# Performance Measurement and Performance Management

Professor Newcomer
for
The Minerva Program
February 2013

1

# Measurement does not ensure use!

- Performance Measurement:
  - the routine measurement of program inputs, outputs, intermediate outcomes, or end outcomes in order to monitor programs services and results.
- Performance Management:
  - **Using** performance measures to inform planning, decision making, resource allocation, and evaluation activities.

4

# Session Objectives

- Discuss the context and expectations of performance measurement/monitoring
- Explore the challenges to performance management in the U.S. federal government
- Identify common challenges to getting value from measurement of programmatic performance

2

# And...

- **Good performance measures**:
  - center on relevant information about either program operations or desired program outcomes;
  - are relevant and useful to managers and stakeholders; and
  - provide valid and reliable information.
- **Good performance management**:
  - connects planning, program management and, perhaps, resource allocation to performance;
  - informs management and policy debates (does not replace role of leaders and bureaucratic politics); and
  - assists in coordination of support systems.

5

# 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/Monitoring) is:**

The routine measurement of program inputs, outputs, intermediate outcomes or longer-term outcomes attributed to a program; and

**Both involve measurement plus judgment!!**

3

# Overall Performance Management Challenges : Major categories of Challenges

- Program Characteristics
- Structural/Bureaucratic Factors
- Political Issues
- Capacity
- Organizational Culture

6

## Program Characteristics

- Are the outcomes easily measurable?
- Are operations/workload visible?
- How easy is it attribute responsibility for work expended?
- How long does is take to produce desired outcomes?
- How easy is it to measure "success"?

7

## Capacity

- Where will the measurement burden fall?
- How well prepared are people to collect and process the data?
- How adequate is the information technology needed to process the data?
- How deep is the buy-in support among the staff charged with data collection and processing?

10

## Structural/Bureaucratic Factors

- How many layers are there in the bureaucracy in which the program operates?
- Which staff( e.g., budget offices) and program offices are involved and/or responsible for the performance management processes?
- How many different governmental units are involved in program delivery?
- How easy is it to attribute responsibility to the different units?

8

## Organizational Culture

- How committed is leadership throughout the organization to using data to inform decisions?
- How effective is top leadership in enforcing leadership throughout the organization? Is there effective mentoring?
- How tolerant is the culture to taking risks and making errors? Are errors treated as opportunities for learning or are they punished?
- Are there tangible or intangible incentives to use data to inform decisions? How visible are the incentives?
- Is there deeply held commitment to achieving mission driven goals throughout the organization?

11

## Political Issues

- How politically sensitive is the problem being address by the program?
- How visible is the governmental unit politically?
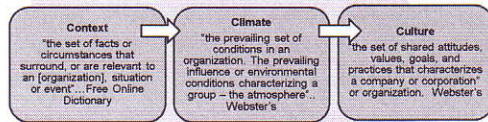- How contentious is the policy arena in which the program operates?

9

## Organizational Culture, cont.

- Is the a clear line of sight to mission from staff at all levels of the organization?
- Are staff dedicated to their work and highly motivated?
- Do staff have the training they need and the information technology they need to succeed?
- Is there a culture supportive of continuous learning?
- Are there any cleavages in the agency, such as across generations?
- Is there a change-centric culture?

12

## Slide 13

### Public and Nonprofit Agencies Work within a Political and Economic Context, but Leaders shape the Internal Climate and Culture

**Context**
"the set of facts or circumstances that surround, or are relevant to an [organization], situation or event"...Free Online Dictionary

**Climate**
"the prevailing set of conditions in an organization. The prevailing influence or environmental conditions characterizing a group – the atmosphere"...Webster's

**Culture**
"the set of shared attitudes, values, goals, and practices that characterizes a company or corporation" or organization. Webster's
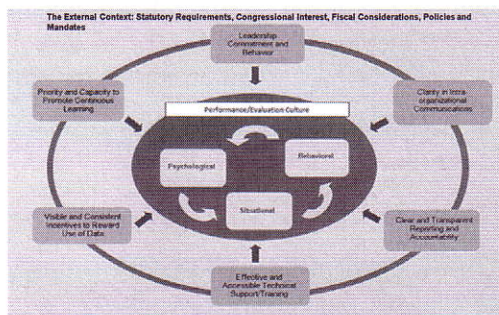
## Slide 16

