet Soup

PPBS stands for Lyndon Johnson's Planning-Programming-Budgeting System, adopted with much fanfare and based on private industry and Defense Department systems to assess and compare public spending options. MBO was Nixon's successor approach for Management by Objectives. ZBB was Carter's more radical initiative for zero-based budgeting to rank all spending options from the ground up in allocating government funds. NPR was Clinton's National Performance Review, which sought to focus government management and budgeting on achieving results. In 1993, a system for strategic results-based management and budgeting was enacted into law under the Government Performance and Results Act (GPRA).

For shock value, at the end of October of this presidential election year, I want to say something good about the newest alphabetically named performance management system. No matter what happens in the presidential election next week, I suggest this slogan, "*Let's not part with PART*," referring to the George W. Bush Administration's "Program Assessment Rating Tool" to compile effectiveness ratings for all federal programs.

THE PART SYSTEM

According to the FY 2005 budget, the Administration has "PART*ed*" (this verb form is used frequently in the budget) 400 programs representing 40 percent of the

budget. Two hundred programs will be added in FY 2006. These evaluations are based on 30 questions to rate each program that is assessed in four areas: 20 percent for its purposes and design being clear, 10 percent for strategic planning, 20 percent for program management, and 50 percent for results. The ultimate goal is to evaluate the performance of all federal programs (over 1,000) in this way.

PART assessments draw one of five conclusions: effective, moderately effective, adequate, ineffective, or results not demonstrated. They are available online.[1] Illustrative ratings are for Head Start "results not demonstrated," for Medicare "moderately effective," and for the Community Development Block Grant (CDBG) "ineffective." Reading PART reports (most of them are about eight to ten pages long), one is struck both by their brevity and their variation. Some, for example on food stamps, are evenhanded and draw on a range of sources from within government and outside. Others are analytically weak.

DISCUSSION

The critical challenge for the PART system and efforts like it is setting performance goals. Where do goals come from? The answers vary. In some cases, it is wishful thinking in the form of political over-promising. In other cases, and these are the ones I care about, performance goals are based on research, policy analysis, and expertise. For this big category, I believe there should be broader consultation and collaboration involving policy officials, program managers, policy analysts, academic policy researchers, and public management experts.

In situations in which the results of definitive social science experiments based on random-assignment studies can be drawn upon, they materially aid policymakers by giving them a high level of confidence about impacts. *Such studies, however, are not available, and indeed could not be conducted, for all types of public programs and for all types of affected groups and policy conditions and needs.*

Terminology is important here. An *impact* means we can show that a particular public program caused something to happen that would not have happened otherwise. An *outcome* is the word customarily used to express a program's performance goals, regardless of whether we know if its activities are additive. Frequently, two types of people are involved: policy researchers, who generally favor experimental studies of program impacts, and management experts, who focus on outcomes. The two groups often have different mindsets and skill sets. As I see it, public policy and public management researchers (especially active members of APPAM among them) should contribute to knowledge about *both* impacts and outcomes.

Leaders in the public service are called upon in many (indeed, most) situations to set and periodically adjust outcome goals for performance management based on the most pertinent public policy knowledge available, and also drawing on expertise, experience, and observation. And yet even with the best of such efforts, governments often do not deal wisely with another critical challenge discussed in this paper, the need to devise politically acceptable and workable performance goals that can't be gamed for undesirable purposes.

While I think the PART system is on the right track in focusing on individual programs as the basic building blocks for assessing managerial performance, major problems are that it is too centralized, too insular, and not sufficiently discriminating. It does not adequately take into account the great differences that exist in the

---

[1] See www.whitehouse.gov/omb. Accessed September 7, 2004. See also Rodriguez, 2004, pp. 56–61.

size, importance, operational character, and settings of different public programs. I have a particular *federalism* problem in this context.

