Center for
Technology in Government

Working Paper Series

Critical Issues and Practical Challenges of IT Tools for Policy Analysis and Program Evaluation

CTG Working Paper No. 04-2008

March 2008

Authors:
Sharon S. Dawes
Theresa A. Pardo

Center for Technology in Government
187 Wolf Road, Suite 301
Albany, NY 12205
Phone: (518) 442-3892
Fax: (518) 442-3886
E-mail: info@ctg.albany.edu
www.ctg.albany.edu

Note: This paper was first presented at the Fall Conference of the Association for Public Policy Analysis and Management (APPAM), October 28-30, 2004, Atlanta, GA, USA.

Critical Issues and Practical Challenges of IT Tools for Policy Analysis and Program Evaluation

Sharon S. Dawes and Theresa A. Pardo
Center for Technology in Government
University at Albany, SUNY
sdawes@ctg.albany.edu

Abstract
Policy makers and public managers want and need to know how well government programs perform, but few have the information to accurately and continuously evaluate them. The dynamic nature of public programs, and the traditional methods used to assess them, compound this problem. Performance measurement and performance-based decisions can be improved by more sophisticated information systems designed for to support analysis and decision making. However, such systems demand close and continuing involvement of program staff, attention to programmatic context, and much better understanding of business processes and the data they generate. Through the use of a case example, the prototype Homeless Information Management System, this paper highlights how attention to these issues can lead to useful and usable performance analysis and evaluation systems.

Introduction

Traditional program evaluations, program audits, and performance measurement programs are all formal, information-based attempts to answer the question "How are we doing?" Most audit and evaluation approaches rely on external reviewers collecting their own data, sometimes supplemented with standard statistical reports and administrative data from the programs under review. Performance measurement activities are usually conducted by people internal to the agency, but separate from those who manage the programs. These efforts typically use existing administrative data or collect separate information specifically for performance reporting. Intended to be "objective," these traditional approaches are designed to be independent of day-to-day program operations. As a consequence, however, the people with the most program knowledge and the discretion and authority to act quickly on many findings – program managers and staff – are generally not involved or enthusiastic about program evaluation.

At the same time, the detailed databases and information systems that support program operations are seldom used routinely or to best advantage for program and policy assessments. Ironically, recent developments in information technology can enable unprecedented access to this information for ongoing program assessment, decision making, and planning. These tools, allow the plan-implement-evaluate cycle to be compressed by providing access to more relevant information more often, and in a more useful and usable way.

These two factors -- uninvolved program staff and unexploited program information -- mean that program performance is assessed only periodically or sporadically and without making the most effective use of either expert knowledge or existing expensive data sources. Consequently, performance information produced by these traditional means is seldom actually used by managers to improve program designs or operations (de Lancer Julnes & Holzer, 2001). This is not to say that formal, independent program evaluation is not needed or valuable, but that alone it is inadequate to the goal of improving program performance.

Drawing on the literature of program evaluation, performance measurement; and management information systems, we explore the changing nature of performance measurement and the emerging use

of advanced technologies to support it. A brief examination of decision making and decision support technologies lays the foundation for an analysis of a case study in which a group of government managers and nonprofit service providers created their own performance information system. This system was designed to support their mutual need to understand, assess, and improve the performance of programs that serve homeless people. We conclude with a discussion of the ways in which such information systems can complement traditional evaluation approaches.

Evaluating government performance

In its strict definition, the term *policy analysis* is used to describe the process of developing and evaluating alternative courses of action before policy decisions are made, although the term is often applied to the evaluation of alternatives to existing policies. *Program evaluation* is the process of after-the-fact review of implemented policies. *Program audits* similarly look at the effects of policy implementation in terms of the cost and effectiveness of program operations. *Performance measurement* focuses on a few key factors that are believed to be the most important indicators of program quality and effectiveness. In practice, however, all these terms tend to be used in similar ways. Any of them might denote an effort to determine how or how well a government program or policy is working. In that sense, they are all forms of evaluation.

Newcomber (1966) classifies formal evaluation efforts into three basic strategies: problem-based investigations, performance assessments, and impact evaluations. Each of these strategies has its own implications regarding focus, evaluator, data requirements, time, resources, tools, and the reaction of program staff. For example, a problem-based investigation may focus on immediate issues in program operations. It may be conducted over a period of weeks by state auditors using agency records and generating a performance audit report which is received without enthusiasm or with outright hostility by program managers. Because performance audits have grown out of the financial audit tradition, they tend to be conducted with a strong emphasis on financial factors and with a focus on finding explanations for documented or alleged poor performance. Given their usually negative connotations, agency leaders and program managers seldom welcome them. By contrast, an impact evaluation conducted by an agency's own research staff might go beyond agency records to gather information from clients and other external sources. It might take a year or longer and investigate questions that cannot be addressed in other ways. Depending on the questions asked and the evaluation and reporting methods used, program staff reaction can range from resistance to enthusiasm. Regardless of there differences in method or data type, all three strategies rely on professional evaluators whether they be research professionals, auditors, or other trained observers. None of the three strategies typically makes extensive use of advanced information technology for data collection, analysis, or presentation.

In the past several decades, many recommendations have been made for methodological improvements that can enhance the credibility (and therefore presumably the use) of findings emanating from these formal evaluations. Improvements in sampling strategies, evaluator credentials, data collection methods, replication studies, documentation tools, and other areas have all been considered in the quest to improve the quality of formal evaluation (Caudle, 1994).

However, evaluation results still tend not to have a concrete, observable effect on decisions or practices. Research has shown that the extent to which any evaluation will affect the program under review depends on a variety of factors including the involvement of stakeholders (Patton, 1986), organizational characteristics (Grifel,1994, Theurer, 1998, Wang and Gianakis, 1999), and evaluation techniques (Williams, et al., 1994). In order to be widely used by program managers and others in the organization, evaluation data must be "valid, complete, and widely accepted by other stakeholders." (Carter, 1994:578). Carter offered several recommendations for increasing the use of evaluation and outcome or performance monitoring including timely reports, detailed breakdowns of data, active participation by program staff, data with high face validity, demonstrated used of outcome information, regular repeated measures, and appropriately designed information systems (Carter, 1994).

Coplin, et al (2002) focused on the role of the evaluator as not only researcher but change agent. They offered guidelines for embracing the broadest understanding of continuous improvement, for adopting the notion of a problem-solving community of both researchers and stakeholders, for resolving the ever-present conflict between rigor and relevance in their reports to different audiences, and for communicating research concepts and logics to lay government officials and citizens.