### Current Context for Performance and Evaluation

- Statutory Requirements
  - GPRA 1993
  - E.O. 13450
    - Performance Improvement Officer (PIO)
    - Performance Improvement Council
  - GPRA Mod Act 2010
- Transparency and Reporting
  - Performance.gov Portal
  - ARRA Stimulus tracking
- Congressional Interest

- Policies/Mandates
  - High Priority Performance Goals
  - Program Evaluation Initiative
  - Emphasis on Impact Evaluations and use of RCTs
- Fiscal
  - Debt/Deficit Choices
  - Fear of Budget Cuts
  - Budget Guidance

## Slide 14

### Leaders Shape the Internal Climate and Culture



The External Context: Statutory Requirements, Congressional Interest, Fiscal Considerations, Policies and Mandates

Leadership Commitment and Behavior

Priority and Capacity to Promote Continuous Learning

Clarity in Intra organizational Communications

Performance/Evaluation Culture

Psychological

Behavioral

Situational

Visible and Consistent Incentives to Reward Use of Data

Clear and Transparent Reporting and Accountability

Effective and Accessible Technical Support/Training

Draft Date: 9/29/2011

## Slide 17

### Government Performance and Results Act of 1993 (Public Law 103-62)

- Part of a series of laws in the 1990s aimed at improving federal government management
- Requires federal agencies to develop
  - **strategic plans** with long-term goals
  - **performance plans** with annual goals and measures
  - **performance reports** on prior year performance
- Also requires a **governmentwide performance plan**, intended to provide a single, cohesive picture of federal performance

## Slide 15

### Performance and Evaluation Culture

- Psychological
  - Individual attitudes, beliefs, and values regarding production and use of performance data and evaluation within the organization

- Behavioral
  - Current practice regarding production and use of performance data and evaluation within the organization
- Situational
  - Reflexive behavior that innately values and relies on performance data and evaluation

## Slide 18

### GPRA Modernization Act of 2010 (Public Law 111-352)

- Among other things, GPRAMA
  - Creates a new government-wide planning and reporting framework
  - Amends agency level planning and reporting requirements
  - Requires leadership involvement and accountability
  - Requires the identification of key performance management skills and competencies
  - Creates an annual process to reduce duplicative and outdated planning and reporting

## GPRAMA Adds Government-wide Planning and Reporting Requirements

- The Act creates a new governmentwide framework including

  - long-term federal government priority goals

  - revised federal government performance plan requirements

  - quarterly priority progress reviews

  - a governmentwide performance website

19

## What have we learned from the Federal Experience with Measurement and Evaluation over the Last Ten Years?

1. There are many different audiences for information about programmatic performance with different expectations.

2. In this era of "evidence based- policymaking" determining what constitutes valid, relevant and reliable evidence for many federal programs is neither easy nor straightforward, and always more challenging than anticipated.

22

## The Current Capacity in the U.S. Federal Government

- A legacy of assets
  - All agencies have strategic plans and performance measures
  - An infrastructure of staff and processes that have been built incrementally to collect and deliver performance data.
- Sophistication in dialogue about programmatic performance
  - within agencies,
  - within the Office and Management and Budget (OMB),
  - within Congress

20

## What have we learned, cont.

3. Attributing measurable outcomes, or impact, to programs is virtually impossible much of the time, despite the political appeal of assessing program outcomes.

4. Advocating a "one size fits all" approach to evaluation (assessment) of federal programs is not effective nor desirable.

5. Collecting performance data and evidence of program results does not translate into improved programs.

23

## The Current Capacity in the U.S. Federal Government, cont.

- Too many measures and too much reporting
- A fairly unwieldly, burdensome and unaligned set of information systems
- A skeptical federal workforce
  - compliance burden mentality
  - few instances of use of performance data to improve government
- Performance Office staff in the agencies that are still battle weary from skirmishes with OMB examiners over PART scoring (The Program Assessment Rating Tool under Bush)
- Inadequate appreciation of linkages between performance management and human resource planning/appraisal and strategic outsourcing

21

## Bush: PART Focus on Program Results

- PART stands for Program Assessment Rating Tool
- A set of about 30 questions addressing program design, management and results is to be answered with "Yes, "Large Extent", "Small Extent," or "No."
- The questions include three on achievement of performance goals, one comparing program to other programs with similar purpose and goals, and one on effectiveness.

