Performance Management and Evaluation of Programs in the Public Sector

Kathryn E. Newcomer, Ph.D.
Associate Director, The Trachtenberg School of Public Policy and Public Administration
The George Washington University
June 3, 2008
Issues Addressed:

1. What are current pressures and opportunities for managing programmatic performance in government?
2. What does performance measurement entail?
3. What are challenges to measuring and evaluating performance?
First things first…

**Program Evaluation** is: The application of systematic analytical (social science research) methods to address questions about program operations and results; and

**Performance Measurement** is: The routine measurement of program inputs, outputs, intermediate outcomes or longer-term outcomes attributed to a program; and

**Both involve measurement plus judgment!!**
• Evaluation and/or programmatic performance measurement of programs is undertaken in order to improve programs and their outcomes -- through providing useful and timely information about programs.

• Performance data may be collected routinely, such as annually, and can address questions about quantity and perceived quality, e.g., “how many clients were served”, and “how satisfied were they with the services?”

• A more extended effort to evaluate programs is need to answer “Why” and “How” and “With what result” questions about services

• Both performance data and evaluation studies undertaken can be useful in the exercise of accountability
And Performance Management?

- Performance Management typically means that performance data are used to inform decision-making about planning, management and resource allocation in public agencies.
Current Pressures for Performance Assessment:

1. International Success Stories, e.g. New Zealand
2. State and Local Governmental Successes in U.S.
3. The Governmental Accounting Standards Board calls for “service efforts and accomplishments” reporting
4. Demands for accountability by nonprofit service providers for both programmatic and community level results by foundations and private funders.
5. TQM, Quality Management initiatives focusing attention on customers
Federal Legislative Initiatives:

2. Government Performance and Results Act of 1993
4. Chief Information Officers Act of 1996
5. Federal Financial Management Improvement Act of 1996
Federal Government Executive Initiatives:

1. The NPR (National Partnership for Reinvention)

2. Office of Management and Budget (OMB) requirements for performance information to accompany budget requests (May 5, 1992)

3. Executive Order 12862 - To survey customers

4. The President’s (George W. Bush) Management Agenda – and especially the PART Process and PART stands for the Program Assessment Rating Tool
Current Pressures on Performance Measurement Efforts in Federal Agencies

- The President’s Management Agenda:
  - Calls for “evidence that programs work,” but no funding for it!
  - Pressure for cost benefit studies
  - Calls to Align Budget Accounts with Programs using the PART instrument to rate programs
  - Calls for “unit costing”
  - Call to evaluate the effectiveness of similar programs in different agencies
PART Focus on Program Results

• A set of questions addressing program results is to be answered with “Yes, “Large Extent”, “Small Extent,” or “No.”

• The questions include three on achievement of performance goals, one comparing to other programs with similar purpose and goals, and one on effectiveness.
Key PART Question on Program Effectiveness

- Do independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results?
Where We Are Today
Distribution of Cumulative Ratings 2002 - 2006

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Results Not Demonstrated</td>
<td>50%</td>
<td>38%</td>
<td>29%</td>
<td>24%</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>Ineffective</td>
<td>6%</td>
<td>11%</td>
<td>15%</td>
<td>15%</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>Adequate</td>
<td>24%</td>
<td>26%</td>
<td>26%</td>
<td>29%</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>Moderately Effective</td>
<td>15%</td>
<td>20%</td>
<td>26%</td>
<td>28%</td>
<td>28%</td>
</tr>
<tr>
<td></td>
<td>Effective</td>
<td>5%</td>
<td>20%</td>
<td>4%</td>
<td>4%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Legend:
- Results Not Demonstrated
- Ineffective
- Adequate
- Moderately Effective
- Effective
Discussion of PART

• The PART process has been very political--
  – Why is it not surprising that 51% of the programs parted thus far were deemed “results not demonstrated?” i.e., failed to reject the null.
  – Why might it be difficult to conduct a mega-PART on all 45 programs on economic development run in five different federal agencies?
  – Why has the PART process not gained widespread Congressional buy-in to use in their budgetary deliberations?
Performance Measurement Systems

What is needed to ensure effective use of a performance measurement system?

• Program mission and objectives
• Valid performance measures clearly linked to program objectives
• Location and support for an operative, reliable performance information system
• Clarity on how performance information will be used
What are the key challenges we face in developing effective performance information systems?

- Communication (vertically and horizontally)
- Measurement
Performance Measurement Systems - Continued

How are useful performance measures identified?

