

Performance Agreements in the U.S. Government: Lessons for Developing Countries

Dr. Prajapati Trivedi

The nonperformance of performance agreements in the U.S. government is among the better known “black holes” in the ever-expanding universe of Reinventing Government initiatives worldwide. This paper goes beyond the explanations based on institutional analysis to identify conceptual flaws in the design of performance agreements and derive the key lessons for developing nations attempting similar reforms.

The article was written from the perspective of belief in the value of performance evaluation in the public sector in general and the concept of performance agreements within government in particular. Thus its objective is not to question the rationale for the U.S. government’s performance agreements policy. Rather, it is intended to improve the quality of implementation efforts and make a good policy even better.

1. Background

The concept of a performance agreement is straightforward: It is intended to be a written agreement between the president of the United States and a cabinet department secretary that describes the mutual responsibilities of the two parties to the agreement. The following paragraph from the introduction to various performance agreements¹ signed by the president with a select group of departmental secretaries succinctly describes the concept.

“The American people deserve a government that works better and costs less. The departments and agencies of the federal government hold vital keys to improving performance and to restoring the faith of the American people in their government. Many changes will need to take place for this broad goal to be realized. The purpose of performance agreements with senior officials is to establish clarity and consensus about the priorities for departmental management. They are intended to improve the management of the Executive Branch and are not intended to create any legally

¹ Appendix I gives the 1994 Performance Agreement between President Bill Clinton and the Secretary of the Interior, Mr. Bruce Babbitt.

enforceable rights. From these agreements should flow the program management priorities of the departments. These agreements represent a beginning, a basis of continuous improvement as we reinvent our government to meet the needs and expectations of the American people.”

The origins of this policy lie in the pioneering Government Performance and Results Act of 1993 (GPRA).² This act requires that at least ten federal agencies launch three-year pilot projects, beginning in fiscal 1994, to develop measures of progress. Each pilot project was to develop annual performance plans that specified measurable goals. The agencies were to produce annual reports showing how they were doing on those measures. At least five pilots would also test so called managerial flexibility waivers—which exempt them from some administrative regulations—to help them perform even better. In exchange for greater flexibility, these agencies were to set higher performance targets. This is exactly the process of measured deregulation (we agree to deregulate you if you agree to be held accountable) that must be the basis of an empowered and accountable government.

GPRA requires that in 1998, after learning from the pilot programs, all federal agencies must develop five-year strategic plans linked to measurable outcomes. By 1999, every agency is expected to craft detailed annual performance plans—that is, plans that describe what they intend to *achieve*, not plans that detail how many pencils they will buy or how many people they will hire. The agencies will have to report their successes and failures in meeting those goals; however, GPRA allows the Office of Management and Budget (OMB) to exempt very small agencies from these requirements.

2. Importance of Performance Agreement Approach

Performance agreements represent the culmination of a long tradition of the central role played by evaluation in public policy.³ Whether from theory or from practice, the message is the same: What gets measured gets done. In large institutions, public and private, things are counted, and whatever is counted, counts. The simple act of defining performance measures is extremely enlightening to many organizations.

Typically, public agencies are either not clear about their goals or are aiming at the wrong goals. The lack of clarity of goals can be attributed to the fact that most public agencies have to deal with multiple principals who have multiple (and often conflicting) objectives. This leads to a fuzziness in the agencies’ perception of what is expected from them. The simple act of defining

² Appendix II gives a summary of the Government Performance and Results Act of 1993.

³ For example, see: R.K. Carter, **The Accountable Agency** (Beverly Hills, Calif: Sage, 1983); J.S. Wholey, **Evaluation and Effective Public Management** (Boston, Little Brown, 1983).

performance measures is enlightening for many public organizations. When they have to define the outcomes they want and appropriate benchmarks to measure those outcomes, this confusion is brought into the open. People begin to ask the right questions, to redefine the problem they are trying to solve, and to diagnose that problem anew.