THE FEDERALISM CHALLENGE OF PERFORMANCE MANAGEMENT

Most domestic programs of the federal government operate by indirection through state and local governments that in many cases contract with nonprofit and private corporations to provide public services. It is beyond the scope of this talk to survey the backward and forward bounces of decentralization in American intergovernmental relations. Suffice to say, as Martha Derthick emphasizes, there are many ways in which members of the Congress and federal Executive Branch officials attempt to influence domestic affairs by adopting narrowly focused grant-in-aid programs and imposing and enforcing conditions, regulations, and guidelines on their operation. On the other hand, there is also a long history of devolutionary efforts to broaden grants (for example, by creating block grants), loosening or not enforcing regulations, granting waivers of federal requirements that are sought by state and local governments and interest groups, and generally by virtue of the fact that policymakers often disagree on public purposes and as a result adopt vague or even contradictory goals for domestic public policies.[2]

Wherever the devolution ball bounces, the point that stands out for me is that state and local governments have to be involved in assessing and improving program performance, especially under federal grant-in-aid programs that are broad gauged and have multiple purposes, as is so often the case. Federal agency officials should work with state officials in ways that are not heavy handed. They should adopt approaches that are continuous, user-friendly, candid, and appropriately intergovernmental.

This recommendation gets me into the consideration of differences between inputs, outputs, and intermediate and end outcomes as the goals of performance management.[3] It is arguable as a general principle that under broad-gauged federal grants-in-aid, the federal government should care most about *organizational outputs,* and in turn work with states to stimulate them to assess and help them in the best ways they can to define and measure the *outcomes for individual participants* of these federally aided domestic programs in the different environments in which they operate. This is what we actually do in more cases than is acknowledged.

Governmental programs have literally thousands of iterations. Reconciling two values, the flexibility (which is inherent to the diverse governmental environment of American federalism) with accountability (which policymakers and administrators should care about and achieve), requires that the PART system be *intergovernmentally sophisticated*. But this is not the only way the performance management movement needs to be complexified.

---

[2] For useful essays on this subject, see Derthick (2001) and also her other writings.
[3] See the attached definitions by Hatry, 2001, p. 19. This part of the discussion concerns "units of analysis," as well as measured goals. For some government programs, performance management goals involve *individuals* as the units of analysis (students, job seekers, sick people). For others, and for reasons that often reflect inherent limitations of measurement, we settle for (or may even decide we prefer) *organizational* goals as the units of analysis: Did service providers do what they were supposed to (for example, social security offices, state food stamp programs)? The subject of organizational performance is treated in Nathan (2000), chapters 6, 8–10, which describe institutional evaluations of the implementation of national welfare reform policies.

THE GAMING CHALLENGE OF PERFORMANCE MANAGEMENT

The Office of Management and Budget in its on-line documentation on the PART system says it is not primarily a budget tool. I agree with this idea. The PART system should not be sold solely as a method for deciding that such and such a program doesn't work so we should cut it, or that it does work and we should add resources. That is not to say that weak programs should be retained or that they should receive more resources. There are situations in which performance findings should influence budgeting. But if PART is used mainly for this purpose, the likelihood is that it will lose its managerial efficacy.

This possibility of gaming raises difficult questions for performance management involving how to set performance goals so that they influence agency behavior in the desired ways. Doing this requires both hard and soft accountability measures. In private industry, soft decision factors are often taken into account in rewarding managers. Economists have affirmed the wisdom of blending objective and subjective performance measures in this way (Baker, Gibbons, and Murphy, 1994). Unfortunately, however, the preponderance of attention and literature on managerial oversight in government has focused on rigid numeric goal setting.

In this intense 2004 pre-election period, it is not easy to envision the kind of consultative and interactive goal setting that could wisely and in evenhanded ways bring qualitative and policy related decision factors into play in performance oversight. This is made harder right now by cuts that have taken place in the managerial staffing of domestic programs, because overseeing such goals is necessarily labor intensive. Moreover, countering the rise in rating-mania by bringing qualitative and political variables into play in performance oversight is hard to prescribe. Yet, the truth is that, like speaking prose, we do it all the time, as suggested by some of the examples discussed next.

SOME EXAMPLES

In the field of employment and training, random-assignment demonstrations have shown that focusing on jobs is effective in aiding low-income, low-skilled people, especially women. However, performance-management goals focused on these dependent variables involving, for example, job placement and tenure can have problems. They can produce cream skimming, that is, selecting the most job-ready people to be aided, people who would have found jobs anyway. Corrections for this problem have been attempted, but can backfire. In one case I know of, the result was to undermine the public employment programs of the 1970s, leading to their eventual demise (Cook et al., 1985).