Williams, et al. (1994) explored three areas of potential conflict between evaluation researchers and the managers of programs being evaluated. One is the different priorities the two groups assign to objectivity vs. subjectivity and analysis vs. action. A second source of conflict is that managers tend to be committed to the "worth" of their work and its ability to succeed, while researchers tend to be committed to objective questioning of the work in terms of both purpose and performance. Third is a conflicting perspective on time. Researchers value a long-term view and "enough" time for analysis; managers view time as short-term and dynamic with demands for action that cannot wait for the research to unfold. All of these issues make formal program evaluations difficult to conduct and weak in their ability to influence change.

Performance measurement as a strategy relies less on formal research programs than on organizational self-assessment. As a management concept, performance measurement has the merits of focus on clear objectives, standards for their accomplishment, and regular feedback about performance into decisions about strategies and practices (Drucker, 1973). As such, performance measures can be used by organizations in two ways: (1) to establish priorities and track progress toward meeting them through regular oversight and (2) to improve the ability of managers to understand their operations and their capacity to improve them (Thompson, 2000). Emphasis on performance-based government stems from the reinventing government movement (Osborne & Gaebler, 1992) which extended these ideas from business into the public sector. Essentially a performance-driven government is expected to deliver more effective programs and make more efficient use of public funds. It is further expected to improve accountability of elected officials and administrators while giving them the information they need to improve operations and outcomes (Thompson, 2000).

In the early 1990s, the National Performance Review launched a broad effort to put performance front and center in government reform efforts. The Government Performance and Results Act of 1993 institutionalized the requirement for performance targets and measures in federal agencies and these have been further reinforced since 2000 by the President's Management Agenda (PMA) which assesses agency performance as on a uniform set of five factors: human capital management, financial management, e-government, competitive sourcing, and integration of budgets and performance. Most states also have legislatively or administratively mandated performance measurement systems (Melkers and Willoughby, 1998).

Despite the benefits usually associated with performance measurement, the practice has also generated controversial and unintended effects. These include a bias toward quantifiable elements of performance and inadequate attention to qualitative aspects, which of ten generates a distorted perspective about what is really important. In addition, successive aggregation and simplification of data as it moves upward to higher levels of oversight weakens its connection with operational relevance. The result can be over-reliance on simplified, symbolic numbers divorced from their context and from the processes of management (Thompson, 2000).

Based on a study of state and local government employees, de Lancer Julnes and Holzer (2001) found that the adoption of performance measurement systems can best be explained by the effects of rational arguments about their benefits along with formal mandates. However, the actual use of measurement systems is shaped by political and cultural factors including the participation of internal stakeholders (i.e. program mangers and line staff) and external interest groups as well as the continued support of elected officials and the public. Successful performance measurement programs involve more than choosing and promoting measures—they also require organizational "readiness," involvement of stakeholders and unions, patience, and emphasis on a culture of performance improvement.

Information technology tools for analysis and decision making

Information technology innovations combined with new strategies for organizational planning and decision making provide alternative ways to capture, store, organize, access, and use data for performance assessment and management. The focus on performance measurement in the public sector (Nyhan and Martin, 1999) and customer-focused marketing in the private sector (Flynn, Curran, and Lunney, 2002) have provided the impetus for organizations to invest in technologies that support these new assessment and management models. Tools such as mass storage devices, growth in processing capability, ubiquitous access to the Web, and new software designed to integrate data from multiple disparate sources are just some of the innovations that can enable organizations to draw on vast data resources for virtually "real-time" use in decision making (Liang and Miranda, 2002).

Managers engage in at least three types of decision making processes; structured routine, semi-structured non-routine, and unstructured non-routine. Different resources are needed to support on the type of decision being made. Management Information Systems (MIS) provide information as a basis for routine, structured decision making, whereas Decision Support Systems (DSS) provide information to support semi-structured, non-routine decision making. A DSS is designed to improve decision making by merging human intuition and judgment with computer systems (Eom and Lee, 1989). Table 1 briefly compares the attributes of MIS and DSS systems.

Table 1. Comparison of Attributes of Management Information and Decision Support Systems

Management Information System	***Decision Support System***
Transaction focused	Decision focused
Difficult to use by non-computer people	Easy to use by decision makers
Batch generated on a pre-planned schedule	User initiated and controlled interactively
Rigid, pre-selected, pre-programmed analysis	Emphasizes flexibility, adaptability and quick response
Records transactions	Combines the use of models and analytic techniques with traditional data access and retrieval functions.
Adapted from Carter et.al., 1992	

Many definitions of DSS can be found in the literature. In essence a DSS is any computer-based information system that supports decision making. One of the key design principles of a DSS is that neither the DSS nor the decision maker alone is as effective as the two combined. A DSS should be designed to improve decision making by merging human intuition-judgment and computer systems (Lee and Eom, 1990). According to Keen and Scott Morton (1978) the merger of intuition and data is required in a decision for which managerial judgment alone would not be adequate and one for which the model or data alone are also inadequate because the solution involves judgment and subjective analysis. One definition that emerged from research on strategic policy making in the public sector characterizes DSS as an electronic aid to improve governance outcomes by facilitating more systematic and accurate identification, analysis, assessments, and linkages of different policy problems, resources, objectives, solutions, costs, benefits, risks, probabilities, priorities, processes, outputs, and outcomes (Cloete, et al. 2003).

A well-designed DSS is useful to an organization as it provides the means to use operational data, managerial expertise, and powerful analytical tools to understand ongoing performance and to make decisions about how to continuously improve it. The basic concepts underlying DSS grew out of managers' dissatisfaction with the limitations of earlier MIS systems with their focus on transactions, pre-defined reports, rigidity, and reliance on computer professionals for access to information. By contrast, a

DSS is useful to, and used by, decision makers because of its flexible, user-controlled methods for displaying and analyzing data and for formulating and evaluating alternative scenarios.

A DSS relies on improved user interfaces, graphical and statistical methods, and simulation and optimization models to support better analysis and decision making (Alter, 1992). However, these same features make a DSS more difficult to design and implement than traditional MIS systems. An effective DSS requires a foundation drawn from the business management processes of the organization, not the technologies, or the tools, or even in the data itself. It requires deep understanding of the business models that frame the organization's operations. Identifying these core processes, explicating their underlying models and assumptions, and developing information resources that can be integrated into those processes is critical to the success of a DSS. When the processes are shared or distributed across multiple organizations, as they often are in government, these difficulties are multiplied.