The theoretical case for performance evaluation in government agencies is supported by the positive results that have accrued whenever an evaluation system has been implemented in the government. As the following example shows, this is true regardless of the level at which it was implemented.

During New York City's fiscal crisis in the 1970s, an independent foundation developed a method, called *Scorecard*, to measure the cleanliness of streets. It then sent out volunteers every month to rate each of 6,000 streets. The sanitation department had always focused on inputs: How many trucks were assigned to each district? How many men were needed on each truck? Now it began to look at the Scorecard information, which rated outcomes: How clean was each street? Using this information, it reassigned its street cleaners and began to reward crews that made the greatest improvements. By 1986, the percentage of streets rated "filthy" had declined from 43 to 4 percent. Nearly 75 percent were rated "acceptably clean."

3. Critique of Performance Agreements in the United States

The performance agreement documents produced by the National Performance Review office headed by Vice President Al Gore are the policy tools crafted to deliver on the promise of the GPRA. These performance agreements have several strengths and weaknesses.

3.1. Strengths

Performance agreements reflect an attempt to instill accountability for results (performance) at the highest levels in the government. While the concept of accountability for performance in the public sector is not a new one, it usually has been implemented at lower government levels. There are several problems with the bottom-up approach. First, it lacks credibility. It is hard for people who are not accountable for results to impose the discipline of results-oriented management on others. It is much easier for accountability to trickle down than it is for it to seep up the bureaucratic chain of command.

Second, if the environment is vitiated by a culture of weak accountability for results, it is difficult to make much of a difference by imposing a system of accountability in one corner of the government. In fact, some economists would argue that to achieve second-best outcomes it may be necessary to depart from the first-best solution of a rigorous system of accountability in one part of the government, if other parts of the government do not have an effective system of accountability. Others would argue that for accountability to work in various parts of the government, it is necessary to create an enabling environment. The U.S. policy of establishing performance agreements at the cabinet level can be viewed as an attempt to create the enabling environment needed to implement

GPRA in all parts of the government.

Third, accountability systems work best when they have the full support of the chief executive in the government. Implementation of performance agreements in the United States was not only spurred by the White House, this policy was clearly seen to have the highest-level political support. Each performance agreement was signed by the President Bill Clinton and the cabinet secretary responsible for the relevant department.

Fourth, the structure of these documents is in line with best practice in this area. Each document contains the following five sections:

1. Introduction
2. Major Objectives of the Department
3. Measurement of Performance
4. Administration Support
5. Terms of Agreement

Finally, the speed and efficiency with which these documents were drafted is commendable. Unlike typical government documents, these documents are readable; they are written in simple, workmanlike language devoid of legal and bureaucratic jargon.

3.2. Weaknesses

In spite of a first-rate effort on most fronts, the performance agreement documents suffer from a fatal flaw: They have a weak methodology to evaluate performance. More accurately, performance agreements do not have any mechanism to evaluate performance against the commitments made by the two parties in body of these document. The rationale for the performance agreement approach is based on the positive benefits that accrue as a consequence of government's ability to measure performance. If the government cannot or does not measure performance, then *ipso facto* it forgoes the benefits. The main weaknesses of the evaluation system can be summarized as follows.

3.2.1. Lack of Prioritization of Objectives and Performance Measures

The performance agreement between the president and the secretary of the interior includes seven objectives and thirty performance measures for fiscal year 1994 (Appendix I). Let us say, for the sake of argument, that at the end of fiscal year 1994, the secretary of the interior reports to the president that he has achieved targets for twenty-four performance measures out of the thirty listed in the performance agreement. How should the president evaluate the secretary's performance?

The short answer is that the president cannot even begin to evaluate performance in an objective manner. It is possible that the secretary did not achieve the agreed targets (performance

measures) in the most critical areas but was able to achieve targets in other, less important areas. Thus, whatever action the president takes, it can be questioned. Clearly, the secretary would like to claim that the twenty four targets he achieved are the ones that really count. His detractors would argue that the six targets he missed were more important. Any method to resolve this conflict is likely to appear subjective and hence undermine the credibility of the evaluation exercise.

