

The **Performance-Based Management Handbook**

*A Six-Volume Compilation of Techniques and
Tools for Implementing the Government
Performance and Results Act
of 1993 (GPRA)*

Volume Four

Collecting Data to Assess Performance



Performance-Based Management Special Interest Group (PBM SIG)
<http://www.ora.gov/pbm>

4

The **Performance-Based Management Special Interest Group (PBM SIG)** is a U.S. Department of Energy (DOE) and DOE contractor funded organization made up of DOE and DOE contractor personnel who have a special interest in performance-based management. The mission of the PBM SIG is to facilitate, promote, and advance the use of performance-based management in DOE. The activities and publications of the PBM SIG are coordinated and administered by the Oak Ridge Institute for Science and Education.

The **Oak Ridge Institute for Science and Education (ORISE)** was established by the U.S. Department of Energy to undertake national and international programs in education, training, health, and the environment. ORISE and its programs are operated by Oak Ridge Associated Universities (ORAU) through a contract with the U.S. Department of Energy. Established in 1946, ORAU is a multiuniversity consortium.

This document was produced under a contract between the U.S. Department of Energy and Oak Ridge Associated Universities.

The Performance-Based Management Handbook

A Six-Volume Compilation of Techniques and Tools for Implementing the Government Performance and Results Act of 1993

Volume 4

Collecting Data to Assess Performance

Prepared by the

Training Resources and Data Exchange
Performance-Based Management Special Interest Group

for the

Office of Strategic Planning and Program Evaluation
Chief Financial Officer

Office of Planning, Budget and Outreach
Assistant Secretary for Energy Efficiency and Renewable Energy

Office of Environment, Security, Safety, and Health
Assistant Secretary for Fossil Energy

Office of Laboratory Policy
Director, Office of Science

Office of Performance Management
Pacific Northwest National Laboratory

Strategic Petroleum Reserves Project Management Office
Strategic Petroleum Reserves Operations

Laboratory Administration Office
University of California

September 2001

Written by

Darrell Beschen
U.S. Department of Energy

Richard Day
U.S. Department of Energy

Gretchen Jordan
Sandia National Laboratories

Howard Rohm
U.S. Department of Energy

Preface	
Preface	i - vi
Volume 4 Overview	
Overview	1
Section I: Determining Data Needs	
‘ Determining Data Requirements	3
‘ “Best Practices” in Data Collection	3
‘ Data Quality	4
Section II: Components of a Data Collection Plan	
‘ Components of Data Collection Planning	5
# Statement of Informational Requirements	5
# Statement of Information Sources	5
# Data Collection Process	6
# Data Collection and Reporting Frequency	6
# Data Collection Costs	6
# Data Protection	6
‘ The Trial Run	6
‘ Ensuring Data Quality	7
# Measuring Data Quality	7
# Data Quality Systems	7
‘ Data Collection Systems and Software	8
# Information Systems Considerations	8
Section III: Data Collection Considerations	
‘ “Data Breakouts”	9
‘ Sampling Techniques	10
# Advantages and Disadvantages of Sampling	10
‘ Bias	10
‘ Cost	11
‘ Data Reliability	11
‘ Data Validity	11

- ' Demographic Diversity/Diverse Populations 12
- ' Geographic Scope/Diversity 12
- ' Level of Accuracy 12
- ' Level of Detail 12
- ' Response Rate 12
- ' Speed 13
- ' Stakeholder Input 13
- ' Data Collection Types and Factors Matrix 13

Section IV: Data Collection Methods

- ' Program and Agency Records 15
 - # Advantages and Disadvantages of Program and Agency Records 15
- ' Site Inspections/Observations 16
 - # Types of Rating Systems 16
 - # Advantages and Disadvantages of Trained Observer Ratings 17
 - # How to Implement a Trained Observer Process 17
- ' Use of Special Technical Equipment 18
- ' Surveys and Interviews 18
 - # Benefits of Surveys and Interviews 18
 - # Steps in Designing and Conducting Surveys 19
- ' Purchasing or Using Statistics from an Outside Source 19
 - # Advantages and Disadvantages of Purchasing or Using Statistics From an Outside Source 19
- ' Peer Review/Expert Panel Evaluation 20
 - # Elements of Successful Peer Review/Expert Panel Evaluation 20
 - # Characteristics of Peer Review/Expert Panel Evaluation 20
 - # Seven Ways to Enhance the Results of Peer Review 21
 - # Advantages and Disadvantages of Peer Review/Expert Panel Evaluation 21
- ' Quantitative Methods for Research and Development (R&D) Activities 22
 - # Publication Counts 22
 - # Patent Counts 22
 - # Citation Counts 22
 - # Integrated Metrics and Mixed Methods 22
 - # Strengths of Quantitative Metrics 22

- # Limitations of Quantitative Metrics 23
- ' Economic Methods 23
 - # Strengths of Economic Methods 23
 - # Limitations of Economic Methods 24
- ' Cost-Benefit/Cost Effectiveness Studies 24
 - # Time Frames of Study 24
 - # Identifying Costs and Benefits 25
 - # Valuing Benefits 23
 - # Converting Costs and Benefits into Present Values 25
 - # Methods to Improve Cost-Benefit Studies 25
 - # Advantages and Disadvantages of Cost-Benefit/Cost Effectiveness Studies 26
- ' Case Studies 26
 - # Types of Case Studies 27
 - # Advantages of Case Studies 27
 - # Designing and Conducting a Case Study 27
- ' Content Review 27
 - # Notes on Coding 28
 - # Example Methodology 28
 - # Advantages And Disadvantages Of Content Review 28
- ' File Review 29
 - # Types of File Documents 29
 - # Use of File Review 29
 - # Advantages and Disadvantages of File Review 29
- ' Focus Groups 30
 - # Uses of Focus Groups 30
 - # Advantages of Focus Groups 30
 - # Designing and Conducting a Focus Group 30
 - # The Cost of Focus Groups 31

Section V: Suggestions for Measuring R&D Activities

- ' Suggestions for Measuring R&D Activities 33
 - # Methods Chosen Depend on Type of R&D 33

Appendix A: Definitions

Appendix A: Definitions A-1

Appendix B: Acronyms

Appendix B: Acronyms B-1

Appendix C: References/Suggested Reading

Appendix C: References/Suggested Reading C-1

Preface

... chart a course for every endeavor that we take the people's money for, see how well we are progressing, tell the public how we are doing, stop the things that don't work, and never stop improving the things that we think are worth investing in."

President William J. Clinton, on signing the Government Performance and Results Act of 1993

Introduction

All high-performance organizations, whether public or private, are, and must be, interested in developing and deploying effective performance measurement and performance management systems, since it is only through such systems that they can remain high-performance organizations. When President Clinton signed the Government Performance and Results Act of 1993 (GPRA) into law, this commitment to quality was institutionalized. Federal agencies were required to develop strategic plans for how they would deliver high-quality products and services to the American people. Under GPRA, strategic plans are the starting point for each federal agency to (1) establish top-level agency goals and objectives, as well as annual program goals; (2) define how it intends to achieve those goals; and (3) demonstrate how it will measure agency and program performance in achieving those goals.

The publication of *The Performance-Based Management Handbook, A Six-Volume Compilation of Techniques and Tools for Implementing the Government Performance and Results Act of 1993* follows a logical progression of resources developed to assist in the effective and efficient implementation of GPRA. In chronological order, these resources are:

- The National Performance Review (NPR)
- *How to Measure Performance—A Handbook of Techniques and Tools*
- *Guidelines for Strategic Planning*
- *Guidelines for Performance Measurement*
- *Executive Guide: Effectively Implementing the Government Performance and Results Act*
- *NPR Benchmarking Study Report Best Practices in Customer-Driven Strategic Planning*
- *NPR Benchmarking Study Report Best Practices in Performance Measurement*
- *The Performance-Based Management Handbook, A Six-Volume Compilation of Techniques and Tools for Implementing the Government Performance and Results Act of 1993*

The National Performance Review

In the same year that GPRA was signed into law, President Clinton and Vice President Gore initiated the National Performance Review (NPR) to reinvent government. One of NPR's reinvention initiatives was to foster collaborative, systematic benchmarking of best-in-class organizations, both public and private, to identify best practices in a wide range of subjects vital to the success of federal agencies in providing high-quality products and services to the American people.

How to Measure Performance—A Handbook of Techniques and Tools

Developed in October 1995, *How to Measure Performance—A Handbook of Techniques and Tools* was the Performance-Based Management Special Interest Group's (PBM SIG's) first handbook. It was produced at a time when DOE personnel were struggling with the concepts and conventions of performance measurement

and has been touted as a very useful guidance document. The handbook describes three different approaches to developing performance measures; provides sections on performance indexing, data analysis, and reporting techniques; and includes a thorough glossary of terms, an inclusive list of references, and a substantial list of sample performance measures.

Guidelines for Strategic Planning

This Department of Energy (DOE) guidance document (DOE/PO-0041) was published in January 1996 by the Office of Policy and International Affairs to help strategic planning teams plan for, organize, and prepare the departmental strategic plan required under GPRA. It provides guidance both to those organizations and personnel starting the strategic planning process for the first time and to those reviewing or updating existing plans. The steps outlined within this document represent a very simplified approach to strategic planning.

Guidelines for Performance Measurement

The DOE Performance Measurement Coordination Team released this guidance document (DOE G 120.1-5) in June 1996. It is often referred to as a companion document to the PBM SIG's first handbook. While both documents cover performance measurement, this document also covers the relationship of performance measurement to organizational operations, presenting topics such as performance linking, tying into departmental systems, and coordinating performance measures.

Executive Guide: Effectively Implementing the Government Performance and Results Act

The U.S. General Accounting Office (GAO) published this document (GAO/GGD-96-118) in June 1996. It resulted from a study done at the request of Congress in which a number of leading public sector organizations that were successfully pursuing management reform initiatives and becoming more results-oriented were studied. Each of these organizations set its agenda for management reform according to its own environment, needs, and capabilities. Yet, despite their differing approaches to reform, all these organizations commonly took three key steps to becoming more results oriented: (1) define clear missions and desired outcomes, (2) measure performance to gauge progress, and (3) use performance information as a basis for decision making. These three key steps are discussed in this GAO executive guide, along with their relationship to GPRA. Also discussed is the role of top leadership and the practices it can follow if it hopes to make GPRA a driving force in an organization. Accompanying the discussion of each practice is a case illustration involving a federal agency that has made progress in incorporating the practice into its operations.

NPR Benchmarking Study Report: Customer-Driven Strategic Planning

In February 1997, NPR published its *Benchmarking Study Report Best Practices in Customer-Driven Strategic Planning*, which documents and details the in-depth processes and approaches of those best-in-class organizations that excel at incorporating their customers' needs and expectations into their strategic planning processes. This study provided public and private leaders and managers with world-class practices and formulas for success in developing and deploying strategic plans and goals for an agency.

NPR Benchmarking Study Report: Best Practices in Performance Measurement

To complement its strategic planning study, NPR commissioned the first-ever intergovernmental benchmarking consortium involving not only U.S. federal agencies, but also local governments and the government of Canada in a collaborative study of performance measurement. As documented in its June 1997 report, the NPR Performance Measurement Study Team found that the best performance measurement and management systems and practices work within a context of strategic planning that takes its cue from customer needs and customer service. They also found that:

- Leadership is critical in designing and deploying effective performance measurement and management systems.
- A conceptual framework is needed for the performance measurement and management system.
- Effective internal and external communications are the keys to successful performance measurement.
- Accountability for results must be clearly assigned and well-understood.
- Performance measurement systems must provide intelligent information for decision makers, not just compile data.
- Compensation, rewards, and recognition should be linked to performance measurements.
- Performance measurement systems should be positive, not punitive.
- Results and progress toward program commitments should be openly shared with employees, customers, and stakeholders.

The Performance Measurement Process Model

To provide them with a useful frame of reference as they studied performance measurement in best-in-class organizations, the NPR Performance Measurement Study Team built a model of the performance measurement process used in the federal context. This Performance Measurement Process Model was published in its June 1997 report. This model is shown in Figure PBM.1 on the following page.