A DSS is built from a number of components; each designed for a specific purpose. Early DSS research by Keen and Scott Morton (1978) identified three; a component that facilitates inquiries, a knowledge component that contains the database, and a problem processing component that contains the models used to generate alternative solutions. Flynn et.al. call these components the dialogue component, the data component, and the model component. (Flynn et.al., 2002)

The inquiry component is the interface between the system and the user. In effect, the system is what it looks like to the user; thus the software interface between the user and the underlying models and databases must be humanized. The likelihood that a decision maker will accept and use a DSS often depends on how it is presented through this interface. (Keen and Scott Morton, 1978) According to Goddard et.al. (2003) this layer provides the interface by which users enter their queries against the data – whether by direct access to a single data set or through a complex query that reaches out to multiple data sets and integrates and aggregates them according to a pre-determined model.

The knowledge component encompasses both data and information (Goddard et. al., 2003). This layer contains the data bases – whether spatial, multi-dimensional, distributed, relational or other. To support the decision making process, the knowledge component of a DSS must contain actionable data necessary to analyze and trace a problem. Data of sufficient quality to support the specific assessments is critical. Data completeness, accuracy, and timeliness must be addressed with respect to the kinds of assessments to be made. Further, explicit agreements about meaning and relevance of data are necessary to determine its usability and this vital but challenging step requires active engagement with users

The problem processing component of a DSS must provide models or frameworks to help users search for root causes of problems and recognize the pre-symptoms of undesirable situations. This ability to support the identification of potential problems and to respond accordingly is seen as the primary benefit of a DSS (McGowan and Lombardo, 1986) In the problem processing component, modeling and other analytic techniques must be provided to enhance information use as a basis for making choices (Andersen and Dawes, 1991) The problem processing component is what most separates DSS from MIS. It is here that DSS supports the merger of intuition and data, providing a well-informed managerial user with access to data, opportunity to query and analyze data, and tools for what-if analysis of alternative decisions.

The fundamental characteristics of DSS technologies have been incorporated into a variety of specific tools to support managerial decision in the last decade. New data management technologies such as multi-dimensional data warehouses, data marts, and Web servers enable large volumes of data to be stored, organized and made accessible for more efficient use. New analytical technologies enable managers to use online-analytical processing and visualization tools such as geographic information systems to explore data. New delivery mechanisms such as Web portals allow them to integrate the readily available operational data into decision making and planning activities (Cloete, 2003, Liang and Miranda, 2001, Goddard et.al. 2003)).

Alternative Performance Assessment Strategies

The foregoing overview of research literature on performance evaluation and the technologies that can support it suggests a framework for comparing different public sector evaluation approaches. Table 2 compares formal program evaluations, performance audits, performance reporting systems, and what we call "self-assessment systems" that are based on the principles of DSS. These four types are compared on the motivation for assessment; their frequency, scope and focus, the kind and quality of information typically used, their cost implications, the nature of program staff involvement, and the usability of findings for program management.

Table 1. Comparison of Selected Evaluation Approaches

	Evaluation Approach			
Characteristic	**Formal program evaluation**	**Performance audit**	**Performance reporting systems**	**Self-Assessment Systems**
Motivation	Planned review or to address a problem, external accountability	To address a problem, external accountability	Response to monitoring & reporting requirements, external accountability, internal management	Continuous performance improvement, internal management
Frequency	Infrequent	Situational	Periodic, usually semi-annual or annual	Ongoing
Scope and focus of evaluation questions	Flexible	Determined by nature of the problem	Fixed by formal rules or directives	Flexible
Kind of information available	Detailed data in context	Detailed data in context	Aggregated, pre-defined data removed from context	Detailed data in context plus expert knowledge
Data Quality	Depends on quality of data sources and understanding of evaluator	Depends quality of data sources and understanding of evaluator	Generally low	Generally high
Cost	Medium to High depending on scope	Medium to High, depending on nature of the problem	High to initiate, low to maintain	High to initiate, medium to sustain
Nature of program staff involvement	Reactive, sometimes defensive	Reactive, often defensive	Little to none after initiation	Fully engaged
Potential usability in program operations	Indirect, delayed	Indirect, delayed	Indirect if at all, delayed	Direct and continuous
Typical IT tools	none	none	Standard reports, statistical packages.	Full range of DSS technologies

All of these approaches are costly and resource intensive, and they are designed to generate information for different audiences and purposes. However, self assessment systems offer some advantages that the other methods lack. They take full advantage of the expert knowledge of program staff and engage them fully in the process of building the system, selecting and defining data, and using it for performance monitoring and improvement. Because these systems can deliver continuous information and ongoing opportunities to identify problems and act on them, they offer more frequent opportunities to assess performance and make needed changes.

Making technology investments to support performance assessment and decision making, a challenge in any environment, is more difficult in the public sector for several reasons. Government programs and government managers need to identify and work within multiple, sometimes competing, agendas. As a result, they have no clearly defined bottom line, but instead need to balance a variety of demands and expectations about performance. These multiple agendas represent the influence of multiple stakeholders both within and outside government. These stakeholders place more constraints on action due to their sheer numbers and the variety of perspectives and priorities they represent. In addition, most government programs rely on interagency or intergovernmental processes or may cross the boundaries of public and private sector organizations, making accountability both more important and more difficult to demonstrate. In the section below, we recount the experience of one government program that designed such a system to illustrate these difficulties and the strategies that can be used to address them. We then conclude with observations about its utility and prospects for performance assessment and program management.

Case: The Homeless Information Management System (HIMS) Prototype

Each night in New York State nearly 29,000 homeless people receive emergency shelter and support services. The 6,400 families and 10,000 single adults require assistance in dealing with their immediate incidence of homelessness as well as assistance in dealing with a variety of other problems including domestic violence, alcoholism or substance abuse, poor parenting skills, mental illness, and a lack of education or employment skills. Many lack the skills to maintain their own housing.

New York State and its localities spend millions of dollars and devote substantial effort in providing both housing and services to these homeless single adults and families. The Bureau of Shelter Services (BSS) manages the temporary housing services program in New York State. The program is comprehensive in that it determines eligibility and need for services, provides case management, direct services, and referrals to outside service providers. The cost to federal, state, and local government programs for the homeless in New York State is estimated to be $350 million annually, of which $130 million is spent on service programs.

Professionals in the homeless services field believe the various service programs they provide to homeless people reduce public assistance costs by helping people achieve independence. But there is little evidence to either support or challenge this belief. Program managers do have quarterly aggregated statistical reports from shelter and service providers regarding the numbers of people being served for payment purposes. However, information about service effectiveness is mostly anecdotal. To understand program effectiveness in useful way, state and local program managers need consistent and complete data across service programs and over time to determine the most effective mix of services for a particular client population. This type of data resides in various separate systems or in paper records that are not integrated. As a result, it is unclear whether self-sufficiency, reduced recidivism, reduced dependence on public assistance, and improved overall life skills are being systematically achieved.