24

## PART: Old wine in new bottles?

- What was not new?
  - The focus on program effectiveness by OMB examiners
  - Staff in the Executive Office of the President providing analytical support in preparing and presenting the President's budget
- What was new?
  - Transparency –they used Expectmore. gov
  - Attempting to link GPRA performance measures and targets with assessment of programs' effectiveness
  - Explicit quantitative assessments with scores
  - Raising hard questions about the need for managerial and even legislative changes in program design
  - Raising the impact (results) question, not simply asking for performance measures and targets
  - Explicit guidance given to the need for rigorous methods to ascertain impact (The RCT push!), and more "hard evidence"

25

## Issues Regarding Agency and OMB Use of PART

Why was it not surprising that:

- 28% (50% first year to 22% in 2006) of the programs PARTED were deemed " results not demonstrated?"
- Agency staff found the PART process to be extremely burdensome and criticized OMB examiners for a lack of consistency in scoring across programs.
- Both agency staff and the OMB Examiners found that the construction of counterfactuals is difficult or even impossible for some programs
- Disagreements arose between the OMB examiners and the agency staff on the extent to which performance measures and goals are reasonable
- OMB Examiners (with very little training in evaluation) found it extremely difficult to determine when program evaluation designs are sufficiently rigorous to establish program effectiveness

28

## Key PART Question on Program Effectiveness

- Do **independent** evaluations of sufficient **scope** and **quality** indicate that the program is **effective** and achieving **results**? (question 4.5)
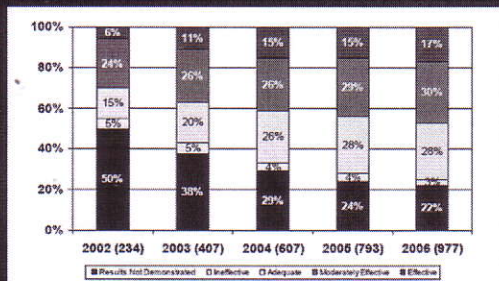
26

## The PART Process Pushed Program Managers and Staff in the Agencies to Think More About Evaluation

- Leveraging resources (e.g. EPA)
- Building evaluation culture

29

### Where We Are Today
#### Distribution of Cumulative Ratings 2002 - 2006



## However...

- The PART process did not involve new evaluations of programs outcomes or impact, but a secondary analysis of existing evidence about both program management and results
- PART scores did not accurately reflect program effectiveness due to multiple questions about other aspects of program operations in the PART instrument

30

## And the Future for Federal Measurement & Evaluation?

- Resources need to catch up with rhetoric - both the collection of data on programmatic effectiveness and the interpretation of such data require substantial resources.
- It is time for a reassessment of performance measurement practice – we need to measure less, and focus more on the relevance, validity and reliability of key measures
- Experience with the PART process both in the agencies and at OMB has provided a great wealth of learning that should be used to design tractable and useful processes for the systematic assessment of program effectiveness in the agencies.

31

## Performance Data Systems Outpace Learning

- Federal MIS systems facilitate the collection of more and more performance measures, but who makes time to make sense of them?



32

## The Obama Administration's Approach to Performance Management

- **Leadership**
- **Prioritization**, and
- **Focus on Use of Performance Data for Improvement**
- **A Performance Portal like Canadas (see http://www.tbs-sct.gc.ca/ppg-cpr/home-accueil-eng.aspx )**

33

## Leadership

- The First OMB Director, Peter Orszag, demonstrated that he places a priority on examining federal programmatic performance-
  - His public commentary on his OMB blog, and
  - OMB memorandum calling for "Increased Emphasis on Program Evaluation," on October 7, 2009
- Appointment of Jeffrey D. Zients, an accomplished management consultant, as Deputy Director for Management of the OMB and 1st ever U.S. Chief Performance Officer, and hire of Shelly Metzenbaum as Zients' Deputy
- Retention of the Performance Improvement Officers (PIOs) within the agencies