Performance Measurement Process Model

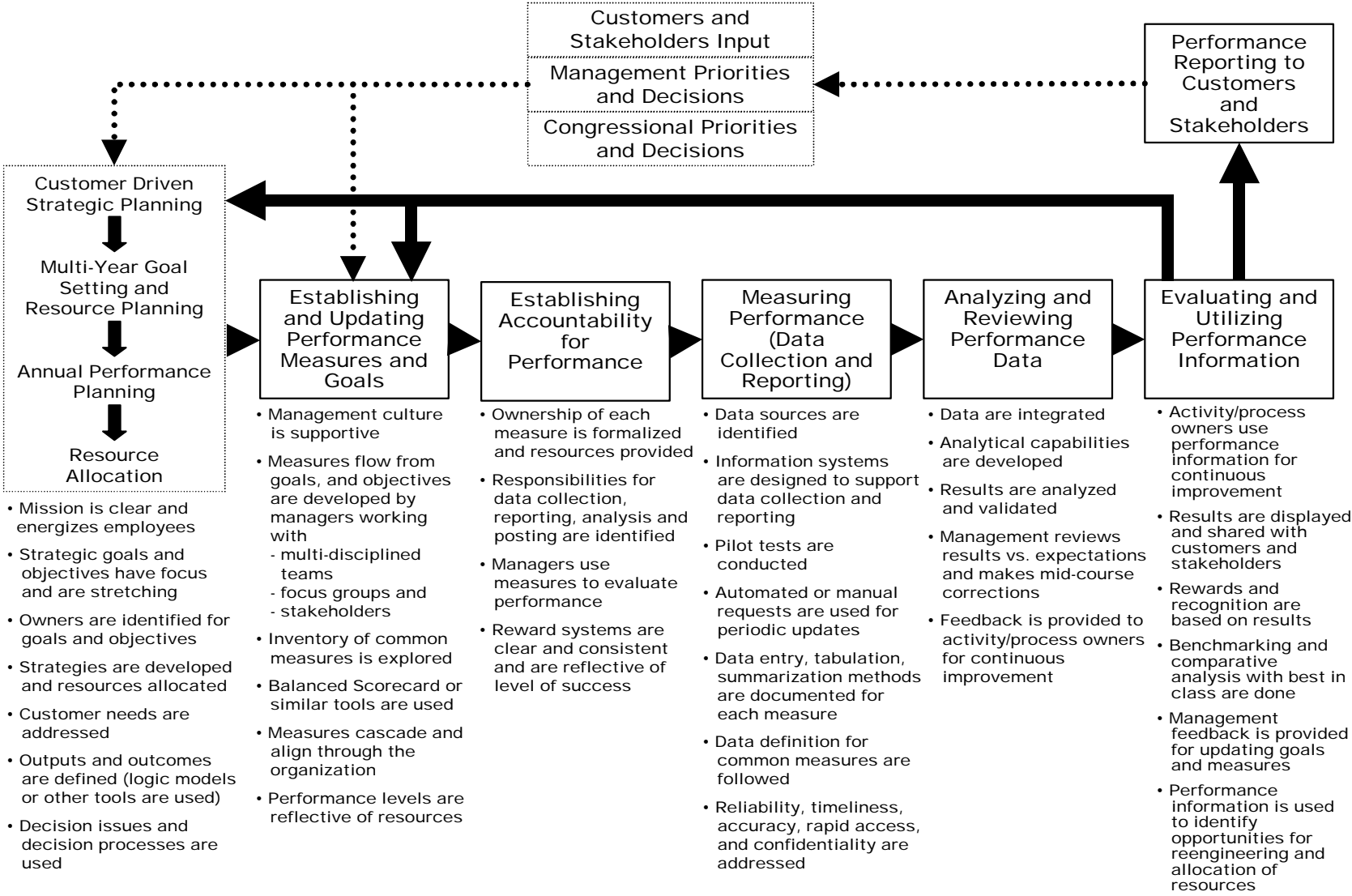
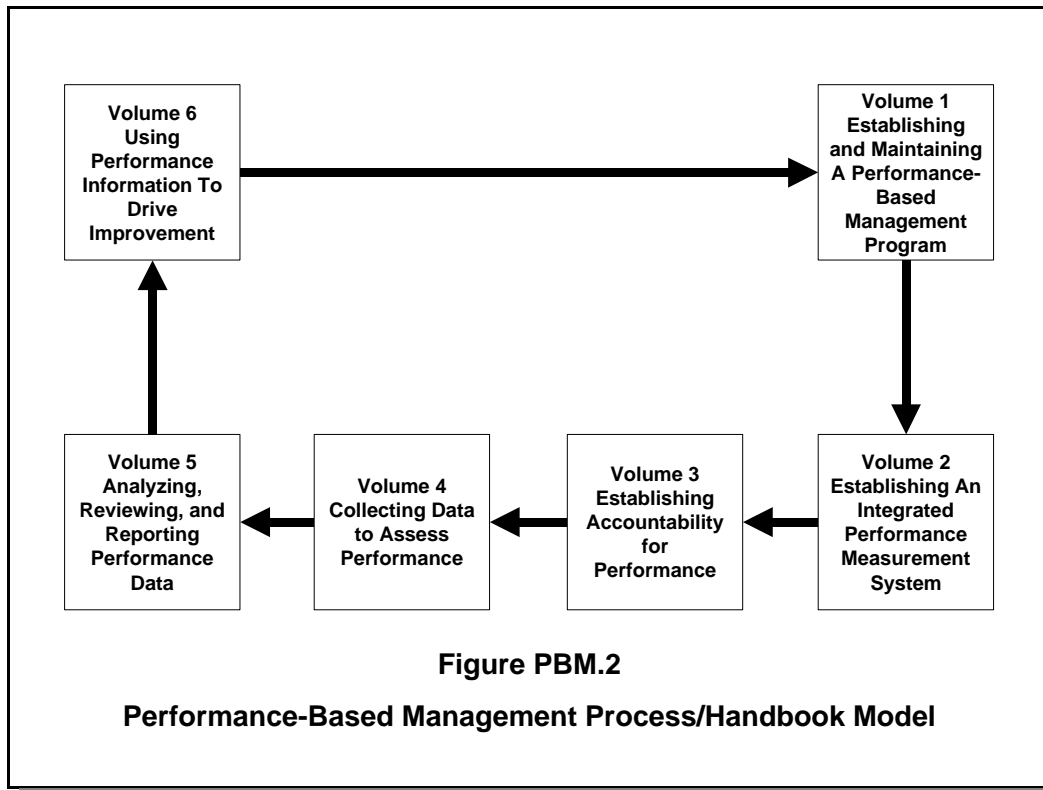


Figure PBM.1

NPR Performance Measurement Process Model

The Performance-Based Management Handbook

The PBM SIG adapted the NPR Performance Measurement Process Model into a performance-based management process model and used this model to structure *The Performance-Based Management Handbook*. The PBM SIG Performance-Based Management Process/Handbook Model is shown in Figure PBM.2 below. Topics covered by each volume are listed after the figure.



Volume 1: Establishing and Maintaining a Performance-Based Management Program

- An Introduction to Performance-Based Management
- Step 1: Define Organizational Mission and Strategic Performance Objectives
- Step 2: Establish an Integrated Performance Measurement System
- Step 3: Establish Accountability for Performance
- Step 4: Establish a System/Process for Collecting Data to Assess Performance
- Step 5: Establish a System/Process for Analyzing, Reviewing, and Reporting Performance Data
- Step 6: Establish a System/Process for Using Performance Information to Drive Improvement
- Maintaining a Performance-Based Management Program

Volume 2: Establishing an Integrated Performance Measurement System

- Understanding Performance Measurement
- Establishing an Integrated Performance Measurement System
- Choosing a Performance Measurement Framework

- Developing Performance Measures—Getting Organized
- Developing Performance Measures—Sample Approaches
- Maintaining an Integrated Performance Measurement System

Volume 3: Establishing Accountability for Performance

- The Concept of Accountability
- Establishing Accountability for Performance
- Accountability Tools

Volume 4: Collecting Data to Assess Performance

- Determining Data Needs
- Components of a Data Collection Plan
- Data Collection Considerations
- Data Collection Methods
- Suggestions for Measuring R&D Activities

Volume 5: Analyzing, Reviewing, and Reporting Performance Data

- Introduction to Data Analysis
- Training Your Organization in Analysis Skills
- Generating Useful Information - Step 1: Question Review
- Generating Useful Information - Step 2: Data Collection and Organization
- Generating Useful Information - Step 3: Data Analysis
- Generating Useful Information - Step 4: Data Presentation

Volume 6: Using Performance Information to Drive Improvement

- Using Performance Information to Drive Improvement
- Benchmarking
- Reengineering
- Continuous Improvement
- Process Improvement

About This Volume

This volume was edited by: Will Artley, Oak Ridge Institute of Science and Education, and Randy LaBarge, Pacific Northwest National Laboratory. Editorial assistance was provided by Phyllis Baker, University of California; Cynthia Eubanks, Bechtel Jacobs Company; Buck Koonce, University of California; and Suzanne Stroh, University of California.

Volume 4 Overview

Facts are stubborn things; and whatever may be our wishes, our inclinations, or the dictates of our passions, they cannot alter the state of facts and evidence.

John Quincy Adams

Webster defines the term datum as "An assumed, given, measured, or otherwise determined fact or proposition used to draw a conclusion or make a decision." Indeed Mr. Adams was correct in asserting that facts or data are stubborn things, for if data were subject to our wishes, inclinations or passions, then the validity of the conclusions or decisions drawn from the data could be questioned by anybody with a differing view of the world. Congress, and through it the American public, expect a certain level of performance for their tax dollars. It is only through hard, factual data and sound analysis at the corporate level that public confidence can be achieved and maintained.

This volume on data collection is a natural progression from the previous volumes where performance-based management systems are developed, performance measures are defined, the organizations responsible for performance measurement are identified, and the individuals responsible for performance are identified and held accountable. Volume 4 builds on these foundations and provides the reader with a systematic approach to the process involved in data collection, from identification of data needs, to the selection of data collection methodologies, to measuring performance of activities. Specifically, this volume addresses the following:

- Determination of data requirements
- Components of a data collection plan
- Considerations in choosing data collection methods
- Data collection methods and their strengths and weaknesses
- Suggestions for measuring research and development (R&D) activities

As with any activity within an organization, the need for management support is critical to the success of the data collection activity. Likewise, senior management support of performance measurement initiatives is paramount to the success of the program. Management must be convinced that the benefits gained from measuring performance outweigh the costs of establishing the program, defining the measures, and collecting, analyzing, and reporting the data.

Most often, the costs associated with data collection, storage and retrieval are the most expensive aspects of a performance measurement program. Additional staff and resources must be directed to the performance measurement program. Thus, avenues must be explored fully to minimize the fiscal and human resources needed to meet data collection needs. It is incumbent upon those involved in performance measurement efforts to effectively communicate to senior management the importance of the activity and to clearly articulate the benefits derived. It should be noted that data collection is not a discrete, one time function. Management needs to be appraised continually of derived benefits and their involvement continually must be solicited.

Section I: Determining Data Requirements

Prior to the actual collection of data, the requirements prompting the need to collect data have to be fully understood and the characteristics for a good performance measurement system must be identified, for it is this system that will provide the collected data. Several performance measurement frameworks were presented in Volume 2 of this handbook to stimulate thought about what should be measured and how the process should be organized. Please refer to that volume for more information on the subject.

Determining Data Requirements

Developers of a data collection plan need to concentrate first on identifying the data that decision-makers use to operate the organization, and then they need to conduct a survey of what data is already available. If they start by defining what is available rather than what is needed, the emphasis is put in the wrong place, and, inevitably, a large data collection exercise is underway before the purpose of the data collection effort is clearly defined. Taking the time to define the purpose of the data collection and reporting effort up-front can yield benefits down the road as well. For instance, line management could have fewer objections to narrowly defined data collection requirements that impact their daily operations. Thus, the first step in data collection is to list the performance objectives and accompanying performance measures for which data is needed.

It is important to look beyond your own organization for performance measures and data requirements. The nesting of strategic and performance plans is vital to assuring that the data collected is consistent across the organization and is provided in a timely manner. The organization's strategic and performance plans are the primary management documents from which performance measures are developed. Within an organization such as the Department of Energy (DOE), there are several subordinate organizations, each with its own strategic and/or performance plans. It is incumbent upon each subordinate organization to develop its plans such that they are in alignment with the organizational plans above them.

An equally important data linkage within DOE is the annual agreement the Secretary of Energy reaches with the President on plans and expectations for the upcoming fiscal year in key priority areas, as well as in business practices. These expectations are communicated in the form of commitments made in each of the Department's five business areas. Usually, these commitments involve a very small number of organizations—possibly one—reporting information to DOE Headquarters that represents the Departmental performance for that commitment. These data are typically outputs of a process rather than the outcome measures associated with the strategic plan.

"Best Practices" in Data Collection

The National Partnership for Reinventing Government [formerly the National Performance Review (NPR)] used the experiences of several U.S. agencies, local governments, and the Canadian government to identify the "best practices" in performance measurement (NPR 1997). The study team documented four principles that were used by all of their partners for gathering the highest quality data. Underlying these "best practices" is the understanding that the people providing the data are motivated to provide good data, that they see to it that the data are used and useful, and that the benefits of collecting the data outweigh the costs. Data collection must be:

- *Focused* on the organization's assessment and improvement needs
- *Flexible* to take advantage of any data source or method that is feasible and cost-efficient
- *Simple and aligned* with the organization's needs to provide clear, relevant information
- *Consistent* to allow comparisons and easy transition from one data set to the next

Data Quality

To ensure that data collection and analysis is applicable to the decisions that will be made from the analysis, it is important to verify and validate the data. Verification ensures that the data collected represents the data desired—that a set of data meets a specified set of criteria. Validation is the comparison of the data with the requirements' specification, i.e., does the data collected accurately represent the condition it is supposed to describe? Together, verification and validation check for any oversights or deviations from requirements, and identifies them.