BSS staff shared growing government-wide interest in outcome-based assessments as well as a general appreciation for how new technologies might support data integration and access. Accordingly, they began to consider the feasibility of a new information resource which they called the Homeless Information Management System (HIMS), to help them assess effectiveness across programs, services, and population groups. The organizations involved included several state agencies, three local governments, and a number of shelter programs operated by nonprofit service organizations. Some of

these are very small operations serving only a few people or families at a time. Others are major programs of large well-established organizations like the Salvation Army and the American Red Cross. The new system was intended to fill an important gap in program management by continuously linking and comparing information on services to information about client outcomes.

The project was a departure from traditional regulatory relationships among the participants in that it attempted to create both a community of practice and a jointly defined, shared data resource for the voluntary internal use of all the participating organizations. It required that these organizations agree on some key performance criteria, jointly define key data elements, understand each others business processes, and look more deeply into the information policies that would govern the use of the data. The HIMS prototype drew its data from multiple existing case management systems and financial systems already in use in the participating organizations. Overall, the project sought to determine whether it was:

- feasible to develop an integrated database from such a wide variety of data sources
- possible to accurately match individual client information across multiple systems
- reasonable to create a system that would allow for the integration of external data sources
- realistic to think that effective partnerships could be formed to support the necessary collaborations to ensure HIMS included the necessary data

BSS staff were very familiar with external program evaluations and with MIS systems to manage programs. However, they envisioned a different approach that would enable them to bring together the vast information resources available in the homeless services provider community and other state agencies to inform day-to-day program management and support continuous program planning and evaluation. BSS faced many challenges in achieving this objective including engaging stakeholders and securing their collaboration, developing service evaluation models that respond to the varied service populations and programs, testing existing information policy frameworks, and wrestling with technical challenges and serious data quality issues. Each of these is discussed below.

Stakeholders and their motivations

Shelter providers were the key stakeholder group to be engaged. These organizations provide shelter and services to homeless people under the regulatory supervision of BSS. They are structured in several ways. They vary greatly in size, specialization, and scope of service. Some are single-site facilities, others are part of a large corporate nonprofit organization. Some have extensive case management systems, others less sophisticated systems, and still others only manual paper records. Each provider its own data dictionary and naming conventions for specific data elements. And, each has individualized business rules that dictate how work is done and what types of data are collected.

About half of the nonprofit homeless service providers in NYC, Westchester, and Suffolk counties are members of a committee of the shelter providers organization called the Technology Committee. The Technology Committee was formed in 1997 to respond to a new information system for case reporting that was being mandated for use in NYC-based shelters by the NYC Department of Homeless Services (DHS). The Technology Committee strongly opposed that system for several reasons. It was a canned commercial system that was selected by DHS without consulting with the shelter providers. The system did not assist providers in case management, but added a new system and reporting responsibility to their existing operations. The system would collect not only demographic information about clients, but also case notes, the highly personal information that case workers collect for purposes of working with clients on their individual problems and needs. The Committee took its concerns to the leadership of the provider community, which successfully brought pressure on the City agency to abandon the effort. Because other information technology questions and opportunities continued to emerge, shelter providers decided to continue this useful forum to jointly address information technology issues.

Given the experience with the City's case reporting system, BSS staff recognized that the success of any new state system would rest heavily on the extent to which providers supported it. BSS has the authority

to mandate compliance with any program it sponsors, but the staff also understood that to achieve a high quality shared information resource they had to pursue a collaborative approach. Consequently, BSS made significant investments in building relationships and trust in the early stages of the project. Many meetings were held to discuss how this initiative would be different from others. The BSS Director made a personal and organizational commitment to the local government agencies and the provider community that "if they don't see value in the system as a tool to support individual providers as well as the community as a whole, then it won't be built." Despite these assurances, and their commitment to high quality service, the Technology Committee members were very guarded in their early participation.

Despite their skepticism, the providers recognized that HIMS could offer them important benefits. They would be able to assess their own programs against their peers. And, the ability to compare programs and outcomes across the whole system would identify the best performers which would probably signal best practices that everyone could share. Through these discussions, providers began to see how HIMS could benefit them directly. Through meetings, presentations, conference calls, and one-on-one discussions with providers, BSS generated growing trust that information the providers shared with the state would not be used to threaten the well-being of clients or used against specific providers or program managers. In this more trusting atmosphere, the group was able to turn its focus to the practical questions of the design, usability, and value of HIMS.

As part of their commitment to partnership, BSS established a regular series of working sessions in NYC with the Committee. New York City and Westchester and Suffolk county staff participated in these discussions where facilitated sessions, led by outside experts, began drawing out and addressing provider concerns which included policy, data, technology, skills, and cost considerations.

Design & platform for HIMS

The HIMS system was not envisioned to replace or replicate a daily transaction process or the case management systems used within the provider community. Its purpose was to provide a historical view of the impact of service programs on the homeless community.

The design and development of this type of system differs from traditional On-line Transaction Processing systems (OLTP) in that it is not transaction based. It is historical in nature and relies on a design process referred to as On-line Analytical Processing (OLAP). OLAP software allows users to quickly analyze aggregated information into multi-dimensional views or hierarchies. It can answer such questions as: "What is the average age of a client in facility X for time period Y?" "In facility Z, how many 18-21 year old females were "first time" residents during October - December, 1999?"

The proposed integrated data repository would test the feasibility of obtaining data from disparate sources, and accurately matching the data so it could be aggregated and analyzed to evaluate services. Demographic data was needed from the homeless service providers who maintained client information in their case management systems. Payment information came from the State's legacy Welfare Management System (WMS), and facility information was provided from BSS's provider certification database. Ideally, medical information would come from the State Health Department's Medicaid Management Information System (MMIS) and data on substance abuse or other services would come from other state agencies. As shown in Figure 1, the prototype design team set out to integrate some of these data sources into a secure Web-enabled system that could be used by all participants.

Four provider organizations from the Technology Committee (Homes for the Homeless, HELP-USA, the Salvation Army, and NYC Human Resource Administration Office of Domestic Violence & Emergency Intervention) volunteered to provide data needed to develop the prototype. The provider data pertaining to family shelters, the data from the BSS facility file, and individual client data from WMS were used to create the Homeless Information Management System prototype.