34

## Prioritization and Transparency

- Spring budget guidance given to the agencies, OMB Director Peter Orszag asked all major Federal agencies to identify a limited number of high-priority performance goals reflecting the near-term implementation priorities of each agency's senior managers
- By the July 31st deadline,19 of the 24 agencies had submitted their list of priority goals, and all released goals with budget

35

## Focus on Use

- Zients announced in September, 2009 that OMB will develop a Federal performance management framework that will be designed "to serve the needs of agency managers, the public and the Congress"
- Zients also said "I am committed to helping and motivating Federal agencies to maximize the productive use of performance information to improve the effectiveness, efficiency and transparency of government. This is a cornerstone of my agenda and one of my highest priority goals." (See http://www.whitehouse.gov/omb/assets/testimony/092409govern ment.pdf24)
- Metzenbaum IBM 2009 report stressed use; she provided 22 recommendations that prescribe practical steps to improve the use of performance data - such as running "goal-focused, data driven meetings" both of cabinet secretaries, and by the agency heads in the agencies, and focusing on trends, not targets

36

## Performance.gov is finally live!

---

## Some Key Challenges

- Effective and Persistent "Cascading" Agency Leadership for Performance Improvement?
- OMB Performance in Calibrating Context and Incentives to Improve Performance?
- Other Fiscal Pressures and the Political Environment?

---

## Obama Approach: A Focus on Goals for Leaders in U.S. Federal Agencies

- Using goals to improve performance and accountability
- Measuring and analyzing performance to find what works
- Delivering better results using frequent, data-driven reviews

---

The Obama Administration's Performance Management Framework

---
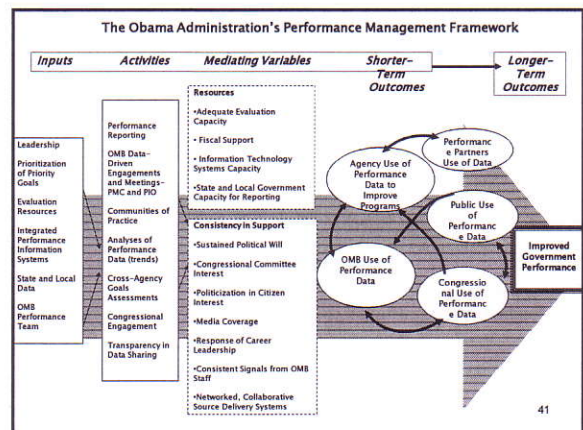
## Some Double-Edged Swords

- Accountability of agency leaders for high priority goals
- "Performance-Stat" type routines for learning (copying the Delivery Unit in the UK under Blair – see the Barber book)
- Focus explicitly on "use" of data
- Transparency of data and goals, e.g. Data.gov
- Congressional Engagement – e.g., HR 2145 and now the Senate Bill on Modernization of the GPRA
- Stretch goals for outcomes
- OMB role as coach rather than police

---

## Chances for Positive Change?

- Pluses: the stress on leadership, prioritization, end of the burdensome PART process, and focus on use for performance improvement (rather than quickly tying to the budget)
- Negatives: **competing priorities**, budget cuts, a legacy of skepticism in the agencies, overly optimistic expectations on:
  - changing the management cultures toward use of performance data
  - addressing cross-cutting programmatic goals, and
  - reducing complexity in reporting systems

## What are common challenges to getting value from measurement of programmatic performance

43

## Types of Measures

- Inputs: measures of the financial and non-financial resources that are applied in providing services.
- Process or service quality: measures that address the merit of a program or service delivery process, including timeliness, customer satisfaction.
- Outputs: measures of the quantity of services provided or the quantity of a service that meets a certain quality requirement.
- Outcome: measures that address the intermediate or long-term results of a program on those receiving a service.

Source: governmental accounting standards board

46

---

## What is the key challenge??

**Ensuring adequate and candid communications**

- **about why measures are needed**
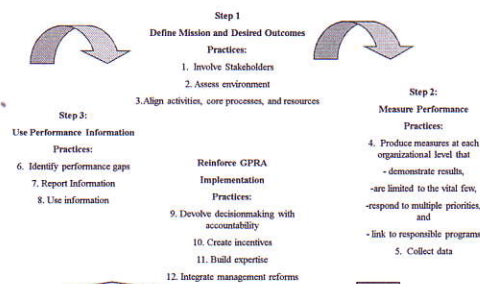- **and how they will be used!**

44

## Types of Measures, cont.

- Effectiveness: Measures of the amount of outcome per unit of input or output
- Efficiency: Measures of the resources used to generate a unit of output
- Explanatory: Relating to factors other than services being provided that may have affected reported performance.