The U.S. General Accounting Office (GAO) looked across all federal agency performance plans and compiled the following advice on ways to improve the quality of performance information as well as meet GPRA requirements for the verification and validation of agency performance information.

1. Foster Organizational Commitment and Capacity for Data Quality

- Communicate support for quality data
- Review organizational capacities and procedures for data collection and use
- Facilitate agency-wide coordination and cooperation
- Assign clear responsibilities for various aspects of the data
- Adopt mechanisms that encourage objectivity in collecting and managing data
- Provide responsible staff with training and guidance for needed skills and knowledge

2. Assess the Quality of Existing Data

- Build data quality assessment into normal work processes, including ongoing reviews or inspections
- Use software checks and edits of data on computer systems and review their implementation
- Use feedback from data users and other stakeholders
- Compare with other sources of similar data or program evaluators
- Obtain verification from independent parties, including the office of the inspector general

3. Respond to Data Limitations

- Report data limitations and their implications for assessing performance
- Adjust or supplement problematic data
- Use multiple data sources, with offsetting strengths and limitations
- Improve the measure by using another source or new methods of measurement

4. Build Quality into the Development of Performance Data

- Use prior research or analysis to identify data elements that adequately represent the performance to be measured
- Gain agreement among internal and external stakeholders about a set of measures that are valid for their intended uses
- Plan, document, and implement the details of the data collection and reporting systems
- Provide training and quality control supervision for all staff who collect and enter data, especially at local levels
- Provide feedback to data collectors on types of errors found by data checks
- Use analytic methods and transformations appropriate for the data type and measure being reported

Section II: Components of a Data Collection Plan

The purpose of data collection is to provide a basis for analysis, in other words, to turn data into information that is used by, and useful to, decision-makers. However, before data can be collected, a data collection plan needs to be developed.

Components of Data Collection Planning

A data collection plan is essential to ensuring that collected data supports the overall objectives of the performance measurement program and provides details to support decision-making by the users of the information. The integrity of the performance measurement program is dependent upon the quality of the collected data. Development of a data collection plan should include the considerations that follow.

Statement of Informational Requirements

The data collection plan should clearly define the informational needs of the performance measurement program. Specifically, the plan should discuss what information will be required as inputs for each of the measures. Clear relationships between the informational requirements and the objectives being measured should be established.

Statement of Information Sources

For each identified measure or data element, the data collection plan should specify a specific data source. The data collection plan should clearly define the relationship between the performance measure, the performance objective, the performance evaluation method, and the data source(s). A simple form to capture this logic is shown in Figure 4.1 below.

This identification should include comments regarding the availability, cost, reliability, and quality of each potential data source. If some data appears difficult to obtain, the data collection plan should outline alternative sources for that data, and specify what impacts the use of the alternative source might have upon the evaluation. (Section IV discusses in more detail possible methods for data collection and what to consider when choosing a method.)

For each performance objective, the data collection plan should include the following elements:

- A collection strategy: method, source, sample (if applicable)
- Development schedule for new data collection efforts
- Administration schedule

A collection strategy may apply to more than one objective. Additionally, it is "best practice" to gather information from multiple sources/strategies for each objective.

Performance Measure	Performance Objective	Evaluation Method	Data Source

Figure 4.1
Identifying Information Sources

Data Collection Process

In any data gathering activity, it is important to establish the reason and process for the data collection, the time period(s) for which the data will be collected, and the form(s) of analysis that will be used with the collected data. Establishing this process will assist in determining what type of data to collect, how to collect and store the data, and what type of analysis may be used by decision-makers with the data. The information used from the data analysis should report what the data reveals about a given inquiry. The reporting of this information should be factual, supported by the data with the proper use of analytical tools appropriate to the type of data collected.

Data Collection and Reporting Frequency

The type of data being collected and the needs of the decision-makers for the timing of the information dictate collection and reporting frequency. The collection and reporting frequency do not have to be the same. Some users may like to see monthly data only once a year, for example, while other users may want trend information such as organizational financial data more frequently. To be useful for management decisions, however, information timing should be matched to the need for decisions and to the cost of collection and processing. Regardless of the timing considerations, ensuring data quality is crucial to delivering useful information to management. Additionally, management needs and relationship to program mission should drive data collection and information development, not data collection convenience.

Data Collection Costs

The more sophisticated the data collection and reporting system, the more expensive it will be to implement. Improved timeliness, depth of understanding, breadth of coverage, and user ease-of-use come at a price. Tight control of the reporting system development process is required, and close attention must be paid to collecting only what is needed, not all that is available. System developers need to resist the temptation to collect the easiest available data at the expense of data that may be harder to collect but is more valuable to decision-makers.

A needs analysis will help define the boundaries of the data collection and reporting system. Design of the information system must be driven from top-down requirements and considerations to ensure that the benefits of collecting data and the subsequent processing into meaningful information is completed in a cost-effective manner.

Data Protection

Protecting sensitive data is of paramount importance to every organization and should be the first consideration when designing a performance information system. Each organization will have information that is for internal management use only and not intended for dissemination outside of the organization. Many organizations will build a two-tiered information system—one tier designed to provide information destined for public consumption and a second tier designed to report internal operating, financial, and other performance information.

The Trial Run

It is important to pilot new data collections methods by doing a trial run. The trial run does not have to involve the entire program, but must:

- Include all untested aspects of the outcome measurement system.
- Involve a representative group of participants.
- Last long enough to span all the key data collection points.

Some options for using a subset of participants in a trial run include:

- For multi-site programs, use only some sites.
- If staff are organized into units, use only some units.
- If participants go through the program in groups, use only some groups.

Ensuring Data Quality

Data quality may be defined as “the extent to which information remains reliable and consistent across the organization.” The issue of data quality often comes down to answering the question, “Is the data collected of use for the intended purpose?” In this case, any data that is incorrect can potentially impact the quality of the data and any decision made from that data. Thus, poor quality data can have a negative impact on an organization since many management decisions are based on quantitative analysis. Incomplete, inaccurate, or missing data increases the risk of incorrect reporting of findings and trend analysis. Moreover, having to scrub data to fix a particular problem is expensive and time consuming. Therefore, it pays to initiate a data quality system because improvements in data quality will lead to more informed management, strategic planning, and decision-making.

Assuring data quality must be a planned activity. It doesn’t just happen on its own. Data must be named and defined consistently across sites, facilities, and departments to support strategic processes and cross-functional analyses. The development and use of a data dictionary will help to ensure proper data standardization and consistent use across the organization.

Measuring Data Quality

The level of data quality is measured against four criteria:

- Accuracy (rate of error)
- Completeness (reporting of data as required)
- Consistency (application of criteria yields similar results)
- Currency (age of data relative to time of collection and collection frequency)

Data Quality Systems

To ensure data quality, an organization needs to establish a data quality system, complete with a data quality policy to standardize the data, to provide definitions, and to establish data naming conventions. This data quality system will measure, analyze, and improve the quality of the data.

The four steps to implementing a data quality system are:

1. Establish a data quality position
2. Formulate a data quality policy
3. Determine objectives
4. Obtain management and employee commitment

When developing the data quality system, make sure you:

- Develop a common architecture with consistent data definitions and formats
- Standardize data elements and data entities
- Determine valid use(s) of the data
- Establish a database map

- Obtain input from credible sources to define and validate data definitions
- Resolve conflicting information
- Create data edit criteria (i.e., acceptance criteria for out of control limits data, accuracy, completeness, currency)
- Create entries to populate look-up tables used to edit data
- Maintain a program-area dictionary/glossary
- Use a single integrated data system

Data Collection Systems and Software

Traditional data collection and reporting involves manually collecting performance data on forms, looking for trends in the data, and summarizing the results in printed management reports. However, this traditional method of reporting is rapidly being replaced by automated software systems that rely on a computer's processing power to help collect, analyze, process, and communicate management information in real-time and in both visual and printed form.

A new class of software—performance information systems—is appearing. Software systems in this category take advantage of more sophisticated features available from client/server-based information systems, such as querying and reporting, data mining and data warehousing, and multidimensional analysis. These more sophisticated systems typically incorporate additional information technology components as well, such as executive information systems, graphical user interfaces, advanced reporting features, “drill-down” reporting to gain access to the underlying data, and linkages to databases at multiple locations. These systems are most applicable in situations where sophisticated trend analysis or forecasting (e.g., leading indicators) are required, or where enterprise-wide information systems are being developed to manage and integrate not just performance information, but other corporate information as well.

The sophistication of the collection and reporting system should be matched to the mission needs of the organization. Clearly, the information technology investment must return benefits to the organization that exceeds the investment costs. These benefits should be mission-related, and they will typically accrue in terms of improved information accuracy, security, accessibility, timeliness, and cost-effectiveness.

Information System Considerations

The ideal time to incorporate information technology into the performance management system is as part of a reengineering effort to improve process efficiency. Indeed, the new generation of work-flow software tools allows for a sophisticated analysis of alternatives to process flow. However, information technology considerations are much broader than work-flow, and the list of items to incorporate into the data collection and reporting system includes:

- Access to legacy mainframe and minicomputer data and systems
- Access to different operating systems and locations
- Internet and intranet access and control
- Data warehouse and executive information system considerations

Section III: Data Collection Considerations

Before selecting a data collection method (as described in the next section), many aspects of the data collection process must be considered. The most notable aspects to consider are:

- "Data Breakouts"
- Sampling Techniques
- Bias
- Cost
- Data Reliability
- Data Validity
- Demographic Diversity/Diverse Populations
- Geographic Scope/Diversity
- Level of Accuracy
- Level of Detail
- Response Rate
- Speed
- Stakeholder Input

These considerations are covered in this section.

"Data Breakouts"

In his book, *Performance Measurement: Getting Results* (1999), Harry Haltriy suggests that planning for breakout groups in data collection and analysis can provide very useful information. Breakouts or disaggregations of data sets can reveal highly useful findings on performance that are hidden by aggregation. Two functions of breakouts are particularly important:

1. Breakouts can distinguish differences in performance among relevant groups. Identifying such differences is the first step toward asking:
 - Why is high performance occurring in some places and not others? Answers to this question can lead to transferring successful practices to less successful work.
 - Why is low performance occurring in some places and not in others? Answers to this question can ensure that appropriate improvement options are identified and addressed.
2. Breakouts can help identify inequities among customer groups—situations in which some groups had significantly better outcomes than others.

The basic types of breakouts for outcome data are shown in Table 4.2 below.

Basic Types of Breakouts for Outcome Data	
Organizational unit or project	Difficulty of workload
Workload or customer characteristics	Type and amount of services provided
Geographical location	Reason for outcome or rating

Table 4.2
Basic Types of Breakouts for Outcome Data

Sampling Techniques

Sampling refers to the measurement or surveying of only a portion of the whole population of interest. Sampling offers a useful way to maximize the benefits of data collection when collection of information from the full population is not feasible. Sampling often attempts to measure a representative group of events and to generalize those results for the population at-large. There are other instances where the ability to generalize is less important than simply sampling some portion of prevailing views.

If the data collection plan includes collecting data from only a portion of the scope of the performance area, then the data collection plan must specify the method of sampling. (There are three primary sampling methods: random sampling, stratified sampling, and systematic sampling.) It must justify the sampling method, provide details of the statistical impact of the chosen sample methodology, and discuss any limitations the sample might place upon extrapolating the results. It also should outline the assumptions used in constructing the samples (especially assumptions about subject variances). Further, the data collection plan should provide information about the amount of data necessary to reach the desired sample size. When including the use of sampling in the data collection plan, it is important to understand any bias that may exist with the sampled data. Sample bias refers to the likelihood that a measurement by a sample of the population does not accurately reflect the measurement of the whole population. Random sampling and the collection of larger samples are effective ways to decrease sample bias.

Advantages and Disadvantages of Sampling

The advantages and disadvantages of sampling are shown in Table 4.3 below

Advantages	Disadvantages
<ul style="list-style-type: none"> • Sampling may significantly reduce the cost of data collection and analysis. 	<ul style="list-style-type: none"> • Sampling does not collect all cases and, although statistically unlikely, may result in misidentification of general impact or opinion.
<ul style="list-style-type: none"> • Sampling may decrease the time needed to complete an evaluation. 	<ul style="list-style-type: none"> • In the case of qualitative data, missing observations may lead to unique perspectives lost (and unaccounted for through sampling techniques).
<ul style="list-style-type: none"> • Sampling is most useful where it may be impossible or inefficient to measure the whole population. 	

Table 4.3

Advantages and Disadvantages of Sampling

Bias

Each method of evaluation contains potential sources of bias in one form or another. The term *bias* refers to the likelihood that data collected may reflect only a portion of the spectrum of relevant opinion. Bias may occur in quantitative and qualitative data collection. It often occurs as the result of the collection of an incomplete or inaccurately weighted sample of data.