Homeless Information Management System (HIMS) Prototype Design

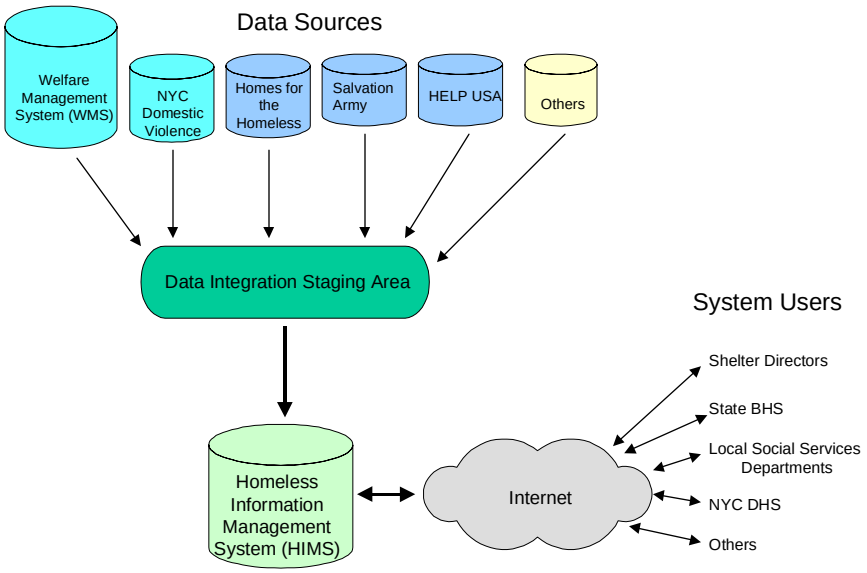

Figure 1. Conceptual Prototype Design

BSS staff approached the choice of technology platform from the perspective of business needs rather than on specific hardware or software preferences. Through facilitated discussion sessions, the project team outlined what capabilities they envisioned this new resource to have. They did not discuss features such as "data warehouse," "SQL Server," or "processing speed of….." Instead, they discussed a number of business process attributes such as: "ability to analyze public assistance and homeless services," "standard template for correspondence," "uniform definition of services," "matching to external files," "ability to do visual analysis and exception reports," and "remote access with appropriate levels of security."

Those capabilities were grouped in a framework of modest, moderate, and elaborate features and functionality. These categories corresponded roughly to the lowest level of functionality that was worth pursuing, to a more robust set of features with more benefits (and costs), to the most extensive system that they could reasonably expect to justify. BSS chose to pursue the moderate level of system functionality, which would meet essential current needs and allow for eventual expansion to include some of the elaborate level features.

Once this was accomplished, the user requirements, business process analysis, and problem definition helped define the selection of the specific technological solution. One aspect of the solution was that the application, housed in Albany, would be accessed by the homeless service provider community via the Internet. By allowing this type of access, the actual platform and training requirements would be reduced—or so the team thought. However, that solution had to be modified as the existing capabilities of the providers were taken into account. In many instances providers either lacked the technological infrastructure (no hardware or limited hardware available within the shelters), or training on how to work within this new environment.

Many had never had access to a PC let alone the Internet. Those who did have PC capabilities often had either no access to the Internet or had policies limiting the access to the Internet. Those providers who were expected to purchase in-house case management systems in the near future were also limited in their knowledge and capabilities to make such a purchase. Few had resources on which to call. Overall

the BSS team found the majority had limited funds, staff, and knowledge on how to access such a system.

BSS staff adopted a developmental strategy to address this situation. They started by working with whatever data was available, usually from the larger nonprofits and the local DSSs, including New York City. Second, they helped the provider community find and encourage software developers to listen to their needs and develop low-cost, easy-to-use, homeless-oriented systems that over time will improve the technological capacities and data resources of the remaining organizations.

Agreeing on how to define and measure performance

A fundamental issue in this case was whether BSS and the homeless shelter providers could agree on a service evaluation model that would satisfy their various assessment needs. In a series of meetings devoted to this question, the group explored the possibility of developing standard service definitions and evaluation measures. They tackled the difficult questions of performance measurement and trying to define, for example, what kind of action, behavior, or outcome constitutes a "success" for a deeply troubled individual compared to a relatively stable family. A simple head count says nothing about these questions. The group began by specifying how HIMS might tell them which services lead to the best outcomes for different categories of clients.

They first focused on simply identifying the various services provided to homeless individuals and families. A list of 66 distinct services was generated. Participants then began to work toward standard definitions of those services considered to be provided by the broadest array of shelters. This exercise demonstrated that it would be possible to develop a manageable, usable set of standard categories and types of services.

Following identification of the services, the group identified the attributes of specific services so that like or comparable services could be compared. Some services were straightforward with few attributes that vary from one place or client type to another; other services had many variables. An attribute such as location of service could be applied to all services and was considered to be a useful factor for comparing the outcomes of similar service programs offered in different locations. But temporary housing, for example, varied considerably across the providers' programs. The discussions identified 26 different forms of temporary housing, described through types of beds, family or single units, or special population characteristics. Temporary housing can be provided through a general shelter bed or in a specialized transitional program bed, in family or single adult facilities, in safe homes, traditional shelters, safe havens, churches or synagogues, emergency housing apartments, hotel rooms, and others. For each attribute, a decision had to be made about how to define and capture the data for future analysis.

After determining that services and their attributes could be standardized, the focus of discussion moved to identifying an explicit set of prioritized outcome goals. The various local government and non-profit providers were surprised to find that their desired outcomes were very similar to each other and to those identified by BSS: a decrease in recidivism, completion of service programs, and self-sufficiency were the most desired outcomes for homeless clients. Several other outcomes were also rated highly including placements in permanent housing, reducing lengths of stay, and increasing sobriety.

Recognizing that professional judgment is critical to the correct interpretation and use of the data, the group then explored factors that are important for interpreting outcome data. Although HIMS would provide access to more robust data than previously available, it would still not tell the whole story. For example, HIMS would allow a user to determine the recidivism rates. However, recidivism rate is not an absolute measure, but must be assessed in the context of each different demographic profiles. The demographic profile of a particular segment of the homeless population might cause them to be less resilient or more likely to be recidivists. For example: a 50% drop in recidivism would be an unrealistic outcome for chronically mentally ill, unemployable single women with one or more children, but should be regularly expected of employable, single women with one or more children.

The group then sought to determine if there was sufficient consistency across their programs and their program implementation and assessment models to move forward with a technology investment. At first glance they appeared to be too different for any standard assessment model to be applicable – however, after many meetings where they talked through their service environments, their implicit and explicit assessment models, and their desired outcomes, they began to see that they did have a core set of services and goals and that sufficient benefit would be achieved in working together to make explicit investments in joint assessment tools.