Sources: Governmental Accounting Standards Board, Montgomery County, Maryland.

47

---

### The U.S.GAO Model of Performance Management

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

45

## Targets

*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 2012."

48

## Illustrative 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

49

## "SMART" Criteria for Judging Performance Measures (World Bank)

- Specific
- Measurable
- Achievable
- Relevant
- Time-bound

50

## Kathy's Criteria for Evaluating Performance Measures

- Clarity of link to program mission
- Timeliness (of measurement of outcomes related to program implementation)
- Vulnerability
- Legitimacy
- Understandability
- Comparability
- Reliability

51

## Which Measures Are More Likely to be Used?

- Those clearly linked to mission achievement
  - clarity in link from mission to strategic goals to measures
  - performance data are more likely to be accepted if the program stakeholders perceive the collection process is clear and unbiased –methodological integrity!

52

## Which Performance Routines Are More Likely to be Used?

- Routine, regularly expected reviews without punishment
- Learning forums that:
  - are not confrontational
  - include a wide array of perspectives from within the organization
  - Include and value both quantitative and qualitative data
  - Embody a culture of collegiality and respect

53

## Use of Performance Data in Agency or Program Reviews

- Performance.stat  sessions
- Dashboards
- Balanced Score Cards

54

## Performance.Stat Initiatives

- Started with Comp.stat in New York City with crime data from precincts reviewed in person with Chief of Police regularly
- Current real-time data are needed to make them work
- Can such a real-time process be used in a national level government agency, why or why not?

## Barriers to Effective Use of Measures, continued

5. Disconnect between community indicators and government performance 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.

## Performance Dashboard



Here NA= Needs Attention; MC=Monitor Continuously; and OT=On-Track

Pressures on Public Managers to Measure Program Performance
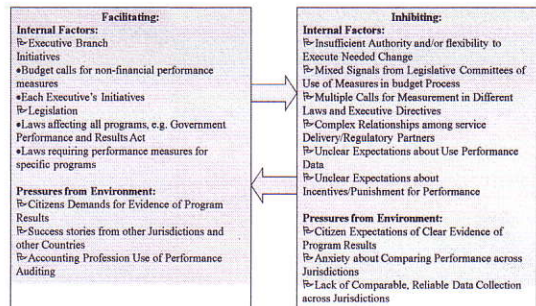
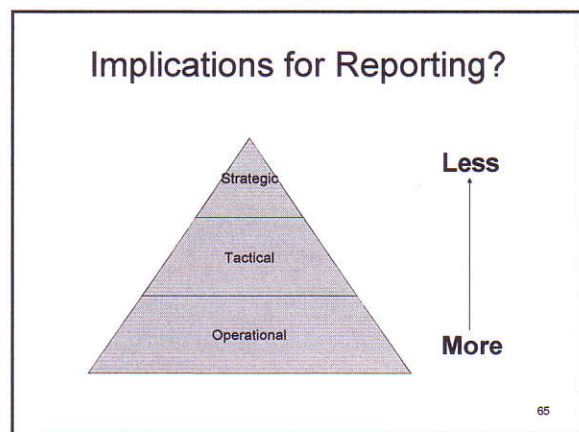## Barriers to Effective Use of Measures

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.

## Performance Measurement and Program Evaluation

- How might performance measurement enhance program evaluation efforts?
- How might program evaluation studies enhance performance measurement systems?
- In what ways might tension between the two arise?

## Slide 61

### The Synergistic Relationship between Program Evaluation and Performance Measurement



**Program Evaluation**
- Design Performance Measures and Systems

**Performance Measurement**
- Refined System and Reporting

**Performance Measurement**
- Ongoing Measurement and Reporting

**Program Evaluation**
- Study / Share promising practices

**Program Evaluation**
- Study trends and address How / Why questions

**Performance Measurement**
- Refined Measurement and Reporting

**Performance Measurement**
- Performance Learning Circles

**Program Evaluation**
- Compare / Contrast delivery variations.
- Improve Measures

61

## Slide 64

# Implications for Information Needs?

**Characteristics of Data Needed?**



Strategic — Fewer metrics, global, trends, comparative, anecdotal success stories, and ??