There are four types of bias:

1. *Response Bias* refers to the bias resulting from an incomplete response to a data collection instrument. For example, those individuals responding may share some common characteristics that those not responding do not share—often appearing as either a positive or negative opinion of the program. Bias in coverage refers to the extent to which subgroups of a target population participate differentially in a program.

2. *Self-lifting Bias* refers to a situation where individual responses are intended to make the respondent look better to the questioner.
3. *Social Desirability Bias* refers to a situation where responses are tailored to mirror the popular social sentiments.
4. *Ego-Threat Bias* refers to a situation where responses (or non-responses) attempt to protect the respondent from divulging sensitive personal information.

Therefore, it is imperative to closely examine data to get a feel for the information and spot potential sources of bias. When interpreting data it is critical to ask how might this data be biased and how might these biases be removed? Often, these answers lie in additional questions that can be asked during data collection to clarify answers that may contain elements of bias.

Cost

As with many other management decisions, cost plays a central role in the evaluation planning process. Data collection techniques vary widely with respect to cost. Program managers should balance the needs of the evaluation with the financial resources available for the evaluation. Some low-cost data collection techniques limit the type of information that can be collected or the quality (i.e., the validity, reliability, and accuracy) of the data. Among the least-expensive data collection methods are program and agency records, file reviews, content analyses, focus groups, mail questionnaires, and telephone surveys.

Data Reliability

The measurement process may be perceived as part of a larger series of measurements. Data collection methods vary in their ability to collect data that points toward similar conclusions. Data is reliable if another analyst employing the same methodology could collect a comparable set of data. If the current evaluation process is part of a pattern of prior evaluations, you should consider the implications to comparability of data should you elect to employ a different data collection strategy, method, or tool. Further, if collected data is meant to form a database of information for future comparison, you should select a data collection method that allows future evaluations to collect comparable data. If consistency over time is important, statistics and well-run surveys generally offer more reliable data sources. Data collected through the other methods will generally not be as generalizable to the whole population and may not compare to earlier or later similar studies.

Data Validity

Validity refers to the accuracy of a measure. A measurement is valid when it measures what it is supposed to measure and performs the functions that it claims to perform. Given the imperfect nature of measurement, validity is a matter of degree—the degree to which collected data accurately measures the factor or parameter the evaluator intended to measure. Since validity is a matter of degree, it is incorrect to say that a measurement is either valid or invalid. All measurements have some degree of validity. It's just that some are more valid than others. The important thing to note is that valid measurements are the ones that take into account all relevant factors, given the whole context of the measurement, and weight them appropriately.

Measurement can be reliable without being valid, and it can be valid without being credible. But if it's valid, it has to be reliable—if the thermometer is valid, it must say 100EC *whenever* placed in pure boiling water at 1 atmosphere pressure and, hence, must agree with itself. That is, it must be reliable.

Demographic Diversity/Diverse Populations

A diverse population refers to the target population of a sampling program whose target audience is non-homogeneous in one or more factors. Program stakeholders may differ with respect to income, size (specifically in reference to organizations), and ethnicity.

Geographic Scope/Diversity

Some sampling programs cover such expansive geographic regions that the types of data readily available will be limited. By the same token, some data collection techniques become prohibitively expensive (or very slow) to execute for programs covering wide geographic areas. Managers may find site inspections and personal interviews difficult to implement for geographically widely-dispersed programs. On the other hand, focus groups, official statistics, file reviews, and questionnaires (mail or telephone) may better serve such widely dispersed programs.

Level of Accuracy

Data collection may yield a body of information that varies in its degree of accuracy. Some collection techniques may produce results that are similar to another overall but differ in degree of focus. For example, if an evaluation attempted to ascertain the amount of energy saved, one data collection method might measure savings for a multibuilding facility, while another might measure energy savings per boiler in a single building or in each of the buildings in the facility. Each body of data will yield valuable information but will differ in its degree of accuracy. Data collected by official statistics is usually more coarse (less accurate) than data collected by individual metering, producing a finer (more accurate) data set.

Level of Detail

Data collection accumulates verifiable facts for use in the evaluation process. Different methods are capable of collecting information of differing levels of detail. Program managers should consider carefully what level of detailed information is required and appropriate and how best to accumulate it. While official statistics and mail questionnaires may provide valuable information for an evaluation, they do not offer much detailed information on the program's operations or on its personal impact. On the other hand, case studies, focus groups, and personal interviews may gather more in-depth information.

Response Rate

Response rate usually applies only to surveys and refers to the ratio of responses received to the number of questionnaires (or other data collection technique) solicited. Response rate is important in determining the bias of the data received from the instrument. If a technique has a low response rate, the evaluator may question who responded. If a survey is being conducted, it may be that one disaffected group is more likely to respond (say those who had a bad experience with the program) and, hence, the data collected may be biased toward a negative review of the program.

There are several ways to improve response rates. For example mail-in surveys traditionally incur low response rates. However, surveys administered in person or over the phone have proven to have significantly higher response rates.

Speed

Several data collection techniques offer rapidly available data. For other techniques, however, the data collection process requires more time. You should carefully consider how quickly the information is needed. Quite often, the data that can be collected more quickly may sacrifice one or more of the other necessary characteristics (i.e. level of detail, or level of accuracy).

Stakeholder Input

Stakeholders are those individuals affected or impacted by (they have a stake in) the outcome of a particular action or series of actions. Stakeholder input typically is collected through formal surveys, informal discussions, or structured forums such as focus groups. Various types of information may be collected through this input, including program structure and operation; potential measures; sources and types of data already collected; and opportunities for program improvement.

Data Collection Types and Factors Matrix

Table 4.4 on the following page provides a convenient matrix for comparing the various data collection techniques with the factors that should be considered when selecting a specific data collection technique. The table and the descriptions of the advantages and disadvantages of the various data collection methods (provided in the next section) summarize the opinions of several evaluation experts. As such, there is room for disagreement, particularly since accuracy, reliability, and other factors will vary depending on the application of the method.

	Cost	Response Rate	Geo-graphic Diversity (Utility)	Bias	Speed	Level of Detail	Stakeholder Input	Reliability	Level of Accuracy	Demographic Diversity (Utility)
Case Studies	Moderate	Moderate	Low	High	Moderate	High	High	Low	High	High
Content Review	Low	N/A	High	High	Moderate	High	High	Potentially Low	Potentially High	N/A
Cost-Benefit/Cost Effectiveness Studies	High	N/A	High	High	Low	Moderate	Low	Moderate*	High	High
File Review	Low	N/A	High	High	High	Moderate	Low	Moderate	Low	N/A
Focus Groups	Low	High	High	Moderate	High	High	High	Low	Moderate	Moderate
Enhanced Peer Review/Expert Panel	Low	High	High	High	High	High	Low	Low	Potentially High	High
Statistics	Low	N/A	High	High	Moderate	Low	Low	High	Moderate-Low	High
Survey Telephone Interview	Low	High	High	Moderate	High	Moderate	High	Low-Moderate	Low	High
Survey Personal Interview	High	High	Low	Moderate	Low	High	High	Low	Moderate	High
Survey Mail	Low	Moderate	High	Moderate	Moderate	Low	Moderate-High	Moderate	Low	Low

Table 4.4
Data Collection Types and Factors Matrix

*This issue is the source of much dispute in most settings. In any case, the quality of cost-benefit analyses rests upon data quality and assumptions made. Two researchers may arrive at differing conclusions using the same data but using different assumptions.

Section IV: Data Collection Methods

Depending on the type of information available and the data and analysis needs, one or more data collection method may be appropriate to support the compilation of performance measurement information, or to support the evaluation of performance when traditional data collection techniques are not feasible. This list is not exhaustive, nor is the treatment given to each technique comprehensive. For additional information on each of the techniques listed, please see the reference section at the end of this volume. Also, please refer to Table 4.4 (on page 14) to view a convenient matrix for comparing the various data collection techniques with the factors that should be considered when selecting a specific data collection technique.

Each data collection method embodies an important set of choices. Just as the different evaluation methods provide expanded options for answering various questions, different collection methods may yield different information and provide differing insights. When selecting a method of data collection, the program manager should keep in mind the many trade-offs between the different types of data collection methods. Consider this list from which the manager can choose:

- Program and Agency Records
- Site Inspections/Observations
- Use of Special Technical Equipment
- Surveys and Interviews
- Purchasing Statistics
- Peer Review/Expert Panel Evaluation
- Quantitative Metrics for Research and Development
- Economic Methods
- Cost-Benefit/Cost Effectiveness Studies
- Case Studies
- Content Review
- Focus Groups

Each method varies in its performance with respect to bias, cost, response rate, speed, level of detail, validity, reliability, and usefulness with demographically diverse populations. When considering the data collection methods, the program manager should assess the limitations of each type of data and the requirements of the evaluation with respect to each of these factors, in order to select the optimal method of data collection. (Note: These factors are discussed in the previous section.)

Program and Agency Records

For administrative purposes, most agencies and programs routinely record data on customers and/or transactions. In his book on performance measurement (Haltry 1999), Harry Haltry suggests that this data collection procedure has been by far the most widely used for producing performance data. In addition to being a source of outcome information, agency records are also the main data source on the amounts of input (both dollars and employee time) and output produced by the program. Records can also be a source of demographic characteristics of customers and other characteristics of the workload for use in providing breakouts of outcome indicators.

Advantages and Disadvantages of Program and Agency Records

The advantages and disadvantages of using program and agency records as a form of data collection are shown in Table 4.5 on the following page.

Advantages	Disadvantages
<ul style="list-style-type: none"> The data are readily available at low cost. 	<ul style="list-style-type: none"> Agency records seldom contain enough service quality and outcome data to create an adequate set of performance indicators.
<ul style="list-style-type: none"> The procedures for transforming the data into indicators are familiar to most program personnel. 	<ul style="list-style-type: none"> Modifications to existing record collection processes are often needed to generate useful performance indicators. Obtaining data from the records of other programs or agencies, which is sometimes needed to calculate an indicator, can be administratively difficult and can raise issues of confidentiality.

Table 4.5

Advantages and Disadvantages of Agency Records

Site Inspections/Observations

Trained observers are used to rate outcome conditions that can be perceived by the eyes or other physical senses of an observer. For trained observer ratings to be a suitable performance measurement procedure, the outcome needs to be measurable by physical observation and to be reliable on a scale that identifies several variations of the condition to be measured, e.g., never, seldom, sometimes, often, always. The goal is to ensure that different observers at different times give the same or very similar ratings to similar conditions. The following procedures will ensure a high degree of reliability of observations:

- Use systematic rating scales that provide well-defined yardsticks against which the observers can assess observed conditions.
- Provide adequate training and supervision of the observers and the process.
- Periodically check the quality of the ratings.

Types Of Rating Systems

Table 4.6 below shows the types of rating systems that can be used by the trained observer.

Type of Rating Systems	Description
Written Descriptions	This type, the simplest and most familiar type of rating system, depends entirely on specific written descriptions of each grade used in the rating scale.
Photographic Rating Scales	Photographic scales can be more precise than written scales in providing clear definitions of each rating grade. Generic photos are used to represent grades on a rating scale. Observers are given (trained in the use of) a set of photos, with several representing each grade on the rating scale.
Other Visual Scales	Visual rating scales can also use drawings or sketches or videos that represent each grade on a rating scale.

Table 4.6

Types of Rating Systems

Advantages and Disadvantages of Trained Observer Ratings

The advantages and disadvantages of trained observer ratings are shown in Table 4.7 below.

Advantages	Disadvantages
<ul style="list-style-type: none"> Trained observer ratings can provide reliable, reasonably accurate ratings of conditions that otherwise are difficult to measure. 	<ul style="list-style-type: none"> They are labor-intensive procedures requiring significant personnel time, including time for training observers. For performance measurement purposes, using an agency's own caseworkers should be avoided, because their potential self-interest in the ratings reduces the perceived, if not the actual, credibility of the results.
<ul style="list-style-type: none"> Periodic ratings can be used to help allocate program resources throughout the year, e.g., street cleanliness. 	<ul style="list-style-type: none"> The need for periodic checks of observers to ensure that they are adhering to the procedures adds costs.
<ul style="list-style-type: none"> The data can be presented in an easy-to-understand format, which is important in reaching officials and the public. 	<ul style="list-style-type: none"> Program personnel may be uncomfortable with observer ratings, as evidenced by their relatively infrequent use.

Table 4.7

Advantages and Disadvantages of Trained Observer Ratings

How to Implement a Trained Observer Process

Implementing a trained observer process requires the following steps:

1. Decide what conditions should be rated.
2. Develop a rating scale for each condition. If possible, adapt an existing scale. Use photographs and written guidelines as appropriate.
3. Determine which facilities or areas should be rated, when, and how frequently. Ratings can be applied to all or selected facilities or areas. If resources are only available to rate a subset of locations, choose them by using random sampling so that the locations chosen will be representative.
4. Select and train observers, who might be program personnel, college or graduate students, or volunteers. Technical ratings, such as safety hazards, will require persons with the requisite professional training.
5. Test the scale and observers on a small number of sites in the facility to make sure that reasonably trained observers give consistent ratings.
6. Establish procedures for supervising the observers and for recording, transcribing, and processing the collected data.
7. Conduct the ratings regularly.
8. Develop and disseminate reports on the finding from each set of ratings for the current period and changes from previous periods. The reports should show the number and percent of locations that fell into particular rating categories. Do not report only the average scores, which can hide very important distributional information.