Information policy concerns

Shelter providers are in the human services business. Their staff interact daily with people who have a variety of personal problems and needs. Many are trained social workers and a strong ethic of client confidentiality pervades the provider community. As a result, one of the first policy challenges was concern from the shelter providers that existing policies would not protect their clients' confidentiality if they shared case management data with BSS. Up to that point the data reported with BSS consisted of quarterly aggregated input and process data rather than outcome data related to programs and specific categories of clients. This concern dominated early discussions.

In this context, several specific concerns emerged. One had to do with unique populations. For the majority of providers, sharing data meant the release and use of client demographics such as name, social security number, age, and address. The Domestic Violence shelter providers had quite different concerns than the rest. Since their clients are in danger of being assaulted or otherwise harmed by people who know them, the most confidential information had not to do with their identity, but with their physical location. Sharing information that linked a particular client to a particular shelter was therefore of great concern to these providers. The group came to understand that different kinds and levels of data security would be necessary to account for these important differences among programs. In this case, all agreed that the facility information and address had to be masked to protect the location of the client.

During the course of the many meetings on this topic, it became clear that the providers were unaware of the stringent requirements and protections already in use by OTDA for other client-oriented systems such as the Welfare Management System. These are based on the New York State Social Services law which requires the agency protect client confidentiality and limit or prohibit the use of data outside the program for which it is collected.

The Director of BSS compiled these documents and sent them to the committee with a cover letter of assurance from the Commissioner of the agency. The material cited specific statutes, regulations, guidelines, and procedures that addressed this threshold concern for providers. The combination of formal documentation with a strong legal basis and the assurance of the agency's top executive allowed the group to move forward to operationalize these policies.

While the issues of confidentiality were being addressed, a new policy concern emerged—How would shared data be used? Shelter providers were concerned that BSS would use the data to publicly measure and report their performance as provider organizations. They wanted the data to be used to assess the impact of specific programs independent of the provider. Here again, history played a role. Around the same time, New York City was developing a family shelter incentive program by which providers could receive a bonus of up to 3% of their budgets if they met specified performance goals. The providers feared that the goals would be unrealistically high and result in negative perceptions of their programs. They carried this concern over to the discussions of HIMS. The BSS staff did not agree to ignore provider-specific information. Instead they pointed out that their existing inspection process already collects the same information, so the risk was no greater with HIMS. The providers acquiesced with the understanding that they would test this concern as the system was built.

Data quality and fitness for use

As the providers began sharing their data for use in the prototype, data quality issues quickly became a concern. Data quality problems have many causes. The most obvious problem is a data entry error. This is typically addressed through internal data entry procedures and audit checks. BSS and its partners, however, faced much more complex and less tractable problems.

One common source of data errors is the stressful situation of the client at the point of entry to a shelter. The decision to go to a shelter is frequently a last resort for a client. The primary concern for a domestic violence client is to stay hidden from an abuser. Some clients have severe mental health or substance abuse problems that make it impossible for them to provide needed information. In some cases, clients may deliberately provide false information in order to protect their anonymity. More commonly, the stress associated with the situation causes clients to forget or have no record of dates, social security numbers, and past histories. Thus the information provided to the case manager at intake can be fraught with gaps and errors. Case managers may choose not to collect all required information in one session. Several providers said they may take up to two or three weeks to complete all the basic information in a client record. In many cases data for a client remains incomplete in some respects.

In an ideal situation, the providers would have the capacity to match client data against a master system, such as the Welfare Management System or New York City's Human Resource Administration system, to verify or complete missing data. However, that was not feasible in this case. Clients are entered under various names, names are misspelled in multiple ways, social security numbers (when used) are often incorrect, and family composition can vary from one date to the next.

For example, one household record contained a female client with two dependent children when they entered the shelter. The next day they moved to another facility and family composition was recorded as a female head of household and three children. Was this a data entry error or was it a correct reflection of the state of this household? As it turned out, both entries were correct. One child was living with a grandparent and reunited with her family on the second day. This is a fairly common pattern. Children are sometimes placed in foster care or with family members while temporary housing is found. Once placement has occurred, the children are reunited with their families.

Data quality tools have limited capacity to address such unstructured quality issues. Data quality tools focus on auditing, migrating, or cleansing the data based on pre-designed business rules (e.g. Name = alpha field, 12 characters in length, value = text string). In HIMS, every record in the prototype was reviewed by the design team and discussed at length with the data provider so key concerns and specific errors could be addressed. Some errors were easily corrected while others needed to be researched with program managers or data technicians.

In the end, all provider data sets were scrutinized, error reports were generated, and data inclusion and exclusion rules were developed. Some data gaps were filled by sending BSS staff into the field to read case records and record the missing elements. Some of the data was "cleaned" with review or filter programs, but this was a minor part of the effort. The human intervention was essential and time consuming—and it required extensive knowledge of the complex program environment. This costly process was feasible only because the prototype data sets were very small. In a fully operational system, much more standardization among the data sources will be required.

Data in context and the importance of expert knowledge

The design team was intentionally made up of program as well as technology experts. Each brought a unique and valuable perspective to the project. As described earlier, it was imperative to have both kinds of professionals involved to assist in the definition of the new system as it related to data, policy, and technology decisions. This joint capacity allowed the team to make decisions quickly and detect gaps in the decision-making process. The provider community's practical perspective gave the team the ability to address operational and policy issues as soon as they were identified. Information management skills such as the identification of data sources, data collection, data security issues, data repository methodologies, and quality control techniques were all necessary in developing HIMS.

At different phases of the project, different skills were needed, and different team members were added to the mix. As technologists or business analysts were needed they were brought in. As their roles were completed, their activity diminished. The one consistent facet of the project team that never changed was the involvement of the BSS staff. They provided the managerial as well as the program focus of the team. Each staff member, acting as liaison to the provider community, could address the goals and the challenges of the project. This provided the continuous, consistent communication that was so important in building and maintaining trust with the providers.

Harmonizing data definitions and the value of meta data

The challenge was not only in obtaining the data but also in finding commonality among the data elements as they were used by the different organizations. The design team needed to understand how the data was collected, what similarities existed among the data sources, and how the data was going to be aggregated in the new system and to document all this with meta data. This required business rules and standards for the new integrated system as it related to the questions the new system hoped to address. While this seemed easy at the beginning, the true complexity emerged as the team wrestled with such seemingly simple terms as 'age' and 'ethnicity.'

The challenge is illustrated well by the process of deciding how a client's age would be calculated. The decision did not lie with conventional data definitions—in most transactional systems a person's age would be calculated based on their birth date and system date. However in this system, age would need to be based on business rules for how a client would be profiled; would age be based on date of entry, on age when referred to a service, or age when a service is rendered or completed? Each decision had to be based on how the data was going to be used in the aggregated form and what questions the system would be used to answer.

