

August 7, 2012

To: Government Performance & Finance Committee

From: Peter Harris, Central Staff

Re: Performance Measurement and Program Evaluation: An Incremental Approach for Consideration

Introduction

The Government Performance & Finance Committee soon will consider a resolution that would establish general standards for performance measurement and program evaluation and state the Council's intent to apply these standards in reviewing the 2013-2014 Proposed Budget. This memo describes the framework for performance measurement reflected in the proposed resolution.

The memo has four sections. The first section presents six ideas underlying the framework. The second describes the framework. The third describes standards for evaluation, which comes into play within the framework. The fourth suggests an incremental approach to implementing the framework beginning with program reviews by the Council.

One reason to develop and apply these standards is to improve the City's accountability to the public. Another is to encourage periodic reassessment of the purposes of City programs and of the City's effectiveness in delivering these outcomes, so that programs are as efficient and effective as possible.

The questions and analyses called for by these standards are not new. Many City staff in line departments, the Budget Office, the Auditor's Office and Council ask these kinds of questions frequently. The ways in which many City programs are expected to produce results, however, are often very complex, especially in the high priority areas of public safety and human services. In some cases there is a significant risk that a program will make no difference or even make things worse. A systematic approach to performance measurement and evaluation would help ensure that programs make things better.

Six ideas about performance measurement

1. Performance measurement essentially consists of defining what one intends to accomplish and determining whether one is achieving those intentions. That is, it consists of defining and measuring goals and the means for achieving them.
2. Within the City, one person's goal is often another person's means. For this reason, performance measurement will not provide single standardized prescription of goals and means for the City as a whole. Instead it is a structured way of defining goals within the specific domain to which they are appropriate and describing the success or failure of the chosen means in reaching those goals.

3. Evaluation is an element of performance measurement. Its purpose is to determine whether the chosen means are in fact making progress toward the goal. Evaluation is what brings evidence to bear.
4. Cost/benefit analysis is another element of performance measurement. It compares the cost of the means for making progress toward the goal to the value of the progress made.
5. Performance measurement is the policy and oversight part of performance management and thus is the appropriate focus of the Council. The other part of performance management, perhaps the larger part, is managing employees to achieve the intentions. This is a role for the Executive branch.
6. Implementing performance measurement is probably best done incrementally. Past City efforts to create and impose universal performance measurement systems have not succeeded, and the City's experience is not unique. An incremental approach would give the Council and the City the opportunity to learn from experience.

A performance measurement framework

Performance measurement has four essential features:

1. A well-defined measurable goal.

What would we like to achieve in the long term? "Measurable" is key, because we must have a way of knowing how far we are from the goal. This does not mean every goal must be defined in precise quantitative terms, only that it be concretely observable. For example, "a livable city" is not a well-defined measurable goal. Often the definition of a goal will involve defining or adopting one or more standards or benchmarks. For example, the goal of "safe drinking water" can be made observable by adopting Health Department standards.

2. Measurement of the status quo with respect to the goal

What is the current state of affairs? How far from the goal are we today? Also, what will happen if we continue our current level of effort? Are factors outside our control driving us toward or away from the goal?

A forecast will be easier for some goals than for others. For example, if the goal is streets with few potholes, the future is a reasonably predictable function of expected wear and tear due to traffic and the nature and timing of street maintenance. If the goal is to reduce aggravated assaults, change in the external drivers will be harder to predict and a forecast of what will happen without new City action will be more difficult.

3. An estimate of the effectiveness of action in making progress toward the goal

What can we do to move closer to the goal? How much will it cost, and how do we know it will work?

This is where evaluation occurs, and also is the beginning of cost/benefit analysis. The purpose of evaluation is to determine whether an action is effective in making progress

toward the goal. The purpose of cost/benefit analysis is to compare the cost of action to the value of the progress it will make, based on the evidence.

When budgets must be cut, the questions take a different form, but the logic is the same. How much do we save by incrementally reducing the current effort? How far does each decrement of action take us from the goal, and what is the value of this loss?

4. A short- or medium-term target

We tend to have ambitious goals, and many of them, and limited means. How much effort can we afford on a given goal now? What is feasible to accomplish in the near future? The cost of achieving the target becomes the budget.

Also, if there is uncertainty about the forecast or the effectiveness of our chosen means, setting an interim target gives us the opportunity to review our progress and reassess the forecast and the means.

This is the place for cost/benefit analysis at the program level. If we have multiple programs in the same domain, program cost/benefit analyses will help us find the most cost-effective mix of programs.

In some cases, when means are limited, the target may be simply maintenance of the status quo. For example, for the next few years we might be satisfied by having no more potholes than today, or only a few more.

Standards for evaluation

The purpose of evaluation is to determine whether a program is effective in making progress toward the goal. Effectiveness means that the program makes a difference in outcomes. This requires defining and measuring the intended outcomes. It also requires a solid basis for knowing what the outcomes would have been without the program.

One way to know whether a program is effective is to conduct a local outcome evaluation. Here are three basic standards for local outcome evaluations:

- (a) The evaluation must define and measure one or more outcomes embodying the purpose of the program.
- (b) The treatment group for the evaluation must include all those whom the program intended to treat, not only those who completed the treatment or who were willing to participate.
- (c) The evaluation must include a control or comparison group for whom the same outcomes are also measured.