Use of Special Technical Equipment

Haltry (1999) also suggests using special technical equipment to collect data for outcome indicators that require scientific measurement, such as noise levels, air pollution levels, water pollution levels, and road conditions (using road meters). There are advantages and disadvantages to using this method, as shown in Table 4.8 below.

Advantages	Disadvantages
<ul style="list-style-type: none"> Appropriate technical equipment usually provides accurate, reliable data. 	<ul style="list-style-type: none"> The equipment can be expensive to procure, operate, and maintain.
<ul style="list-style-type: none"> It may be the only reasonable way to achieve completely credible information on important environmental outcomes such as those listed above. Programs can obtain subjective outcome data using trained observers or user surveys to assess the quality of roads, water, air, and so on, but such information is likely to lack the credibility provided by technical measurement. 	<ul style="list-style-type: none"> The information obtained must be interpreted to be useful to program personnel and outsiders. For example, air pollution measurements need to be converted into overall air quality levels understandable to the public, such as good, moderate, approaching unhealthful, and unhealthful.

Table 4.8

Advantages and Disadvantages of Technical Equipment

Surveys and Interviews

Surveys provide a method of gathering information from stakeholders and others by directly questioning them. Generally speaking, a survey is simply a planned effort to collect data through direct questioning. This questioning can produce either qualitative or quantitative data, or both. Some methods allow free-form input, while others seek categorical responses. Questionnaires used to gather data may take the form of personal interviews, telephone interviews, or mail questionnaires. Each offers advantages and disadvantages and is capable of collecting different types of information.

Benefits of Surveys and Interviews

Three benefits of surveys and interviews are:

- Surveys and interviews can be used to gather small or large amounts of data quickly. Samplers can query a large number of respondents—or a statistical sampling—in a wide geographic area.
- Samplers can obtain thoughtful and/or quick responses from respondents and can provide confidentiality on data collection with some methods.
- Surveys offer the opportunity to involve stakeholders in both the program assessment and the program improvement processes. A well planned and conducted survey presents an efficient method of collecting personal information and perceptions of individuals impacted by a program. Surveys may provide a leading indicator of what changes may become necessary in the near future.

Steps in Designing and Conducting Surveys

How a survey is designed and conducted can be critical to the success of the survey and to the collection of a body of data that is of adequate size and quality for the intended purpose. The following steps can help you optimize your survey.

1. Define the areas for evaluation and develop applicable questions.
2. Establish a survey plan. Identify the target population and designate a comparison group, if applicable.
3. Develop a sampling protocol that includes a well thought out method of data collection, sampling techniques and method of analysis.
4. Develop the questionnaire.
5. Field test the questionnaire, the individual questions, and the time it takes to administer the test.
6. Distribute the questionnaire to respondents with a stated return date. They may need prompting to get it in on time.
7. Provide a follow-up contact with nonrespondents, if the sample size is small enough to be able to track nonrespondents individually.
8. Analyze data and share the results with stakeholders.
9. Report the results.

Purchasing or Using Statistics from an Outside Source

Statistics purchased or used from an outside source can be inputs to analyses. For example, statistics from outside sources are often used in cost-benefit studies and regression analyses. Additionally, secondary statistics may offer an opportunity to assess the effectiveness of a program by comparing data for participants and nonparticipants. Types of available statistics include:

- **Government Statistics** - Data collected and published by the United States government, through any of its entities, offers a great resource for evaluation data. Official statistics generally provide data that are relatively reliable and consistently collected. For the data analysis to provide value, the agency must follow a standard collection methodology. Some data analysis techniques may require a stream of data. In these cases, government tends to collect certain statistics on a more regular basis than many private sources.
- **Privately Published/Collected Data** - Private organizations, such as trade associations and advocacy groups, collect data that may be valuable to an organization's performance measurement effort. Unfortunately, this kind of data is prone to interrupted collection, irregular methods, nonuniformity, and uncontrollable bias. The careful program evaluation professional will only use data that conforms to the researcher's needs and will specify data limitations or seek to apply multiple lines of evaluation methods when any data is in doubt.

Advantages and Disadvantages of Purchasing or Using Statistics from an Outside Source

The advantages and disadvantages of purchasing or using statistics from an outside source are shown in Table 4.10 on the following page.

Advantages	Disadvantages
<ul style="list-style-type: none"> • Many statistics can be collected with little or no assistance from the program participants. 	<ul style="list-style-type: none"> • Available statistics may not exactly measure the desired characteristics, and the proxy measures force evaluators to either create an additional measure or simply accept the measure.
<ul style="list-style-type: none"> • Collecting data from government agencies may cost little or nothing for Federal agencies. 	<ul style="list-style-type: none"> • The bias of some privately collected data could influence the evaluation in unintended ways.
<ul style="list-style-type: none"> • This type of data can often be assembled quickly, depending on data collecting agency. 	<ul style="list-style-type: none"> • Statistics alone, in many cases, offer little evidence of program performance.
<ul style="list-style-type: none"> • Official statistics provide generally reliable sources of data, with consistency over time and across geographic areas. 	<ul style="list-style-type: none"> • Since most of the measures were originally conceived for other purposes, they may not measure the phenomenon to the desired level of detail.

Table 4.10

Advantages and Disadvantages of Purchasing or Using Statistics from an Outside Source

Peer Review/Expert Panel Evaluation

A traditional approach to research and development (R&D) program evaluation has been peer review. This approach involves the reviewing of one's work by those with expertise in the field. Peer review is premised upon the assumption that a judgement about certain aspects of science, for example its quality, is an expert decision capable of being made only by those who are sufficiently knowledgeable about the cognitive development of the field, its research agenda, and the practitioners within it.

Several governmental agencies, including the National Science Foundation (NSF) and the National Institutes of Health (NIH), currently use peer review to determine funding considerations. These reviews may take the form of highly structured or unstructured processes. Authors have recently begun to press for evaluation of basic scientific research through an enhanced peer review process that broadens both the definition of peer and the evaluation issues addressed.

Elements of Successful Peer Review/Expert Panel Evaluation

Three approaches to designation of peers have been developed. Barry Bozeman (1993) posited that peers should, whenever possible, include members of the applicable "invisible college" who study the program or area to be studied. These potential investigators may be those within an organization such as NIH or NSF or those knowledgeable professionals working in the field (professors, engineers, etc.). A different approach would be to use those evaluated to evaluate their own work, although this method remains open to criticisms of its objectivity. A more recent approach is to use stakeholders to evaluate the program or work.

Characteristics of Peer Review/Expert Panel Evaluation

Here are three characteristics of peer reviews/expert panel evaluations:

1. They can produce a list of rankings, a formal report of results, and/or a metric-driven output.
2. They can review either specific projects or an organization's entire effort.
3. They may evaluate, prospectively, works in progress or a completed project.

Seven Ways to Enhance the Results of Peer Review

Bozeman (1993) proposes the following seven ways to enhance the results of peer reviews:

1. Use peer review in conjunction with other evaluation techniques.
2. Use peer review evaluation for R&D activities that are in the public domain. Peer review needs a high level of detailed information regarding the existing work. For some kinds of valuable patentable research or industrial process information, researchers may not be willing to open their research to outsiders.
3. Peers must be readily identifiable. For this kind of evaluation to work, there must exist a body of researchers sufficiently acquainted with the technical work. In small or newly emerging fields, there may only be a few with sufficient expertise to evaluate the project, and they may be actively involved in the project itself.
4. Avoid internal peers. With peer review, part of the benefit comes through the examination of projects by technically competent outsiders. If the panel does not contain outsiders, they may tend to produce biased or incomplete findings and recommendations.
5. Guard against dysfunctional dynamics. Since the value of peer reviews comes from the open discussions of the members of the peer review panel, group dynamics play an important role (group-think, bandwagon, and domination by individuals). Some peer reviews use online, anonymous discussions among the members and a separate individual evaluation (in addition to the group output).
6. If scales are used, test the validity and reliability of those scales. In metric-driven peer reviews, the metrics must accurately measure the phenomenon. Typical evaluations of this type request the evaluator to assess several dimensions with simple quantitative scales (i.e. very good = 1, good = 2, average = 3, bad = 4, very bad = 5). A good scale should record the relative merits of projects with a high level of consistency.
7. Provide a bias statement for reviewers. During the NSF review process, panelists stipulate their biases and potential conflicts of interest. Such a statement can mitigate some of the pitfalls of peer review.

Advantages and Disadvantages of Peer Review/Expert Panel Evaluation

Table 4.9 below shows the advantages and disadvantages of peer reviews/expert panel evaluations.

Advantages	Disadvantages
<ul style="list-style-type: none"> • Peer review may offer a way of evaluating very complex matters or to provide an insightful ranking of technical alternatives. 	<ul style="list-style-type: none"> • Peer review may possess several biases that may preclude its use in many cases when alternative methods exist. There are methods to correct some of the biases. • Very rarely is peer review used to evaluate the impacts of programs because of its subjectivity and potential for generating inconsistent results
<ul style="list-style-type: none"> • Peer review can evaluate projects that are not near a mature stage or projects that may produce immeasurable outputs (basic research produces knowledge but not necessarily patents). 	
<ul style="list-style-type: none"> • A peer review may quickly accumulate expert opinion or advice for use in developing an evaluation framework or measurement methodology. 	

Table 4.9

Advantages and Disadvantages of Peer Review/Expert Evaluation

Quantitative Methods for Research and Development (R&D) Activities

Quantitative methods for collecting performance data include the use of indicators of input such as funding and human resources, and indicators of output such as publications, citations, and patents, to identify research program effectiveness and productivity (Cozzens, et al, 1994, NRC 1994, Hauser 1997, Geisler 1999). Economic methods are also quantitative but will be discussed separately in this section. The validity of quantitative performance indicators is based on the argument that research must be subject to peer review before it can be funded or published and that research impact can be extrapolated from how frequently patents and publications are cited (Cozzens 1989, NSTC 1996).

Quantitative indicators of research quantity are most appropriate for assessing the research produced in organizations that encourage publication and whose main research output is knowledge in the form of publications and patents (Cozzens 1989, NRC 1994, Hauser 1997, Geisler 1999). Hauser (1997) states that organizations can effectively use market-outcome indicators of cost and net present value to evaluate applied projects, if they adjust for projects that are small scale, low risk, and short term. However, to evaluate the long-term value of technology development programs, organizations should combine market outcome metrics with more direct measures of research such as publications, patents, and citations.

Publication Counts

Publication counts provide useful information when combined with a larger, richer set of indicators and analyses. Their use alone or without sufficient information about other aspects of performance and the circumstances of the research can produce an incomplete, if not inaccurate, picture (NSTC 1996).

Patent Counts

Counts of patents, new devices, computer programs, and other inventions do not say much about whether a program is conducting world-class science at the frontier of knowledge, but some agencies may use them to gain insight about connections between their program and the agency mission (NSTC 1996).

Citation Counts

Counts of the number of times a publication is cited must be used with caution because high numbers of citations may indicate a negative evaluation, citation rates vary among fields, and citation counts typically only include journal references (NSTC 1996). Citation counts are often expanded with co-citation and cross-citation analysis to illustrate links between researchers (Cozzens 1989).

Integrated Metrics and Mixed Methods

These methods are a more comprehensive way of assessing organizational effectiveness to balance the strengths and limitations of multiple individual methods. Integrated approaches often incorporate a framework for describing an organization and a menu of metrics that include a combination of objective and subjective measures. The government of Canada (Canada 1993) describes using partial indicators that subjectively incorporate combinations of quantitative information as a method of assessing socioeconomic outcomes. These integrated metrics provide a higher level of reliability and flexibility, but they are complex and require more time and effort to implement (Cozzens, et al, 1994, Werner and Souder 1997).

Strengths of Quantitative Metrics

One of the strengths of using quantitative indicators, such as measures of research input and output, is that the data can be easy to collect (Cozzens 1989, Canada 1993, GAO 1997, COSEPUP 1999, Geisler 1999). Individual researchers and organizations typically maintain reliable information on publications, patents, employee demographics, R&D spending, and funding. Information on publications and citations can also be obtained using library and online databases. Another strength of quantitative indicators is that they can be readily analyzed. The GAO (1997) states that one of the strengths of R&D

spending data is that it reduces the innovation process to a single figure for the purposes of assessment and discussion. Narin, Olivastro, and Stevens (1994), in their article entitled "Bibliometrics/Theory, Practice, and Problems," state another benefit of quantitative metrics is that they are completely objective and they can be used to track research ideas and identify the key players in a particular discipline.