This effort involved more than the technical staff. Each question had to be considered from the program and business perspectives as well as the technical perspective. Because the system was being created as a data mart, the transformation (aggregation) of the data was a crucial factor. The lowest level of aggregation for each data field had to be considered. For example, if age were aggregated by ten-year groupings (e.g. birth-10 yrs old, 10-20 yrs old, 20-30 yrs old) the users would not be able to a profile clients who are 18 years of age, an important milestone birthday for many public programs.

For each data element, the team had to agree on common definitions and consider how these definitions would affect the inclusion or exclusion of data elements into the integrated system. Each decision made needed to be revisited with each additional data source. These questions helped define the business rules that shaped the system. While they were easily addressed from either a data management or a technology perspective, the more global policy perspective was both more difficult and more important. It not only provided the policy framework for the entire system, but also assured that the system would provide data that would support informed decision making.

In another example, the HIMS design rested on the ability to establish a unique identifier for every record within the integrated system. Unfortunately, but not surprisingly, each provider assigns a different unique identifier to a client. Originally, the design called for the use of a client's social security number (SSN) as the identifying code. SSN is recorded in some database systems, but some providers never ask for this information.

Certain providers know that even if their clients have a social security number, it is unlikely that they would recall it during the intake process. Some providers use an identification number assigned by NYC DHS, while others assign a system-specific identification number. The WMS legacy system assigns a Client Identification Number (CIN) at intake but also collects SSNs and Human Resources Administration (HRA) Case Numbers. When domestic violence is an issue, the social security number is recorded but may be changed to protect the client. The domestic violence database maintains no cross-reference system so the social security number cannot be used to match clients in the domestic violence database with any other database.

The design team had the task of developing a common identifier for the prototype or a specific procedure so that data could be cross-referenced with WMS and across provider systems. The design team found that each record contained either the social security number or the HRA number for the head of household. Once this was discovered, the team was able to match either number in the WMS system to obtain a client's CIN number. The CIN number can also be used in the future to match to other legacy systems so that medical assistance and public assistance information can be obtained. To do this, a matching program had to be written for each system feeding the prototype. While labor intensive, it could be reused by each provider once the initial program was generated.

Data usability for different purposes and users

Even though the design team had cleansed the data, created a consistent structure and format for use across all systems, created a data dictionary that explained the type of data that was to be brought from each provider's database for each field, and had access to all of the original data, it was still difficult to completely understand the data and its potential use. In addition to the data consistency issues discussed above, each data code for the integrated system needed to be reviewed in the context of related programmatic issues.

In some cases, data was collected based on unique policies or business rules specific to the provider. For example, each system contained information regarding a client's ethnicity. Usually ethnicity had five categories, while in one instance the provider registered 12 different categories. Conventional wisdom would collapse the 12 into five generic codes and then the five codes would be used in the integrated system. However, the 12 ethnicity categories were extremely important to that particular provider because they are tied to federal regulations and funding requirements for its programs. Therefore, the system design needed to incorporate a translation table that would feed data to HIMS, while retaining the ability to provide data back to the provider without losing the provider's expanded categories.

Transforming data into information

As each data set was added to HIMS, the human intervention required to review each data element was multiplied. The project team, comprising 10 people, spent seven months in what is called the discovery phase of system development. The discovery phase is characterized by many discussions of the specific business processes the system is designed to support and the business rules it will follow.

The actual prototype development, the technology component of the project, took two months time for two developers, a majority of which was spent on data transportation and transformation rather than creation of the actual application.

Both BSS and provider staff, along with expert consultants reviewed data sets to address:

- data quality issues
- data transportation issues (moving the data sets to a staging area in preparation for transformation)
- data transformation issues (changing data sets based on new business rules governing the new system)
- data exportation issues (moving the new aggregated data sets from the staging area to the new system)

This process absorbed an enormous amount of project staff and consultant time. As discussed above, the contextual knowledge required during this phase was imperative to ensure correct business decisions were made. Substantial time was spent in drafting new business rules and making data inclusion decisions—time seldom considered in the cost estimation models.

Discussion

The HIMS prototype represented a new information resource for state, local, and nonprofit program managers. In the process of designing it, representatives from all three groups came to have a much more detailed and nuanced understanding of their clients, programs, processes, and relationships. The

system allowed them to answer questions about performance that could not be answered before. They developed a deep appreciation for the power of good information to help them design and operate effective programs and a similar appreciation for how difficult and time-consuming it is to assure that the data quality was sufficient for their purposes.

In the end, HIMS did not progress beyond the prototype stage for lack of funding. Without the resources to complete the data acquisition strategy, train and equip all users, and refine the user interface, it did not go into operation. Its more lasting legacy was a much more effective set of relationships among the participating organizations and new-found appreciation for how information and technology could support their work.

Nevertheless, systems like HIMS can fill an important gap in performance information. Because they are the result of deep involvement by program managers and staff, they are much more likely to be used to make decisions about processes and performance. They allow mangers and line staff to assess their own programs frequently (or continuously) and to act on the data with confidence in its quality and relevance. As shown in Figure 2, self-assessment systems offer great opportunity for continuous intelligence about program performance and frequent opportunities to make adjustments in design or operations. In the time it would take for the typical formal evaluation to unfold, such a system could help managers make a series of assessments and related improvements. Moreover, because program managers have invested in the creation of the system and its data, they will be likely to have far more confidence in it than they typically have in external evaluation results.