In the field of K–12 education, the emphasis on numeric goal achievement has caused unintended consequences that have required political jockeying and managerial recalibration. To paraphrase Tip O'Neil, all education is local. No matter how elegantly formed, the nature of the educational process requires that national oversight goals for local schools be focused on simple and general indicators, such that they tell only part of the story about school performance. For this reason it is important that performance goals for local schools be viewed as motivational every bit as much as being viewed as tools for the close and fulsome calibration of institutional effectiveness.

In the field of health policy, experts have urged the adoption of financial incentives for effective practice and quality enhancement, but at the same time acknowledge that the processes for providing and overseeing such standards face impedi-

ments (Epstein et al., 2004). Furthermore, we are told that such goals are unlikely to achieve their intended results unless substantial amounts of new money are provided (Epstein et al., 2004). Where is the money going to come from?

Citing these examples is not meant to suggest that setting performance goals and assessing how well they are achieved is unwise. The point is that performance management is difficult, can be expensive, and worst of all can backfire. It has to be smart and it has to be flexible, adaptive, and subtle. Stimulating efforts to ratchet up program performance on the part of the armies of governmental and quasi-governmental workers at every level of government who implement public policies requires setting and treating performance goals so that they serve *both* as targets for managers and as symbols for the public, in the latter sense, symbols that are well and widely understood and accepted. Raising student performance by using performance goals to focus attention on shared values about what our schools should teach has a good effect, and yet in the process of doing this we have learned about pitfalls of over-specification and have had to make adjustments. This is as it should be; undue rigidity in setting performance goals can undermine their effectiveness.

A good illustration of the latter point is shown by the recent success of welfare reform in reducing dependency and facilitating employment. The 1996 national welfare reform law dramatically changed bureaucratic behavior. Seemingly, this was rooted in strict and specific goals about jobs, hours of work, etc. But in reality, there was a big loophole, the "caseload reduction credit." This provision of the law, plus others, enabled states to advance their work-first purposes in a manner that reflected varied state and local values and conditions. The process was incremental and typically American. The success achieved was not the result of crafty planning by calculating policymakers. It came about serendipitously in ways that were surprising to us in our implementation studies of the 1996 national welfare reform law (Nathan and Gais, 1999; Gais et al., 2001, pp. 35–69).

MODIFYING THE PART SYSTEM

To reiterate, what is appealing to me about the PART system is that it focuses on individual programs as the basic building blocks, more so than on strategic and often overly elaborate purposes as advocated under previous federal government management reforms (Radin, 2000, pp. 111–135). What is needed, however, is a more candid and flexible treatment of goals in ways that involve a range of scholarly and expert perspectives.

Ann Blalock and Burt Barnow have called for a *partnership* for connecting demonstration and evaluation studies conducted by academic experts with performance management systems:

> Our recommendation is that competent evaluation research, or applied social science research, must be coordinated with or integrated within performance management systems if precise, valid, reliable information about social programs is to be made available to decisionmakers. (Blalock & Barnow, 2001, pp. 487–519)

Blalock and Barnow point out that what they call the "evaluation movement" was developed in the crucible of academia, while the "performance management movement" has its roots in public administration and in administrative bureaucracies. They believe, and I agree, that collaboration between these two movements would yield important benefits.

> We recommend that the major direction for the future is to coordinate evaluation research with performance management systems more fully, moving toward full integration of evaluations within performance management. Such integration will require that performance management systems treat evaluators not as aliens from outer space, who land only periodically to study and give advice, but as part of an interdisciplinary team. It will require that evaluators become more sensitized to managers' needs, to have ongoing information for tracking outcomes, and to express the benefits of their professional roots with greater humility. (Blalock & Barnow, 2001, pp. 487–519)

The two "movements" as described by Blalock and Barnow have different disciplinary bases. The impact/experimental movement is essentially Weberian in its assumptions about bureaucracy and implementation: Programs should be precisely controlled and replicated. By contrast, performance management seeks to establish goals as reference points for managers who are encouraged to innovate and change what they are doing in response to continuous feedback.

Just as there are many players in the bargaining processes for policymaking in American political pluralism, multiple players should be involved in deciding and adjusting the outcome goals for performance management.