Limitations of Quantitative Metrics

The main limitation of quantitative methods is that they do not include any measure of research quality (NSTC 1996, GAO 1997, COSEPUP 1999, Geisler 1999). This limitation is significant because, although a research group may receive a high level of funding, produce a large number of publications, or be cited in many publications, there is no way of identifying if the research is of poor quality or if the citations are negative. Cozzens (1989), COSEPUP (1999), and Geisler (1999) also point out that, for bibliometric analysis, all publications are treated equally regardless of whether they present a long-term study or a short research note. Another limitation of bibliometric methods is that their results cannot be compared across fields. There is enough difference between the publication rates of different disciplines that a comparison would be improperly weighted toward the field that published more often (Cozzens 1989, Narin, et al, 1994, COSEPUP 1999).

Narin, et al, (1994) emphasize that, even with their limitations, quantitative methods can be used to objectively track the development of a research project and monitor the activities of researchers. Cozzens (1989) provides several guidelines to improve the use and effectiveness of quantitative methods. She states that indicators of publication, citation, and patent activity should be used to assess the work of research groups and not individuals, should only be used in conjunction with other methods that also indicate research quality, and should be used for evaluating the research produced in organizations that encourage publication as the main output.

Economic Methods

The value in using economic methods is to estimate how much benefit or return organizations receive from initial investments in R&D. From these estimates, organizations can make better decisions about budget and human resource allocation. The results of economic methodologies typically indicate that R&D activities produce high overall rates of return (GAO 1997). Assessment methods that are available to measure many of the economic facets of an organization's research include (Link 1993, Averch 1994, Cozzens, et al, 1994, NSTC 1996, Tassej 1996, GAO 1997, Geisler 1999):

- **Rates of Return** - Rates of return estimate the actual economic value derived from investment in R&D.
- **Production Functions** - Production functions are mathematical functions that incorporate values for technology input and output to estimate R&D impact via increased productivity.
- **Customer Surplus** - Customer surplus is an estimation of economic impact that makes use of how much a consumer is willing to pay for a product.
- **Social Rate of Return** - Social rate of return assesses the sum of the social benefits from technology changes as compared to cost of the technology investment.

Economic methods can indicate the value of applied or developmental research projects more easily than the value of fundamental research projects (Averch 1994, Hauser 1997). Both Averch (1994) and Cozzens et al (1994), however, suggest that the economic value of public and fundamental R&D also can be measured using variations on production function analysis, surplus methods, and social rate of return.

Strengths of Economic Methods

The primary strength of economic methods is that they give organizations some indication of the economic benefit or value derived from their R&D investments (Link 1993, Geisler 1999). Another strength is that economic metrics are quantitative and can be objectively calculated and easily analyzed

(Link 1993, COSEPUP 1999). Cozzens, et al, (1994) found that using economic methods at an aggregate level could help organizations estimate whether or not their fundamental R&D activities were meeting goals and making valuable overall contributions.

Limitations of Economic Methods

Economic methods are limited because they make use of mathematical functions, models, and equations that require discrete values for research input and output. It is often difficult to find accurate values to incorporate into these functions because an organization's economic data may be uncollected, unreliable, or proprietary (Averch 1994, Cozzens, et al, 1994). Another significant limitation is the complexity of factors involved in derivation of economic benefit. Using a single value to represent the many variables that contribute to the economic impact of research, including economic, social, technological, and behavioral issues, may distort the importance of each individual factor (Link 1993, Cozzens, et al, 1994, GAO 1997, Geisler 1999). Finally, economic calculations are hindered by the long time period between initial R&D investment and final realization of benefit (Tassey 1996, GAO 1997, COSEPUP 1999, Geisler 1999). By the time benefit can be determined, isolation of the inputs that led to the benefit may be difficult or impossible.

As long as the limitations in acquiring accurate data and overcoming complexity are kept in mind, economic methods can be used to help understand the underlying processes and effects of fundamental research (Cozzens, et al, 1994). One of the most important guidelines for successfully using economic methods to evaluate the value of R&D performance is to initially structure data collection methods to obtain the right information (Averch 1994). Tassey (1996) also makes several suggestions for using economic methods. In his experience evaluating R&D programs for NIST, he found that many of the limitations of economic methods could be overcome if organizations:

- Integrate measurements of short-, mid-, and long-term projects.
- Monitor the data quality, sample sizes, and metrics to ensure the results will be comparable.
- Combine the results of economic metrics with other qualitative and quantitative measures.

Cost-Benefit/Cost Effectiveness Studies

There are significant and varied data collection requirements to develop cost-benefit studies and cost effectiveness studies. Here are the differences between the two:

- *Cost-Benefit Studies* attempt to evaluate the costs of programs together with the benefits they foster. To accomplish this evaluation, these analyses convert all of the benefits and costs into dollar values. The results of a cost-benefit study attempt to determine if the costs of an action is justified by the value of the benefits that would result.
- *Cost Effectiveness Studies* are a type of cost-benefit study in which either (1) the costs of the program being evaluated are identical so that it is necessary to compare only the benefits, or (2) the benefits of the programs are identical so that only the costs need be compared, or (3) benefits are not transformed into monetary terms.

Time Frames Of Study

When conducting cost-benefit/cost effectiveness studies, it is important to consider the time frame of the study. Three Time Frames are:

1. *Life-Cycle* - This time frame, or scale, of study evaluates all benefits and costs of the program from inception through program termination. These studies attempt to analyze whether or not the program's entire costs are justified by all the possible benefits that will occur.
2. *Prospective* - This time frame, or scale, of study includes only those costs and benefits that will occur in the future. These studies do not include the costs associated with program start-up costs. These

studies attempt to review whether or not the program's future operation costs are justified by its future benefits.

3. *Retrospective* - These studies review only the costs and benefits that occurred in the past. Although this type of study has been widely performed, it does not provide justification for continuing the program.

Identifying Costs and Benefits

Direct costs and benefits are the easiest to compile. These measures typically include all program operation costs, administrative expenses, and overhead expenses on the cost side, while benefits usually include only the impact the program was designed to create. Program costs and benefits often include indirect, or tangential, impacts or expenses. Indirect costs and benefits accrue as byproducts or as spillovers from the program's operations. Following these chains of costs and benefits further out from their source requires careful analysis due to the difficulty in directly linking them to the program operation. Indirect costs may include overhead costs, tax impacts, productivity slowdown, costs to the private sector, and/or increased costs to the government. Indirect benefits may include technology spin-offs, secondary job creation, private market activity, and/or further research.

Valuing Benefits

Very often program benefits, either direct or indirect, include factors not expressed in terms of dollars. Common benefits include energy saved, emissions reduced, lives saved, time saved, increased productivity, increased jobs, and/or improved aesthetic features. These varied benefits must be converted to a dollar figure to allow comparison with the costs of the programs, if using cost benefit analysis. There are several methods, specific to each type of benefit, to convert these benefits into dollar measures. There is a wealth of literature describing benefit valuation and conversion. Common methods of conversion include: market prices, "shadow prices", hedonic price estimation, and/or contingent valuation. Often evaluators will utilize sensitivity analysis (completing several cost benefit studies with differing values for the benefits in order to understand the net change for each unit of input) to test the resilience of the evaluation findings to changes in the conversion of benefits into dollars.

Converting Costs and Benefits into Present Values

The cost-benefit study converts the streams of future costs and benefits into current costs and benefits by a "present value" calculation. Under such a system, future benefits and costs are mathematically discounted by some percentage. The selection of the percentage will significantly impact the analysis. A low rate will increase the relative importance of future benefits and review more favorably programs with high up-front costs. A high rate will minimize the value of benefits accrued in the future and exaggerate the costs for programs with high up-front costs. Results of these studies are often presented as "Net Present Value" (the value of present benefits minus the present costs), a benefit-per-cost ratio (which is the present value of benefits over the present value of costs), or the Return-on-Investment (a method that calculates the rate necessary for present costs and benefits to be equal).

Methods to Improve Cost-Benefit Studies

Two ways to improve cost-benefit studies are to:

1. Present a range of assumptions. When the results of the study hinges upon a few key assumptions, it is imperative to clearly state those used in the study. Once the data is in place, the evaluator can easily change these assumptions to present a sensitivity analysis.
2. Consider using cost effectiveness studies when dealing with sensitive benefits such as lives saved. When critics disagree with the conversions, it may be useful to evaluate the effectiveness of the program operation rather than to attempt to perform a cost-benefit study.

Advantages and Disadvantages of Cost-Benefit/Cost Effectiveness Studies

Table 4.11 below shows the advantages and disadvantages of cost-benefit/cost effectiveness studies.

Advantages	Disadvantages
<ul style="list-style-type: none"> • They identify the schedule of expected benefits and costs. 	<ul style="list-style-type: none"> • Small changes in program assumptions may cause dramatic swings in the results of a cost-benefit study.
<ul style="list-style-type: none"> • They measure the overall impact of the program upon both the intended beneficiaries and those unintentionally impacted by the program. These evaluations may provide support for programs that create a large share of their benefits through indirect means. 	<ul style="list-style-type: none"> • They do not often produce results that depict the distributional impacts (costs and benefits attributable to specific groups).
<ul style="list-style-type: none"> • They allow programs with vastly differing goals and results to be compared and insightfully ranked for spending priority during times of funding scarcity. 	<ul style="list-style-type: none"> • Cost-benefit studies assert causality for secondary impacts. It does not prove that the program actually influences these phenomena. It may be difficult to trace how the program impacts are distributed. As such, secondary benefits are often scrutinized by critics as not the result of the programs.
<ul style="list-style-type: none"> • They evaluate the program benefits and costs that will occur over time. These analyses allow valuation of future benefits to be included in consideration of program costs (start-up and operations). 	<ul style="list-style-type: none"> • They do not produce results that facilitate program review and examination, except at the macro level. They do not identify the cause of program impacts, only that they occur.
	<ul style="list-style-type: none"> • They may require a great deal of data to assess programs with varied impacts.
	<ul style="list-style-type: none"> • Cost-benefit studies may require more time than other methods of evaluation

Table 4.11

Advantages and Disadvantages of Cost-Benefit/Cost Effectiveness Studies

Case Studies

The case study is a research method that uses extensive description and analysis of a complex situation studied in its context to answer questions about the efficiency and effectiveness of current programs. Good case studies:

- Provide illustrative examples of program successes and failures
- Identify problems within current programs
- Show the interaction between program objectives and real life constraints
- Test specific theories and strategies
- Identify common program elements for further study and comparison
- Provide significant, credible, and comprehensible data
- Demonstrate justifiable and compelling evidence for recommendations that can be generalized to other programs

Types of Case Studies

The six types of case studies are:

1. *Program's Effects* examines the results of specific actions undertaken.
2. *Illustrative* uses descriptive details to familiarize specific aspects of the sample.
3. *Exploratory* shows the results of a program and gauges what needs have not been met.
4. *Critical Instance* examines a distinctive feature of the program to test or challenge key hypotheses, strategies, or problems.
5. *Cumulative* incorporates the findings from several case studies to answer new queries based on description, judgements, or cause and effect analysis.
6. *Program Implementation* acts as a normative investigation of an operation's execution.

Advantages of Case Studies

The advantages of case studies are that they:

- Furnish insight into a complex situation from alternative perspectives of analysis.
- Analyze cause and effect conditions.
- Allow a variety of aspects to be incorporated into an evaluation.
- Are adaptable in their ability to survey programs, conditions, events, regions, functions, agencies of various types, and scales.
- Use extensive descriptions and analysis to provide a comprehensive understanding.
- Provide information that is taken as a whole and kept in context.

Designing and Conducting a Case Study

Here are the steps to designing and conducting a case study:

1. Define the problem and formulate the scope and objective of the query with specific attention toward the nature and the context of the subject.
2. Identify the samples to be used in the study. Samples should address the representational needs of the range of data being evaluated in addition to showing the relevance of the study.
3. Select the type of case study most appropriate to the needs of the program. Be sure to understand the constraints of the data collection technique selected (See Table 4.4 on page 14).
4. Collect the data to be analyzed through a combination of sources. These sources may include direct observations, interviews, surveys, and direct examinations of relevant literature and documents.
5. Analyze the data, accounting for rival explanations, reproduction of findings, internal validity, plausibility, ability to generalize, and overall coherence.
6. Evaluate the results with regard to your overall ability to generalize and with regard to the internal validity of the data.
7. Write the report and share the findings.

Content Review

Content review refers to the codification and analysis of qualitative data. By coding and classifying the qualitative data sources, this technique attempts to develop an understanding of the meaning of large volumes of qualitative analyses. Data sources for which content review may be useful include free-response surveys or questionnaires, case studies, site inspection reports, focus group reports, or literature searches.

Notes on Coding

Content review relies upon the classification of written communication into categories. Evaluators using content analysis must consider the entire message being communicated when classifying it. The primary activities of content review are unitizing communication (breaking communication into the smallest units that are comprehensible as a message) and categorizing (placing these messages into categories). Thus, content review relies upon a great degree of subjective analysis.