Tactical — All of above, programmatic outcomes linked to budgets, and??

Operational — All of above, and disaggregated across client characteristics, and ??

64

## Slide 62



Program Evaluation Across the Programmatic Life Cycle → ...ece Organizational Learning

62

## Slide 65

# Implications for Reporting?



Strategic

Tactical

Operational

**Less**

**More**

65

## Slide 63

### What are the implications of the different types of use of data: internal versus external, and strategic, tactical and operational?

- Strategic?
- Tactical?
- Operational?
- And at each level – external or internal users??

63

## Slide 66

### What difference does it make whether the information or data are used internally or externally?

- Can the same data inform both external accountability requirements and management decision-making?
- What are common games played in using data for accountability?

66

## What is needed to produce <u>useful information</u> about program performance?

- An understanding of
  - how a program is supposed to work and
  - the ability to ask useful questions about the program operations

- Methodological Integrity in the production of the answers to the questions!

67

## SO?

Evaluation tools must are needed to accompany routine performance measurement to address the **Why** and **How** questions!

**But** measuring outputs and outcomes helps raise the right questions!

70

## The Range of Questions to Address with Performance Measures, may include:

- **Where and when is performance higher (or lower) than expected?**
- **Are there inequities among beneficiaries– across sites or different types of services?**
- **How does performance compare with –**
  - **Baselines?**
  - **Targets?**
  - **Other providers?**
  - **And???**

68

## Brainstorming on Use/Nonuse of Performance Measures

- What helps, and what hinders?
- Hint: For the third memorandum how I see <u>Moneyball</u> being used…

71

## The Range of Questions that Performance Measures May Raise – but NOT Answer:

- Why is performance higher (or lower) than expected?
- Why are there inequities among customer groups – across sites or different types of services?
- Why does performance fail to meet targets?
- Why aren't outputs resulting in the expected outcomes?
- How are external factors – outside of the control of the program staff affecting the ability of the program to produce the intended outcomes?

69

## Output versus Outcome Controversy?

- The New Zealand outlook
- The politicians' preference
- The cost dimension

72

## Evaluating Outcomes: Targets?

- Benefits of setting targets:
  - Demonstrating "stretch goals" to funders
  - Motivating staff
- Risks involved with setting targets:
  - Setting unrealistic targets based on hunches
  - Converting achievement into a pass/fail scenario

73

## Key National Indicators

- What are advantages of a dash board of a set of social/economic/environmental indicators for a country"
- What are the challenges to making such as an effort viable?

74

## And the Use of Balanced Scorecards?



75

## References

- Barber, Michael. 2008. <u>Instruction to Deliver:  Fighting to Transform Britain's Public Services</u>. London: Methuen Publishing Ltd.
- Government Accountability Office (GAO). 2009. Government Performance:  Strategies for Building a Results-Oriented and Collaborative Culture in the Federal Government. GAO-09-1011T. Washington, DC: U.S. Government Accountability Office.
- Metzenbaum, Shelley H. 2009. Performance Management Recommendations for the New Administration. Presidential Transition Series. Washington, D.C.: IBM Center for the Business of Government.
- Newcomer, Kathryn and F. Stevens Redburn. 2008. Achieving Real Improvement In Federal Policy And Program Outcomes: The Next Frontier. Washington, DC: National Academy of Public Administration.
- Newcomer, Kathryn. 2010. Putting Performance First: A New Performance Improvement and Analysis Framework. In Jonathan D. Breul, Kathryn Newcomer, Joseph P. Goldman, Paul L. Posner, and Steven Schooner. *Framing a Public Management Research Agenda*, 7-16. Washington, D.C.: IBM Center for the Business of Government.
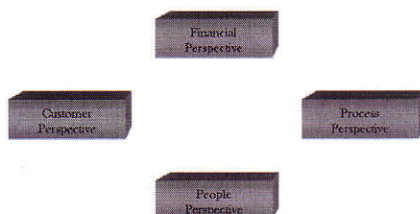
76