To overcome this problem, it is common to agree on an explicit prioritization of performance measures. In the example given here, this could be done by attaching weights to each of the seven major objectives specified in the Interior Department's performance agreement and dividing them among the thirty performance measures listed under these objectives. Table 9.1 gives a hypothetical example of how this could be done for the Department of the Interior. This table simply takes the indicators mentioned in the Department of the Interior's performance agreement and restates them in the suggested format.

3.2.2. Lack of Clarity with Respect to Targets

This category includes three kinds of problems. First, for some performance measures it is not clear how the performance will be measured. Take, for example, the following performance measure under the objective "Act as a Partner with Indian Tribes."

"The secretary will emphasize continued consultations with the tribes and Congress to develop an appropriate vehicle for the management of Indian trust funds including alternatives other than operations within the Bureau of Indian Affairs."

This performance measure does not make clear how to determine whether the secretary did indeed "emphasize continued consultations." For instance, will a simple letter from the secretary to this effect fulfill the requirement to "emphasize continued consultations?" Would it not be far better to require the secretary to develop an appropriate vehicle for management of trust funds than to ask him to merely "emphasize continued consultations?" Conventional wisdom regarding performance evaluation suggests that output-oriented performance measures must be given preference over process-oriented indications.

The second problem relates to the way in which some of the targets have been stated. For example, take the following criteria under the objective "Strengthen the Commitment of the National Park Service to Employees and the American Public."

"The secretary will make significant progress in FY 1994 to develop a mechanism to provide seasonal and temporary employees greater benefits ..."

"... concession reform will commence with new negotiations of current contracts early in FY 1994."

In both examples, the target is loosely defined. In the first example, the phrase “significant progress” allows great latitude for convenient interpretation. Similarly, the phrase “early in FY 1994” introduces fuzziness in evaluation.

Finally, it is not clear how the deviation from the targets would be measured. For example, the 1994 performance agreement between the president and the secretary of veteran’s affairs included the following target:

“Accelerate automation of patient clinical records and give medical providers more rapid access to clinical information in order to improve service to veterans by implementing automated exchange of patient data between 135 medical centers.”

Assume that at the end of 1994, the secretary for veterans affairs reported to the president that he managed to implement automated data exchange between 120 centers, as against the target of 135. How should the secretary’s performance have been evaluated? In the absence of any explicit agreement on how to measure deviations from the targets, any conclusion regarding the secretary’s performance could be questioned. For this reason, current best practice requires agreement on a scale of targets rather than on a single point. This eliminates any subjectivity at the end of the evaluation cycle. For example, a five-point scale for this target might look like the one shown in Table 9.1.

Table 9.1: Example of a Target Scale

Criterion	Unit	Criterion Values				
		5	4	3	2	1
		Excellent	Very Good	Good	Fair	poor
Implementation of Automated Data Exchange System	No. of Medical Centers	140	135	130	125	120

This five-point scale permits a precise judgment of performance. The actual number of points on the scale is not important. However, it is important that the two parties to the performance agreement agree to a specific scale when the agreement is signed.

3.2.3. Lack of a Methodology for Overall Evaluation

Prioritizing objectives (by assigning weights that add up to 100 percent) and defining the

targets more precisely (using a scale) are the ingredients of a proper evaluation. However, these ingredients have to be processed to achieve fair and objective evaluation. Therefore, it is equally important to agree on a performance evaluation methodology.

The following example illustrates how the performance agreements could be reformatted to enable overall performance evaluation. To simplify exposition, three criteria are included in the hypothetical performance agreement for the Department of Veterans Affairs signed at the beginning of the year (see Table 9.2).

**Table 9.2: Recommended Structure for a performance agreement
(Signed at the Beginning of the Fiscal Year)**

Criterion	Unit	Weight	Criterion Values				
			5	4	3	2	1
			Excellent	Very Good	Good	Fair	poor
1. Implementation of Automated Data Exchange System	No. of Medical Centers	.50	140	135	130	125	120
2. Reduce internal Directives	Percentage of Reduction	.20	11%	10%	9%	8%	7%
3. Train 50% of the Employees in Total Quality Management	Months	.30	9.5	10	10.5	11	11.5

Assume that at the end of the year the performance reported by the secretary of the Department of Veterans Affairs to the president is as shown in Table 9.3.