The evaluator should clearly establish what constitutes a unit of communication (even down to a sentence or multi-sentence chunk), in order to create some degree of consistency in the unitization process. Coding categories must be flexible enough to cope with changing data collection and changing analysis methodologies (broader categories capture more cases in each category, but preclude detailed analysis). When possible, select meaningful, descriptive coding category names with which the program managers are familiar. It may be necessary to create terms for patterns that emerge in the course of the review, but select meaningful names so that the program participants and managers can easily understand them.

Only code text for information valuable for use in the specific evaluation. There are several computer packages that can code narrative text. Among the most widely used are ETHNO, Text Analysis Package (TAP), QUALPRO, TEXTBASE, ALPHA, THE ETHNOGRAPH, and HYPERQUAL.

Example Methodology

Below is an example methodology that can be used when utilizing the content review data collection technique.

1. Determine the data source.
2. Establish the coding categories.
3. Code the text.
4. Analyze category frequencies, correlation, and patterns.
5. Write the report.

Advantages and Disadvantages of Content Review

The advantages and disadvantages of content review are shown in Table 4.12 below.

Advantages	Disadvantages
<ul style="list-style-type: none"> • It allows an evaluator to quantitatively analyze qualitative data sources that may provide information not easily collected in another format. 	<ul style="list-style-type: none"> • Improper unitizing or classification can lead to inefficient results.
<ul style="list-style-type: none"> • It allows the evaluator to digest a vast amount of information and distill the commonalities shared across many cases. 	<ul style="list-style-type: none"> • Coding necessarily obscures some of the rich, detailed information imbedded in the text.
<ul style="list-style-type: none"> • It provides a potentially more consistent method of qualitative data analysis. 	<ul style="list-style-type: none"> • The process is still open to claims that it is too subjective to provide generalizable results.

Table 4.12

Advantages and Disadvantages of Content Review

File Review

Reviewing data that has been previously collected and is present in program files or other program documentation may provide information necessary and pertinent to the evaluation of a program. This type of review offers a relatively quick method of discovering what data has already been collected, with an eye toward minimizing the need for additional data collection and the costs associated with that data collection effort.

Types of File Documents

The types of files that should be considered for file review include the following:

- Authorizing legislation, congressional testimony, and comments of legislators.
- Documents related to the regulatory implementation of the legislation (internal review of legislation, public comments, and final rule support documents).
- Budget documents, administrative documents, and meeting minutes.
- Program participant data collected as part of their interaction with the program under review.

Uses of File Review

Although files vary widely between programs, file reviews are particularly helpful in the areas of demographic information on program participants, such as location, size, and type of participant, and information on the types and/or frequencies of services provided. A careful file review may discover key performance measures or required reporting.

Advantages and Disadvantages of File Review

The advantages and disadvantages of file reviews are shown in Table 4.13 below.

Advantages	Disadvantages
<ul style="list-style-type: none"> • Data may be collected without assistance from program participants. In some cases data may be quickly collected for presentation. 	<ul style="list-style-type: none"> • Unfortunately, the information in program files may be limited to administrative information not very useful for some portions of the evaluation process.
<ul style="list-style-type: none"> • File review may offer an inexpensive method for collecting information to evaluate a program. 	<ul style="list-style-type: none"> • Individual program files may be incomplete (depending upon program data collection process) or difficult to locate (particularly for geographically dispersed programs).
<ul style="list-style-type: none"> • A well thought out data collection plan can collect a great deal of useful information to complete an evaluation and minimize additional expense and time. 	<ul style="list-style-type: none"> • Data may not have been collected by the same individual and may reflect multiple subjective judgements. • Data collected in program files may not be very accurate or may contain unknown biases of a program.

Table 4.13
Advantages and Disadvantages of Files Review

Focus Groups

Focus Groups are small, group-facilitated sessions, designed to quickly gather in-depth information while offering stakeholders a forum for direct participation. They are usually facilitated by an outside third party and can yield invaluable information.

Uses of Focus Groups

Focus groups can be used to gather a wide variety of information, including to:

- Obtain general background information about a topic of interest.
- Generate research hypotheses that can be submitted to further research and testing using more quantitative approaches.
- Stimulate new ideas and creative concepts.
- Diagnose the potential for problems with a new program, service, or product.
- Generate impressions of products, programs, services, institutions, etc.
- Learn how respondents talk about their interests to facilitate the design of questionnaires, survey instruments and other research tools.
- Interpret previously obtained quantitative results.
- Provide a forum for stakeholders to present their views and participate in the process.

Advantages of Focus Groups

The advantages of focus groups are that:

- They provide data from a group of people quickly and at a lower cost than individual interviews.
- Researchers interact directly with respondents allowing follow-up questions, clarification of responses, contingent answers, and observation of nonverbal responses.
- They allow respondents to build on responses of other group members.
- They allow flexibility to examine a wide range of topics and subjects.
- They provide a forum to respond to answers more quickly than when written answers are required.

Designing and Conducting a Focus Group

To conduct a successful focus group, that is, to optimize the amount of information that can be gathered from a focus group, the following methodology is proposed:

1. Define the problem and formulate the research question. An internal focus group—a group that is composed of members of the same organization—promotes interaction between different levels of power and measures the group opinion on a topic. An external focus group is composed of participants from inside and outside an organization and can bridge the gap between professionals and the target audience.
2. Identify the sample population for the study. When using a single focus group, the size of the group should range from eight to 12 participants and should represent a cross section of opinions. In a larger study, divide the participants into groups of similar opinion or organizational level, using group data to form a composite picture.
3. Carefully choose a facilitator. A good facilitator is well trained in group dynamics, familiar with a variety of interview techniques, genuinely interested in hearing other's thoughts, and is expressive of his or her own feelings and biases.

4. Generate and pre-test the interview guide. A structured study may be based on steering, leading, testing, obtuse, factual, “feeling,” or anonymous questions. An unstructured or unfocused group is based on the introduction of ideas followed by group discussion.
5. Recruit the sample. Although meal and transportation costs are often sufficient, financial incentives—ranging from \$25 to \$75 per participant—are frequently used to attract participants.
6. Conduct the group. Because sessions typically last 1½ to 2 hours, the focus group should be conducted in a quiet room with comfortable furniture and few distractions. For an external study, facilities geared toward focus groups (often including the use of one-way mirrors, sound equipment, and video recording devices) can dramatically enhance the data recorded.
7. Analyze and interpret the data. Use transcripts and recordings to recreate the discussion and use content analysis and other standard data analysis techniques to glean the greatest amount of information from the material provided by the participants.
8. Write the report. Share and use the findings.

The Cost of Focus Groups

If internal staff prepare the focus group and if internal facilities are used, then costs can be minimal. However, the preparation of a focus group using professional consultants can cost between \$2000 and \$4000 per group.

Section V: Suggestions for Measuring Outcomes of R&D Activities

Evaluators have found that assessing R&D performance is difficult because of specific qualities that are inherent to research and the scientific process. No one can predict what discoveries will be made as a result of R&D activities because they are neither routine nor do they have specific outputs. Any goal that was predicted would most likely shift before the time to evaluate the goal had arrived. Cozzens (1999) outlines four factors that pose particular challenges for meeting the requirements of the Government Performance and Results Act of 1993 (GPRA):

1. The attributes of research that can be tracked and measured are not always important.
2. Significant research events occur unpredictably and cannot be subject to schedules.
3. Many sources of funding and contribution are often integrated within a single research program.
4. There is no easy, accurate method to objectively evaluate research quality or result.

These challenges and the risk of researchers falling back on safe, short-term projects with easily identified and reported outputs, make performance reporting a task that requires constant vigilance and maintenance. Successfully joining the international trend toward accountability and results-oriented research management requires U.S. science agencies to integrate their abilities and goals with public interests and transition from an initial short-term focus to longer-term strategies. Cozzens (1999) concludes that the new accountability could allow agencies to publicly highlight their achievements, communicate more effectively with regulators and the federal government, and directly address the concerns and needs of their stakeholders.

Methods Chosen Depend on Type of R&D

For research and technology development programs, as for other types of programs, most authors stress that, due to the limitations inherent in the individual methods, a complete picture of performance can be shown only if multiple methods are used in combination to balance the subjective versus objective aspects of each and compensate for weaknesses. Quantitative objective metrics are based on numerical measures of R&D input and output—staff count, R&D cost, time spent, number of publications, cost reductions, and goals met. Quantitative subjective metrics are based on nonnumerical judgements that are converted into numeric values and ratings via profiles, scaling models, checklists, and scoring models. Qualitative metrics measure human resources and other aspects of R&D performance using self-assessment, supervisory rating, peer rating, and external audit. GPRA gives agencies that can not define quantitative performance goals for their programs the option of using measurable qualitative goals and measures, if approved by OMB (OMB 1999).

A 1993 study by the government of Canada (Canada 1993) looked at the data collection methods for measuring the socioeconomic impact of R&D. The assessment methods that were determined to be the most applicable for the Canadian government's purposes are shown along with their relative cost in Table 4.14 (on the following page). The Canadian study team concluded that, although more work is needed, available methods can provide a limited assessment of the impact of government R&D activities, with some providing "reasonably accurate" quantitative assessments. Their report suggests using cost-benefit methods as a way of analyzing applied research but not basic research. The methods they discuss include modified peer review, user surveys, benefit-cost methods, case studies, partial indicators, and integrated partial indicators. Modified peer reviews incorporate assessment of research quality by scientific peers with assessment of the socioeconomic value of the research. Partial indicators are quantitative information on program inputs (i.e., funding, human resources), outputs, activities, or impacts. The information obtained using partial indicators can be analyzed further with integrated partial indicators. These assign an overall score for the project based on the outcome of the individual partial indicators.

The Canadian study team suggests that determination of which method to use is dependent on three factors:

1. Whether the assessment is occurring before or after the R&D is completed.
2. What type of R&D is involved (basic/strategic research, applied research, product/process development).
3. The purpose/category of the R&D.

Included in the categories are R&D that was undertaken to:

1. Contribute to society's R&D infrastructure.
2. Provide information for policies or standards.
3. Contribute to attainment of government policies and economic development.
4. Support industrial innovation.

Methods	R&D Time Frame	R&D Type	R&D Purpose*	Relative Cost
Modified Peer Review	Past, ongoing, and future	All	All (least well suited for Category 1)	Low/Medium
User Surveys	Past and ongoing	Applied research and development	Categories 2 - 4	Medium
Benefit-Cost Methods	Past (ongoing and future in certain circumstances)	Applied research and development	Category 4 (2 and 3 in certain circumstances)	High
Case Studies	Past	Applied research and development	Categories 2 - 4	Medium
Partial Indicators	Past and ongoing (limited future)	All	All	Low
Integrated Partial Indicators	Future	Applied research and development	Categories 2 - 4	Low
*Categories:				
1. R&D Infrastructure 2. Policy Development 3. Policy Attainment 4. Industry Support				

Table 4.14
Methods for Assessing Socioeconomic Outcomes of R&D

Appendix A: Definitions

Because people often associate different meanings to “common” terminology, definitions are always tricky and controversial. Such may be the case with the definitions given herein. Please remember that many of these definitions are applicable with respect to the U.S. Department of Energy and its operations. The intent here is to define terminology such that the reader can get a general understanding of it. We do not intend to be prescriptive or inflexible, nor do we admit to being the highest source of information.

Accountability

The obligation a person, group, or organization assumes for the execution of assigned authority and/or the fulfillment of delegated responsibility. This obligation includes: answering—providing an explanation or justification—for the execution of that authority and/or fulfillment of that responsibility; reporting on the results of that execution and/or fulfillment; and assuming liability for those results.

Activity

Actions taken by a program or an organization to achieve its objectives.

Assessment

An all-inclusive term used to denote the act of determining, through a review of objective evidence and witnessing the performance of activities, whether items, processes, or services meet specified requirements. Assessments are conducted through implementation of activities such as audits, performance evaluations, management system reviews, peer reviews, or surveillances, which are planned and documented by trained and qualified personnel.

Baseline

The initial level of performance at which an organization, process, or function is operating upon which future performance will be measured.

Benchmarking

- 1.) To measure an organization’s products or services against the best existing products or services of the same type. The benchmark defines the 100 percent mark on the measurement scale.
- 2.) The process of comparing and measuring an organization’s own performance on a particular process against the performance of organizations judged to be the best of a comparable industry.

Bottom Up

Starting with input from the people who actually do the work and consolidating that input through successively higher levels of management.

Cascaded Down

Starting with a top level of management, communicated to successively lower levels of management and employees.

Characteristics

Any property or attribute of an item, process, or service that is distinct, describable, and measurable.

Continuous Improvement

- 1.) The undying betterment of a process based on constant measurement and analysis of results produced by the process and use of that analysis to modify the process.
- 2.) Where performance gains achieved are maintained and early identification of deteriorating environmental, safety, and health conditions is accomplished.

Corrective Action

Actions taken to rectify conditions adverse to quality and, where necessary, to preclude repetition.

Criteria

The rules or tests against which the quality of performance can be measured.