Who should be the multiple players in outcome goal setting? Program managers at the appropriate levels of government (federal, state, and local) bring a needed perspective to bear about the likely effects of different performance goals in different settings. Researchers and policy analysts can also play an important role drawing on research that shows what programs are likely to work best and what their effects are when they do. But, it isn't enough to know if a program works. The responsible goal-setting officials also need to work their way through the hard questions raised earlier about how to express such findings in politically acceptable ways in performance goals that can't be gamed for undesirable purposes. Another important group of players is elected and appointed political officials; they clearly have a role to play in setting performance goals that reflect executive branch and legislative policies.

There is still another overarching aspect of this question about who should participate in performance oversight involving *auspices*. The GPRA law seeks the joint role of Congress and the Executive, whereas the Bush Administration's PART system gives the strong lead role to the Office of Management and Budget. While someone (that is, some agency) has to be in the lead, the challenge, as I see it, is not so much a challenge involving agency roles as one involving *transparency*. There needs to be a high level of transparency in sharing information about performance oversight. This is needed in the case of the PART system as it applies to other control agencies besides OMB, so that they can participate in performance oversight. The Congressional Budget Office, the General Accounting Office, and the Congressional Research Service, along with state and local budget and management offices and outside evaluators, need to have access to the underlying data that are used and understand how they are used in PART performance evaluations.

In line with the "complexifying" theme of my talk, this challenge of sharing data is made more difficult as more data are brought to bear, and more data *should be* brought to bear in performance management. An obvious and important opportunity is the availability of administrative data. We are virtually drowning in administrative data. Almost all domestic programs have reporting requirements that are extensive and detailed. Yet, despite this fact, and despite the fact that the quality of administrative data varies, not enough effort has been made to clean up these data, to scrutinize them and compile the best sources. Here, the underlying federalism

condition that states are different—that is, laboratories not just of democracy, but also of data availability—creates an environment in which multi-state studies can take advantage of opportunities to use administrative data wisely and more extensively.[4] Making fuller use of the best existing administrative data sets can have the important additional advantage of enabling evaluators to break out data for different groups of program participants in different kinds of program settings.

Another step that I believe would advance the art and practice of performance management would be to *relax* a little, loosen up the rating games. Instead of widely advertised *U.S. News and World Report*–type scorecards and report cards, people inside America's governments would do well to present performance results in more detailed and nuanced ways, explaining why and how they operate under different programs and conditions.

As a political scientist, I feel obliged at this juncture to go more deeply into my reason for putting so much emphasis in this talk on the dynamism, diversity, and fragmentation of the political milieu in which public services are provided. The reason is that so often in government *administration is policy*. Policymaking and administration are intrinsically linked. The day-to-day promulgation of rules and guidelines; the issuance of policy interpretations; the approval of grants; appointments and staffing decisions, all involve the transmission of values in ways that are more than routinely ministerial (Nathan, 1983).[5] The public administration literature is deficient in recognizing that policies and programs are changing all the time. Performance management must take this dynamism into account. It has to be seen as a *continuous* process. It has to be carried out by trial and error. It cannot be accomplished by fixed "one-size-fits-all" managerial formulas.

This is not to say that we should stop working hard on performance rating, only that these ratings should be presented in thorough and nuanced ways, and that they should be explained in statements that describe the strengths and limitations of the technology and data on which they are based.

It was not my intention in this talk to make specific recommendations for modifying the PART system. Moreover, my sense is that this should not be done in a rigid and formal way. Discussions about who is consulted and what kinds of inputs are used in setting performance are necessarily *situational*.[6]

Opening up the PART system (and state and local PART-like performance management systems) to more players and greater variations can also achieve something that is very practical and low tech. It can help deal with the information explosion in public affairs. We are virtually flooded with reports, e-mail messages, books, articles, conferences, and harangues. Performance management systems like PART, which pull together what is known about the effects of public programs, can inform and clarify debates about what America's governments do and the immense, diverse, and complex ways their actions affect social and economic conditions.

I realize that this plea for nuanced discourse, greater selectivity, flexibility, and transparency is wishful. "Complexifying" performance management in these ways

---

[4] Some large cities, notably New York City, have innovative performance management systems, which can provide valuable lessons and insights.

[5] On management tracking, I suggest Heinrich, 2002, pp. 712–725.