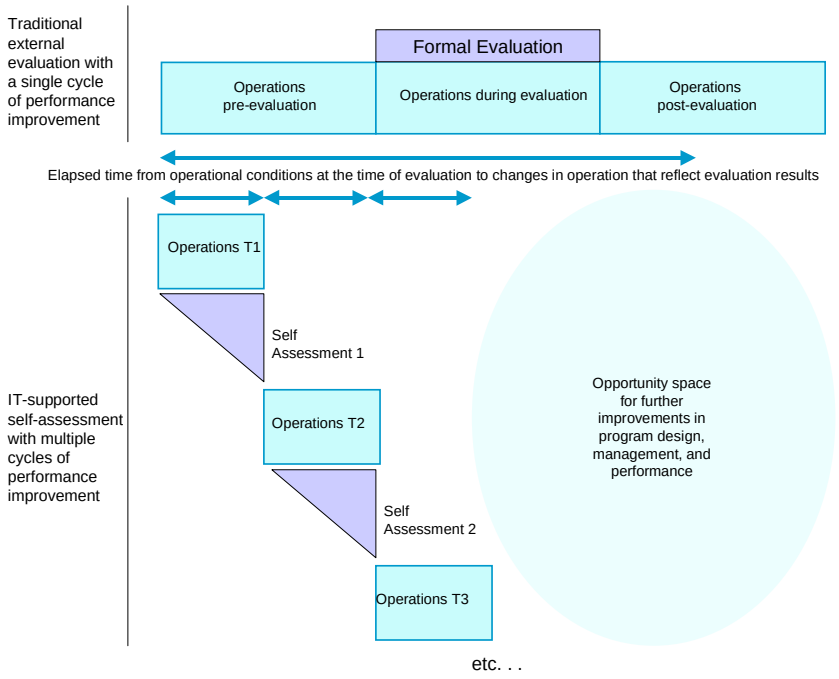

Figure 2. Traditional evaluation vs. self-assessment systems

However, such systems are not easy or inexpensive to create and should therefore be approached with a clear understanding of their costs, benefits, and limitations. Traditionally, estimating the cost of any new system takes into consideration the initial design and start-up costs. The development costs (such as hardware, software, and consulting) and the production system costs (such as hardware, software, and dedicated technical staff) are readily quantifiable. Unfortunately, few consider the cost of the program staff time, especially the time required in the early stages of problem definition and relationship building. These are often hidden costs that do not go into the investment calculations, yet they are essential,

substantial, and continuous. Getting early relationships into a more trusting mode and constantly reinforcing these relationships consume large amounts of time and managerial attention. In the case of HIMS, regular large group working sessions, weekly status meetings, reaching out to potential data providers in other agencies, and building working relationships within its home department were all costly, but essential project activities for the BSS managers.

In addition, as the case amply demonstrates, choosing relevant performance measures and selecting data to support them demands expert knowledge sharing and putting data in context. This requires careful process analysis, harmonizing data definitions across different programs and organizations, improving poor quality data so it can be used with confidence, and often acquiring or sharpening analytical skills.

While agencies are often able to obtain appropriations for the technology costs of new systems, these staff and data management costs must often be embedded in regular operations. Therefore they must generate operational benefits as well as evaluation results. This is where self-assessment systems can have a distinct advantage over traditional evaluation approaches. They are designed by program managers for the purpose of improving program management. As such, they have a legitimate and compelling claim on the time and resources of the people who build and use them.

Despite its benefits, this approach has limitations. Self-assessment systems are beneficial for ongoing monitoring and frequent adjustments to the factors that managers have some discretion to change. However, because they are internal, they are not likely to be effective for fostering major, highly visible changes in program policies or design. Those kinds of changes need the financial and political resources that come from legislative or public support -- they demand the political credibility that internal actions cannot command. For this reason, traditional external evaluation approaches with their public accountability benefits are also needed. Self-assessment systems should take a place along side formal external evaluation methods as a complementary investment in better government performance.

References

Alter, S. (1980). Decision Support Systems – An Application Perspective. Addison Wesley.

Carter, Reginald. (1994). Maximizing the use of evaluation results, in Wholey, J.S., Hatry, H.P., and Newcomer, K.E., eds. (1994). The handbook of practical program evaluation. San Francisco: Jossey-Bass.

Caudle, S. (1994). Using qualitative approches, in Wholey, J.S., Hatry, H.P., and Newcomer, K.E., eds. (1994). *The handbook of practical program evaluation.* San Francisco: Jossey-Bass.

Cloete, F. (2003). Assessing governance with electronic policy management tools. *Public Performance and Management Review,* 26(3): 276-290.

Coplin, W.D., Merget, A.E., and Bourdeaux, C. (2002). The professional researchers as change agent in the government-performance movement. *Public Administration Review*, 62(6): 699-711.

De Lancer Julnes, P. D, and Holzer, M. (2001). Promoting the utilization of performance measures in public organizations: an empirical study of factors affecting adoption and implementation. *Public Administration Review* 61(6): 693-709.

Drucker, P. (1973). *Management: tasks, responsibilities, practices.* New York: Harper & Row.

Lee, S.M., and Eom, H.B. (1990). "Multiple Criteria Decision Support Systems: The Powerful Tool for Attacking Complex, Unstructured Decisions," *Systems Practice* (3:1) pp 51-65.

Flynn, P., Curran, K., and Lunne, T. (2002). A decision support system for telecommunications. *International Journal of Network Management*. 12:69-80.

Goddard, Steve, Sherri K. Harms, Stephen E. Reichenbach, Tsegaye Tadesse, and William J. Waltman. (2003). Geospatial Decision Support for Drought Risk Management. *Communications of the ACM* 46(1): 35-37.

Grifel, S. (1994). Organizational culture: its importance in performance measurement. *Public Management* 76(9): 519-23.

Keen, P. and Scott Morton, M.S. (1978) *Decision Support Systems: An Organizational Perspective*. Addison Wesley.

Liang, L.Y. and Miranda, R. (2001) Dashboards and Scorecards: Executive Information systems for the Public Sector. *Government Finance Review*.

McGowan, R.P. and Lombardo, G.A. (1986). Decision Support Systems in State Government. *Public Administration Review*.

Melkers, J. and Willoughby, K. (1998). The state of the states: performance-based budgeting requirements in 47 out of 50. *Public Administration Review* 58(1): 66-73.

Newcomber, K. E. (1998). Evaluating public programs, in Perry J.L., ed. *Handbook of Public Administration (2nd edition)*. San Francisco: Jossey-Bass.

Nyhan, R.C. and Martin, L.L. (1999) Comparative Performance Measurement: A Primer on Data Envelopment Analysis. *Public Performance and Management Review* 22(3): 348-364.

Osborne, D. and Gaebler, T.. (1992). *Reinventing government: how the entrepreneurial spirit is transforming the public sector.* Reading, MA: Addison-Wesley.

Patton, M.Q. (1986). *Utilization-focused evaluation*. (2nd edition). Newbury Park, CA: Sage.

Theurer, J. (1998). Seven pitfalls to avoid when establishing performance measures. *Public Management* 80(7): 21-24.

Thompson, J. R. (2000). The dual potentialities of performance measurement: the case of the Social Security Administration. *Public Productivity and Management Review*, 23(3): 267-281.

Wang, X and Gianakis, G. (1999). Public officials' attitudes towards subjective performance measures. Public Productivity and Management Review 22(4): 537-53.

Wholey, J. (1999). Performance-based management: responding to the challenges. *Public Productivity and Management Review* 22(4): 299-307.

Wholey, J.S., Hatry, H.P., and Newcomer, K.E., eds. (1994). *The handbook of practical program evaluation.* San Francisco: Jossey-Bass.

Williams, F. P., McShane, M.D. and Sechrest, D. (1994). Barriers to effective performance review: the seduction of raw data. *Public Administration Review,* 54(6): 537-542.