Goal

- 1.) The result that a program or organization aims to accomplish.
- 2.) A statement of attainment/achievement, which is proposed to be accomplished or attained with an implication of sustained effort and energy.

Guideline

A suggested practice that is not mandatory in programs intended to comply with a standard. The word "should" or "may" denotes a guideline; the word "shall" or "must" denotes a requirement.

Impact

Characterization of the outcome of a program as it relates to specific objectives.

Item

An all-inclusive term used in place of the following: appurtenance, sample, assembly, component, equipment, material, module, part, structure, subassembly, subsystem, unit, documented concepts, or data.

Lessons Learned

A "good work practice" or innovative approach that is captured and shared to promote repeat application. A lesson learned may also be an adverse work practice or experience that is captured and shared to avoid recurrence.

Line Manager

Includes all managers in the chain of command from the first-line supervisors to the top manager.

Management

All individuals directly responsible and accountable for planning, implementing, and assessing work activities.

Measurement

The quantitative parameter used to ascertain the degree of performance.

Metric

A standard or unit of measure.

Objective

A statement of the desired result to be achieved within a specified time.

Occurrence

An unusual or unplanned event having programmatic significance such that it adversely affects or potentially affects the performance, reliability, or safety of a facility.

Outcome

The expected, desired, or actual result to which outputs of activities of an agency have an intended effect.

Outcome Measure

An assessment of the results of a program activity or effort compared to its intended purpose.

Output

A product or service produced by a program or process and delivered to customers (whether internal or external).

Output Measure

The tabulation, calculation, or recording of activity or effort and can be expressed in a quantitative or qualitative manner.

Performance-Based Management

A systematic approach to performance improvement through an ongoing process of establishing strategic performance objectives; measuring performance; collecting, analyzing, reviewing, and reporting performance data; and using that data to drive performance improvement.

Performance Expectation

The desired condition or target level of performance for each measure.

Performance Indicator(s)

- 1.) A particular value or characteristic used to measure output or outcome.
- 2.) A parameter useful for determining the degree to which an organization has achieved its goals.
- 3.) A quantifiable expression used to observe and track the status of a process.
- 4.) The operational information that is indicative of the performance or condition of a facility, group of facilities, or site.

Performance Measure

A quantitative or qualitative characterization of performance.

Performance Measurement

The process of measuring the performance of an organization, a program, a function, or a process.

Performance Objective

- 1.) A statement of desired outcome(s) for an organization or activity.
- 2.) A target level of performance expressed as a tangible, measurable objective, against which actual achievement shall be compared, including a goal expressed as a quantitative standard, value, or rate.

Performance Result

The actual condition of performance level for each measure.

Process

An ongoing, recurring and systematic series of actions or operations whereby an input is transformed into a desired product (or output).

Process Improvement

A set of management techniques for controlling and improving the effectiveness and efficiency of a process. In order to be measured, monitored, and analyzed, the process must be repeated frequently, perhaps weekly or monthly at a minimum. It must also have measurable inputs and outputs, and the process must be controllable.

Program Evaluation

An assessment, through objective measurement and systematic analysis, of the manner and extent to which Federal programs achieve intended objectives.

Quality

A degree to which a product or service meets customer requirements and expectations.

Quality Management

The management of a process to maximize customer satisfaction at the lowest cost.

Reengineering

The radical redesign of current business processes with the intent of reducing cost and cycle time resulting in increased customer satisfaction.

Root Cause

The basic reasons for conditions adverse to quality that, if corrected, will prevent occurrence or recurrence.

Root Cause Analysis

An analysis performed to determine the cause of part, system, and component failures.

Self-Assessment

A systematic evaluation of an organization's performance, with the objective of finding opportunities for improvement and exceptional practices. Normally performed by the people involved in the activity, but may also be performed by others within the organization with an arms-length relationship to the work processes.

Senior Management

The manager or managers responsible for mission accomplishment and overall operations.

Situation Analysis

The assessment of trends, strengths, weaknesses, opportunities, and threats, giving a picture of the organization's internal and external environment to determine the opportunities or obstacles to achieving organizational goals. Performed in preparation for strategic planning efforts.

Stakeholder

Any group or individual who is affected by or who can affect the future of an organization, e.g., customers, employees, suppliers, owners, other agencies, Congress, and critics.

Strategic Planning

A process for helping an organization envision what it hopes to accomplish in the future; identify and understand obstacles and opportunities that affect the organization's ability to achieve that vision; and set forth the plan of activities and resource use that will best enable the achievement of the goals and objectives.

Task

A well-defined unit of work having an identifiable beginning and end that is a measurable component of the duties and responsibilities of a specific job.

Total Quality Management

- 1.) A management philosophy that involves everyone in an organization in controlling and continuously improving how work is done in order to meet customer expectations of quality.
- 2.) The management practice of continuous improvement in quality that relies on active participation of both management and employees using analytical tools and teamwork.

Validation

An evaluation performed to determine whether planned actions, if implemented, will address specific issue(s) or objective(s).

Verification

- 1.) A determination that an improvement action has been implemented as designed.
- 2.) The act of reviewing, inspecting, testing, checking, auditing, or otherwise determining and documenting whether items, processes, services, or documents conform to specified requirements.

Appendix B: Acronyms

ABM	Activity-based management
AOP	Annual Operating Plan
APQC	American Productivity and Quality Center
ARL	Army Research Laboratory
ASQC	American Society for Quality Control
BMOP	Business Management Oversight Pilot
CEO	Chief Executive Officer
CFO	Chief Financial Officer
CIO	Chief Information Officer
COO	Chief Operating Officer
CPI	Consumer Price Index
CRT	DOE Contract Reform Team
CSF	Critical success factor
DOE	U.S. Department of Energy
ES&H	Environment, safety and health
EVA	Economic value-added
FY 19xx	Fiscal Year 19xx
FY 200x	Fiscal Year 200x
GAO	General Accounting Office
GPRA	Government Performance and Results Act of 1993
IBM	International Business Machines
IRG	Initial Review Group
ISO	International Standards Organization
JIT	Just-in-time
JPL	Jet Propulsion Laboratory
MBNQA	Malcolm Baldrige National Quality Award
M&I	Management and Integrating
M&O	Management and Operating
NAC	National Advisory Council
NASA	National Aeronautics and Space Administration
NIH	National Institutes of Health
NPR	National Performance Review
NRC	Nuclear Regulatory Commission
NSF	National Science Foundation

OMB Office of Management and Budget
OSHA Occupational Safety and Health Administration
PBM SIG Performance-Based Management Special Interest Group
PDCA Plan-Do-Check-Act Cycle
POCMs Performance objectives, criteria, and measures
QCDSM Quality, cost, delivery, safety, and morale
R&D Research and development
ROI Return on investment
S&T Science and technology
SAI Strategic Alignment Initiative
SPC Statistical process control
TQM Total Quality Management
UC University of California
UCOP University of California Office of the President
URL Universal Resource Locator
WWW World Wide Web

Appendix C: References/Suggested Reading

- Allen, Sophie, *Name and Address Data Quality*, MasterSoft Research Pty Ltd, Sydney Australia, 1996. Available on the Massachusetts Institute of Technology's (MIT's) Total Data Quality Management (TDQM) Web site at: <http://web.mit.edu/tdqm/papers/other/allen.html>.
- American Evaluation Association, "Links of Interest to Evaluators" Web site. Available at: <http://www.eval.org/EvaluationLinks/links.htm>.
- Averch, Harvey, "Economic Approaches to the Evaluation of Research," *Evaluation Review*, 1994.
- Bozeman, Barry and Julia Melkers (Eds.), *Evaluating R&D Impacts: Methods and Practice*, Kluwer Academic Publishers, 1993.
- Cozzens, Susan, "Are New Accountability Rules Bad for Science?" *Issues in Science and Technology*, National Academy of Sciences, Summer 1999.
- Cozzens, Susan, "Literature-Based Data in Research Evaluation: A Manager's Guide to Bibliometrics." Report to the National Science Foundation, 1989.
- Cozzens, Susan, Steven Popper, James Bonomo, Kei Koizumi, and Ann Flannagan, *Methods for Evaluating Fundamental Science*, DRU-875/2-CTI, Critical Technologies Institute, RAND Corp., October 1994.
- Donaldson, Scott and Stanley Siegel, *Cultivating Successful Software Development: A Practitioner's View*, Upper Saddle River, NJ: Prentice Hall PTR, 1997.
- English, Larry, *Help for Data Quality Problems—A Number of Automated Tools Can Ease Data Cleansing and Help Improve Data Quality*, InformationWeek Labs, Issue No. 600, October 7, 1996. Available at: <http://www.informationweek.com/>
- Fetterman, David, Shakey Kaftarian, and Abraham Wandersman, *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*, Sage Publications, 1996.
- Firth, Chris, *Data Quality in Practice: Experience From the Frontline*, Citibank Singapore, 1996. Available on the Massachusetts Institute of Technology's (MIT's) Total Data Quality Management (TDQM) Web site at: <http://web.mit.edu/tdqm/papers/other/firth.html>.
- Geisler, Elie, "The Metrics of Technology Evaluation: Where We Stand and Where We Should Go from Here." Presented at the 24th Annual Technology Transfer Society Meeting, 1999. Available at: <http://www.stuart.iit.edu/workingpapers/index.html>.
- Government of Canada, *Methods for Assessing the Socioeconomic Impacts of Government S&T*, Working Group on S&T Financial Management and Mechanisms, 1993.
- Haltry, Harry, *Performance Measurement: Getting Results*, Urban Institute Press, 1999.

- Hauser, John, "Research, Development, and Engineering Metrics," *Working Paper for The International Center for Research on the Management of Technology*, 1997.
- Indian and Northern Affairs Canada (INAC), *Facts From Stats: Improving the Quality of Data*, Issue No. 5, June 1995. Available at: http://www.ainc-inac.gc.ca/nr/nwltr/sts/1995fs-5_e.html.
- Link, A., "Methods for Evaluating the Return on R&D Investments," in *Evaluating R&D Impacts: Methods and Practice*, Kluwer Academic Publishers, 1993.
- Mohr, Lawrence, *Impact Analysis for Program Evaluation*, Sage Publications, 1995.
- Narin, Francis, Dominic Olivastro, and Kimberly Stevens, "Bibliometrics/Theory, Practice and Problems," *Evaluation Review*, Vol.18, No.1, February 1994.
- National Academy of Sciences, Committee on Science, Engineering, and Public Policy (COSEPUP), *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*, NAPA Press, 1999.
- National Partnership for Reinventing Government (formerly National Performance Review), *Serving the American Public: Best Practices in Performance Measurement*, June 1997. Available at: <http://www.npr.gov/library/papers/benchmrk/nprbook.html>.
- National Research Council (NRC), *Quantitative Assessments of the Physical and Mathematical Sciences: A Summary of Lessons Learned*, National Academy Press, 1994.
- National Science and Technology Council, Committee on Fundamental Science, "Assessing Fundamental Science," July 1996. Available at: <http://www.nsf.gov/sbe/srs/ostp/assess/start.htm>.
- Patton, Michael Quinn, *Utilization-Focused Evaluation: The New Century Text*, Sage Publications, 1997.
- Rossi, Peter, Howard Freeman, and Mark Lipsey, *Evaluation: A Systematic Approach*, Sage Publications, 1999.
- Segev, Arie, *On Information Quality and the WWW Impact: A Position Paper*, Fisher Center for Information Technology and Management, Haas School of Business, University of California at Berkeley, 1996. Available on the Massachusetts Institute of Technology's (MIT's) Total Data Quality Management (TDQM) Web site at: <http://web.mit.edu/tdqm/papers/other/segev.html>.
- Sylvia, Ronald, Kathleen Sylvia, and Elizabeth Gunn, *Program Planning and Evaluation for the Public Manager*, Waveland Press, 1997.
- Tassev, Gregory, "Rates of Return from Investments in Technology Infrastructure," *National Institute of Standards and Technology (NIST) Planning Report 96-3*, NIST 1996.
- Treasury Board of Canada, "Program Evaluation and Review," Web site. Available at: <http://www.tbs-sct.gc.ca/pe/contents.htm>.

- U.S. General Accounting Office (GAO), *Measuring Performance: Strengths and Limitations of Research Indicators*, GAO/RCED-97-91, 1997. Available at: <http://www.gao.gov/archive/1997/rc97091.pdf>.
- U.S. General Accounting Office (GAO), *Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information*, GGD-99-139, 1999. Available at: <http://www.gao.gov/archive/1999/gg99139.pdf>.
- U.S. General Accounting Office (GAO), "Special Publications, Policy, and Guidance Documents" Web site. Available at: <http://www.gao.gov/special.pubs/publist.html>.
- Wholey, Joseph, Harry Haltry, and Kathryn Newcomer (Eds.), *Handbook of Practical Program Evaluation*, Jossey-Bass Publishers, 1994.
- Wolf, Fredric, *Meta Analysis: Quantitative Methods for Research Synthesis*, Sage University Paper Series on "Quantitative Applications in the Social Studies," Series No. 07-001, Sage Publications, 1986.