The difference between the outcomes for the treatment group and the outcomes for the control or comparison group is the measure of the effectiveness of the program.

A second way to know whether a program is effective is to replicate a model program that has been evaluated elsewhere and shown to be effective. Here are some standards for replication:

- (a) The model program being replicated must have been shown to be effective by one or more evaluations meeting the standards for local outcome evaluations.
- (b) The model program must have been successfully implemented in multiple sites.
- (c) The local program must accurately follow the model. This means the local program must adhere to the model program design; the persons or situations receiving the local program treatment or services must receive the same dosage or exposure to treatment or services as in the model program; and the local program must be delivered with the same quality as the model.
- (d) The local program must measure the accuracy with which it follows the model. This is known as measuring implementation fidelity.

Sometimes we may want to replicate a model program and also conduct a local outcome evaluation. We may want to replicate a program that is promising, but which does not yet have solid evidence or otherwise has not met replication standards. Replicating a program for which the evidence is somewhat weak or which has been implemented in only one location may offer a better chance of success than developing a new program from scratch. In this case a local outcome evaluation would be required, and an evaluation of implementation fidelity may be useful.

One objection sometimes raised to these standards for local outcome evaluation is that they call for true experimental designs with randomly selected treatment and control groups in every case, and that this often is completely infeasible or, if feasible, unreasonably expensive. Random controlled trials are indeed the gold standard for evaluation, but are not necessarily too expensive, and in any event are not required by these standards. Good quasi-experimental designs can suffice.

In many instances a control or comparison group may be readily at hand or implicit. For example, if we wanted to know whether a new way of channeling traffic on a stretch of road is successful in reducing speeds and collisions, we might start by comparing speeds and collisions before and after the new channels are created. In itself this comparison would not meet the standards because there is no explicit control group. But in fact it may include an implicit comparison group, in the form of neighboring similar roads with stable speeds and collision rates.

Another common objection is that the cost of any evaluation reduces the amount available for the service itself. But if the service is not effective, none of the money is doing any good. Also, evaluations generally are one-time and services continue. A \$100,000 evaluation, for example, might seem expensive for a new program that costs \$200,000 per year. But after ten years the program will have cost \$2 million. Thus the choice would be between spending \$2 million for unknown results or \$1.9 million for a program that we know makes a difference. In addition, evaluation can not only give us confidence that the cost of a program is not being wasted, it can also give us the ability to select the program that has greatest possible benefits for the problem at hand.

There are also ethical considerations. Sometimes the City represents to the public that a program is effective, but has little basis for this claim. This is compounded by the risk that an unevaluated

program is not merely ineffective and a waste of public resources, but makes things worse. This risk is real, as the Council soon will see when it reviews the upcoming Auditor's report on the George Mason University assessment of the City's crime prevention programs.

The Council's role in an incremental approach to performance measurement

One reason to implement this framework incrementally is that the nature of goals and means varies greatly across the many lines of business in the City, and the framework will be useful only as it is tailored to the specific goals and means in a given line. Another is that the framework is more likely to make a difference if it is shown to be useful case by case than if it is widely imposed as an administrative requirement.

A third reason is that making it useful will require learning by all. These general ideas, drawn as they are from standard concepts of policy analysis and often applied already in many places in the City, nevertheless are not how the City routinely describes what it is trying to accomplish, and applying them will not always be easy. Where programs have been focused for years on outputs, it may take work to define and get agreement on the desired outcomes. Where goals are clear but effectiveness is uncertain, evaluation will be the focus. Departments and the Council will have to learn together what is most useful.

But if performance measurement is important for public accountability, the Council should not merely suggest this framework and wait for line managers to use it. What should the Council do?

One place to start would be in the Council's own review of programs, one at a time, as they become issues for the Council in the normal course of business. The Council's role would include confirming or revising the Executive's judgments on the definition and importance of the long-term goal, on the value or benefits of making a given amount of progress toward the goal, and on the appropriate short- or medium-term target to be reflected in the budget. The Council would want the Executive to provide a reliable description of the status quo, reliable forecasts of future conditions when feasible, and reliable estimates of the cost and effectiveness of the actions needed to achieve the target.

It would be reasonable for the Council to seek all of the information described in the framework for the programs it chooses to review in this way, but it may not be reasonable to expect all this information to appear at the outset. We should expect the City to be able to describe what it is trying to accomplish and whether it is succeeding. The framework outlined here raises the standards for what counts as a good answer to these questions.

Conclusion

Ultimately this framework will matter only if it is used in making decisions about the course of programs and the resources devoted to them. If the Council uses it this way, departments will follow. A few successes will count more than many declarations of intent.

An incremental approach could take a number of forms. The proposed resolution says the Council will apply the framework in upcoming budget deliberations on programs that are new or significantly changed. Councilmembers may seek the same kind of information about some continuing programs. Beyond the first cases, if the framework is useful, a question for the Council and the Executive

together will be how to develop capacity in the City for this kind of work. On one hand there is value to independence in measurement and evaluation. On the other, the more that a program would benefit from this approach, the more the approach should be integrated into the core of the program by the line department responsible. These and other questions of implementation would benefit from more and broader discussion.

I hope these thoughts are useful. If you have any questions, or would like further explanation, please let me know.