Table 9.3: Achievements at the End of the Fiscal Year

50

Criterion	Unit	Achievement
1. Implementation of Automated Data Exchange System	Nos. Of Medical Centers	135
2. Reduce internal Directives	Percentage of Reduction	8.5%
3. Train 50% of the Employees in Total Quality Management	Months	9

Table 9.4 illustrates a methodology for evaluating the overall performance of the Department of Veterans Affairs based on the comparison of ex-post performance with the ex-ante agreement on the targets and their relative priorities. Similar methodology is used for evaluating performance in the public sector in several countries such as South Korea, Thailand, Costa Rica, India, Pakistan, and Colombia.

Table 9.4: Overall Performance Evaluation

Criterion	Unit	Weight	Achievement	Raw Score	Weighted Raw Score
1. Implementation of Automated Data Exchange System	No. of Medical Centers	.50	135	4	2
2. Reduce internal Directives	Percentage of Reduction	.20	8.5%	2.5	.5
3. Train 50% of the Employees in Total Quality Management	Months	.30	9	5	1.5
COMPOSITE SCORE					4.00

The calculation of the overall performance score (composite score) involves three steps. First, by comparing the actual achievement with the agreed criterion values, we get a raw score. The raw score for Criterion 1 in Table 9.4 is “4” because 135 is listed under the column number 4 in Table 9.2. Since we cannot add these scores because each has a different level of importance (weights), we call them raw scores. The second step involves multiplying each raw score by its respective weight

to obtain the relevant “weighted raw score” in Table 9.4. Finally, by adding all the weighted raw scores we obtain the “composite score” representing overall performance of the Department of Veterans Affairs.

The beauty of the composite score is that it gives a precise overall evaluation on a five-point scale. This composite score can be used to compare various departments and thereby generate benchmark competition among government departments and agencies. While the commitments, targets, and weights for these departments will be different and cannot be compared on a one- to-one basis, the departments’ ability to meet these commitments can be compared using the composite score.

3.2.4. Lack of Clarity on Consequences of Failure of Administration Support

Each performance agreement has a section entitled “Administrative Support.” Not only is this in line with international best practice, but it is a *sine qua non* for performance agreements. As mentioned earlier, a performance agreement is supposed to be an agreement between two parties in which the obligations of both parties in achieving the agreed targets are mentioned.

There is much to be said for the way some of the items under the “Administrative Support” section are listed. For example, let us examine the following item in the performance agreement for the Department of the Interior:

“The Office of Management and Budget will provide generic clearances for customer surveys within two weeks of submission of requests by the Department of the Interior.”

This is a good way to list conditions that are required to achieve the targets. Unfortunately, the problem with this formulation is the absence of any agreement on the consequences of failure on the part of the administration to meet this condition. The appropriate way to handle this situation is to list which targets will be affected and how they will be affected by the failure to meet this condition.

In addition to the problem just mentioned, there is the problem of lack of clarity with respect to some of the items listed. For example, in the following item listed in the performance agreement for the Department of the Interior, the term ‘support’ is not clearly defined.

“Support passage of authorizing legislation to establish the National Biological Survey.”

3.2.5. Unclear Institutional Mechanism for Evaluation

The mechanism for performance evaluation is unclear. performance agreements are agreements between the president, representing the administration, and cabinet secretaries, representing their

respective departments and agencies. However, the evaluation is expected (implicitly) to be done by the president, and this presents an apparent conflict of interest. It does not seem appropriate for one party to the agreement to be the judge as well. In other places where this type of institutional arrangement has been tried, these instruments have tended to be one-sided documents and the weaker parties have felt the treatment to be unfair.