Need both --

– Strategic plans
– Clear understanding of program objectives
The U.S. General Accounting Office Model of Performance Measurement

Step 1
Define Mission and Desired Outcomes
Practices:
1. Involve Stakeholders
2. Assess environment
3. Align activities, core processes, and resources

Step 2:
Measure Performance
Practices:
4. Produce measures at each organizational level that
   - demonstrate results,
   - are limited to the vital few,
   - respond to multiple priorities, and
   - link to responsible programs
5. Collect data

Step 3:
Use Performance Information
Practices:
6. Identify performance gaps
7. Report Information
8. Use information

Reinforce GPRA Implementation Practices:
9. Devolve decisionmaking with accountability
10. Create incentives
11. Build expertise
12. Integrate management reforms
Definitions of terms:
• Inputs
• Outputs
• Outcomes
• Performance indicators
• Performance targets
GPRA Definitions

Input measure: measure of what an agency or manager has available to carry out a program or activity: i.e., achieve an outcome or output. These can include employees (FTE), funding, equipment and facilities, supplies on hand, goods or services received, work processes or rules.
GPRA Definitions

Output measure: a tabulation, calculation, or recording of a program activity or effort that can be expressed in a quantitative or qualitative manner.

Outcome measure: an assessment of the results of a program compared to its intended purpose.
GPRA Definitions

**Performance target:** a target level of performance expressed as a tangible, measurable objective, against which actual performance can be compared, including a goal expressed as a quantitative standard, value, or rate, e.g., “Improve maternal and child health on tribal reservations to meet 95% of the national standards for healthy mothers and children by 2010.”
So how well does the Model work?

• To what extent are performance measures being used by management to:
  – Support budget requests?
  – Improve program operations?
  – Guide planning efforts?
  – Provide useful benchmarking?
Potential Consequences of Performance Reporting for External Audiences?

• “Do you count what can be counted rather than what counts?” (Einstein)
• Is the rush to measure expanding our capacity or is capacity shaping measurement?
  – Are we adequately auditing validity and reliability of data?
• Are we interpreting the numbers out of context? (any systems thinking?)
• Are calls for “hard evidence of effectiveness” in some areas even within reason?
• What is the impact of setting targets?
  – Threshold effects?
  – Outputs distortion?
• What about rankings?
  – Validity of criteria?
  – Reliability of data used?
Figure 1

The Potential Role of Programmatic Performance Measurement and Reporting
Figure 3
Potential Outcomes of Programmatic Measurement and Reporting

- Ongoing Refinement in Performance Measurement and Reporting
- Knowledge of Program, Process and Results
- Informed Program Management
- Informed Exercise of Oversight
- Improved Link Between Performance and Allocation of Resources.
Where do program managers start in deciding what to measure?

- The use of program logic models has spread across the various governmental agencies and nonprofit agencies.
- The logic model is a tool to identify what the theory underlying a program should be.
A Program Logic Model

Contextual Factors

Program Inputs → Program Activities → Program Outputs → Intermediate Program Outcomes → Longer-Term Program Outcomes
Figure 5. Program Logic: *PreventionWorks* as Gateway to Substance Abuse Treatment

<table>
<thead>
<tr>
<th>INPUTS</th>
<th>MEDIATING FACTORS</th>
<th>OUTPUTS</th>
<th>MEDIATING FACTORS</th>
<th>SHORT TERM OUTCOMES</th>
<th>INTER-MEDIATE OUTCOMES</th>
<th>LONG TERM OUTCOMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Prevention-Works” Staff</td>
<td>Availability of payroll and volunteer effort</td>
<td>Needle exchanges performed</td>
<td>Client ability and willingness to accept referral</td>
<td>Increases in numbers of IDUs presenting for drug treatment</td>
<td>Intermediate term decreases in number of clients injecting drugs</td>
<td>Long-term reductions in levels of injection drug use in the community</td>
</tr>
<tr>
<td>Funding Sources</td>
<td>Numbers of substance abuse counselors and cross-training of outreach workers</td>
<td>Substance abuse treatment referrals made</td>
<td>Client willingness to enter treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability of Substance Abuse Treatment Slots</td>
<td>Levels of private donations and other support</td>
<td></td>
<td>Client ability to get to treatment site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Service Agencies</td>
<td>Political will and governmental budgetary support</td>
<td>Level of coordination and competency between agencies</td>
<td>Treatment provider ability to admit client without delay</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management Information Systems</td>
<td>Other existing IDU prevention programs (contextual variable: impacts entire system)</td>
<td></td>
<td>Ability to track client through system</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. “Mediating Factors” are those which, given variations in levels of their intensity and scope, will positively or negatively affect relevant processes. The red vertical dotted line indicates the point beyond which it would be impractical for *PreventionWorks* to conduct an effective and affordable evaluation.
Other Relevant Performance Measurement Trends

• Comparisons between/among agencies’ performance became institutionalized with:
  – state level and city level benchmarking efforts, e.g., “Oregon Benchmarking”
  – ICMA/Urban Institute performance consortium
  – rankings of federal agencies’ strategic plans, and then performance reports under GPRA
  – ratings of health maintenance organizations (HMOs) by such groups as the California Cooperative HEDIS Reporting Initiative
Comparisons!

• “An organizational report card is a regular effort by an organization to collect data on two or more other organizations, transform the data into information relevant to assessing performance, and transmit the information to some audience external to the organization themselves.” (Gormley and Weimer, 1999)
Consequences of Cross-Agency Comparisons of Performance?