[6] The best book that shows how public management is *situational* is *Bureaucracy* by James Q. Wilson (2000 edition). Wilson describes organizations as needing "to acquire sufficient freedom of action." Based on his teaching and research, Wilson said he discovered that what was most needed was to permit an organization "to define its tasks as it saw best and to infuse the definition with a sense of mission." See chapter 2, "Organization Matters," pp. 14–28.

requires the application of judgment, and this invariably opens opportunities for political maneuver. Still, the fact of the matter is that we can never depoliticize performance management, nor should we even think of trying to do so. In the final analysis, what is needed is to bring more players into the process in ways that create a wise balance of expertise and politics. The long-term result would be to create a setting in which there is a more realistic focus on program results that gives citizens an honest, believable, and realistic portrayal of what their governments do and how well they do it.

RICHARD P. NATHAN, director, Nelson A. Rockefeller Institute of Government, State University of New York.

ATTACHMENT

Performance Management Definitions

* Inputs: Resources (that is, expenditures or employee time) used to produce outputs or outcomes.
* Outputs: Products and services delivered. Output refers to the completed products of internal activity: the amount of work done within the organization or by its contractors (such as a number of miles of road repaired or number of calls answered).
* Intermediate Outcomes: An outcome that is expected to lead to a desired end but is not an end in itself (such as service response time, which is of concern to the customer making a call but does not tell anything directly about the success of the call). A service may have multiple intermediate outcomes.
* End Outcomes: The end result that is sought (such as the community having clean streets or reduced incidence of crimes or fires). A service may have more than one end outcome.
* Efficiency, or Unit-Cost Ratio: The relationship between the amount of input (usually dollars or employee-years) and the amount of output or outcome of an activity or program. If the indicator uses outputs and not outcomes, a jurisdiction that lowers unit cost may achieve a measured increase in efficiency at the expense of the outcome of the service.
* Performance Indicator: A specific numerical measurement for each aspect of performance (for example, output or outcome) under consideration.

Source: Hatry, 1997, pp. 1–4.

REFERENCES

Baker, G., Gibbons, R., & Murphy, K.J. (1994). Subjective performance measures in optimal incentive contracts. The Quarterly Journal of Economic, 109(4) November, 1125–1156.

Blalock, A.B., & Barnow, B.S. (2001). Is the new obsession with performance management masking the truth about social programs? In D.W. Forsythe (Ed.), Quicker better cheaper: Managing performance in American government. Albany, NY: The Rockefeller Institute Press, pp. 487–519.

Cook, R.F., Adams, C.R., & Rawlins, V.L. (1985). Public service employment: The experience of a decade. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Derthick, M. (2001). Keeping the compound republic: Essays on American federalism. Washington, DC: The Brookings Institution Press.

Epstein, A.M., Lee, T.H., & Hamel, M.B. (2004). Paying physicians for high-quality care. The New England Journal of Medicine, 350(4), 406–410.

Gais, T.L., Nathan, Richard P., Lurie, Irene, & Kaplan, T. (2001). Implementation of the Personal Responsibility Act of 1996. In The World of Welfare, R., Blank and R. Haskins (Eds.), Washington, DC: Brookings Institution Press.

Hatry, H.P. (1997). Comparative performance measurement: FY 1996 data report (adapted). Washington, D.C.: International City/County Management Association.

Hatry, H.P. (2001). What types of performance information should be tracked. In D.W. Forsythe (Ed.), Quicker better cheaper: Managing performance in American government, Albany, NY: The Rockefeller Institute Press.

Heinrich, C.J. (2002). Outcome-based performance management in the public sector: implications for government accountability and effectiveness. Public Administration Review, 62(6), 712–725.

Nathan, R.P. (1983). The administrative presidency. New York, NY: Macmillan.

Nathan, R.P. (2000). Social science in government: The role of policy research. Albany, NY: The Rockefeller Institute Press (reprinted).

Nathan, R.P., & Gais, T.L. (1999). Implementing the Personal Responsibility Act of 1996: A first look. Albany, NY: The Nelson A. Rockefeller Institute of Government.

Radin, B.A., (2000). The Government Performance and Results Act and the tradition of federal management reform: Square pegs in round holes. Journal of Public Administration, Research and Theory, 10(1), 111–135.

Rodriguez, J.F. (2004). The arrival of performance budgeting. The business of government, pp. 56–61 (summer).

Wilson, J.Q. (2000). Bureaucracy. New York, NY: Basic Books.