One possible alternative would be to have the agreement signed between the vice president and various cabinet secretaries. In that event the president would be the appellate authority that would be expected to act as a neutral third party.

However, before we conclude that this would be the best arrangement, it is important to point out that there is a positive side to the current institutional arrangement. The president's signing of the agreement lends the entire exercise enormous prestige and could in many circumstances help compensate for the problems resulting from an agreement between two unequal parties.

4. Current Status of the Performance Agreement Policy

In spite of the fanfare with which this policy was launched, there was no clear follow-up. The performance agreements signed in 1994 have never been evaluated (or at least the evaluation has not been made public). There seem to be several reasons for this. To start with, as this paper has tried to establish, it would have been impossible to evaluate performance on the basis of the 1994 performance agreements even if the Administration had wanted to do so. The technical elements required for a successful ex-post evaluation were missing from the structure of the 1994 performance agreements.

Another reason appears to be the political risk involved in evaluating cabinet secretaries. The Administration was worried that the performance evaluations based on the 1994 performance agreements would be used by the Republican opposition for narrow political purposes.

In a recent best-seller entitled *Banishing the Bureaucracy*, the father of the Reinventing Government movement, David Osborne, offers an explanation as to why the United States and Canada have not managed to install performance evaluation systems at the highest levels of government while New Zealand, the United Kingdom and Costa Rica have done so. According to Osborne, the top executive in the government must be willing to devote his personal attention and political capital to the task of improving government performance. He must be willing to invest the time, energy, and resources—the blood, sweat, and tears—it takes to reinvent government. In short, the head of the government must have enormous courage.

5. Lessons for Developing Countries

Developing countries have tried to reform their bureaucracies for a long time. The common

jargon for the standard approach is “civil service reforms.” Unfortunately, these reforms have focused largely on the input side: reducing the number of civil servants, reforming the salary structures, modifying civil service recruitment and promotion procedures and so on. Many of these reforms have been driven by fiscal imperative of reducing the expenditures on the government machinery. The current wave of reforms symbolized by the performance agreement approach in the United States shifts the debate from *quantity* of government to *quality* of government and from the *cost* of government machinery to *cost effectiveness* of the government machinery.

The evidence from successful reforms suggests that the returns from this type of reform are well worth the effort required. This is particularly true in developing countries, where the current crisis is not one of lack of ideas, but one of poor implementation of ideas. Good ideas and policies do not implement themselves; they require a dedicated cadre of public servants. This dedication comes not from some selfless pursuit of public good, but from a well-designed system of accountability for results and from incentives for public servants to respond in the appropriate manner. A well-designed government machinery is capable of implementing any policy effectively, whereas a poorly functioning government machinery is likely to botch up any policy, however well thought out it may be. In other words, reform of the accountability system creates a tool that enables the government to implement all other reforms effectively.

The U.S. experiment has several useful lessons for developing nations attempting this paradigm shift. First, accountability for results must start at the very top. If top-level officials are held responsible for results, they will hold those below them accountable for results. It will take time, but soon the culture of accountability will trickle down to all levels of the government machinery. Even though the 1994 performance agreements apparently were not evaluated, they set in motion a process of greater focus on end results within the government agencies.

Second, the government must start by defining its objectives at the highest levels and develop the capacity to evaluate whether the people in charge of achieving them are succeeding. As the old saying goes, “What gets measured gets done.” Furthermore, if they are to have any effect on the behavior of the people being evaluated, the results of this exercise should be made public.

Third, all other systems in the government—financial, procedural and organizational—should be adapted to increase the success of achieving these objectives. That is, the criterion for judging reforms in the government machinery should be whether the proposed changes help the government become more successful in achieving clearly measurable objectives through an objective performance evaluation system.

Finally, the documents must be structured so that they enable objective ex-post performance evaluation. This article has concentrated on this aspect because in the final analysis the benefits of this approach depend on the ability to evaluate performance. The vast documented experience in this field should allow developing countries to adopt best-practice methodologies and avoid reinventing the wheel.

Appendix I

Appendix II