• On behavior of oversight officials?
• On behaviors of clients or consumers?
• On behavior of program staff:
  – Selection of what/who to count?
  – Change in focus to what counts?
  – Other dysfunctionalities?
City of Charlotte Scorecard

• Customer, rather than Financial Perspective is at the top (results level)
In Local Governments, this means:
Defining and achieving the outcomes that are important to your community and delivering responsive, accountable local government with demonstrated effectiveness and efficiency
Prince William’s Outcome Indicators - Effective Government 1996-2000

By the year 2001:

80% citizens satisfied with value for tax $

# adopted County program outcomes achieved will increase by 10% each year

75% citizens satisfied with mix of services

70% citizens who trust County government

92% citizens satisfied with County services

89% citizens satisfied with efficiency and effectiveness of County government
Other Types of Performance Measures

- Accuracy
- Adherence to schedule
- Complaints
- Consumption rate
- Customer satisfaction
- Design specifications
- Efficiency
- Error rates
- Inventory fill
- Maintenance and repair intervals
- Mean failure rates
- Milestone and activities schedules
- Percentage coverage
- Responsiveness rates
- Transactions
- Utilization rates
- Workload
Valid Performance Information

What is needed to ensure valid performance information?

... relevant, competent, and sufficient evidence on program performance
The Rules of Evidence*

- **Relevance:** Does the evidence address the question?

- **Competence:** Was the methodology used to collect the evidence competently executed by competent professionals?

- **Sufficiency:** Is the evidence convincing to the customers/to a reasonable person?

* The Rules of Evidence from the GAO Yellowbook
What constitutes competence?

- *Measurement validity* -- Are we accurately measuring what we really intend to measure?
- *Internal validity* -- Are we able to definitely establish whether there is a causal relationship between a specified cause and effect?
- *Reliability* -- Do our data collection procedures consistently measure the same phenomena?
What constitutes competence? - Continued

... and, if samples are used:

• **External validity** -- Are we able to generalize from the collected data?

• **Statistical conclusion validity** -- Do the numbers we generate accurately detect the presence of a factor or relationship of a specified magnitude?
Criteria for Evaluating Performance Measures

• Clarity of link to program mission
• Timeliness (of measurement of outcomes related to program implementation)
• Vulnerability
• Legitimacy
• Understandability
• Comparability
• Consistency
• Reliability
### Facilitating:

**Internal Factors:**
- Executive Branch Initiatives
- Budget calls for non-financial performance measures
- Legislation
- Laws affecting all programs, e.g. Government Performance and Results Act
- Laws requiring performance measures for specific programs

**Pressures from Environment:**
- Citizens Demands for Evidence of Program Results
- Success stories from other Jurisdictions and other Countries
- Accounting Profession Use of Performance Auditing

### Inhibiting:

**Internal Factors:**
- Insufficient Authority and/or flexibility to Execute Needed Change
- Mixed Signals from Legislative Committees of Use of Measures in budget Process
- Multiple Calls for Measurement in Different Laws and Executive Directives
- Complex Relationships among service Delivery/Regulatory Partners
- Unclear Expectations about Use Performance Data
- Unclear Expectations about Incentives/Punishment for Performance

**Pressures from Environment:**
- Citizen Expectations of Clear Evidence of Program Results
- Anxiety about Comparing Performance across Jurisdictions
- Lack of Comparable, Reliable Data Collection across Jurisdictions

---

**Figure 1 Pressures on Public Managers to Measure Program Performance**
Barriers to Effective Use of Measures

1. Time and cost of collecting performance measures.
2. Perceived fear of exposure, accountability.
3. Difficult to get PM to be seen as important and owned by both employees and management.
4. Outdated information systems and methods.
5. Disconnect between community, program measures.
6. Difficulty making fair comparisons, and finding appropriate benchmark partners.
7. Cannot say why results are occurring.
8. Lack of continuity of elective officials.
9. Pressure to set targets before actual performance is known.
Effective Measurement Practices

1. Involve staff, management, elected officials, and citizens as stakeholders.

2. Create a climate/conceptual framework to support performance measurement.

3. Align performance measures with missions, goals, and objectives.
Effective Measurement Practices, continued

4. Encourage leadership in order to sustain performance measurement efforts.

5. Create linkages to other government processes (strategic planning, budget, etc.).

6. Relate performance measures to existing data, and existing data collection systems.
Effective Measurement Practices, continued

7. Use an iterative approach to allow measures to be refined based on experience and utility.

8. Limit the number of measures to a group of key indicators.

9. Emphasize positive, not punitive results.

10. Avoid setting targets and linking to HR system before system is established.
Nagging Issues in Performance Measurement

1. Measures for Administrative Support?
2. Targets OR NOT?
3. Stakeholder involvement?
4. Time Frame?
5. Top-down vs. Bottom-up?
6. Mission or Political Leadership Change / Conflict?
7. Individual Accountability?