

The Challenges of Measuring Employment Program Performance

William S. Borden

Mathematica Policy Research

November 2009

Prepared for the November 2009 conference

What the European Social Fund Can Learn from the WIA Experience

The Challenges of Measuring Employment Program Performance

Both the Workforce Investment Act (WIA) reauthorization process and the planning efforts of the European Social Fund (ESF) would benefit from a review of the recent experiences of performance management of employment training programs in the United States. This paper presents an operational perspective on how performance systems are designed and implemented. It also discusses the challenges to effective performance management—challenges that are little known except to the state and federal staff managing the performance systems, and that are often not clearly understood. There is very little that is easy and straightforward about measuring program performance. Seemingly simple concepts such as enrollment, exit, employment, earnings, and who to count in a measure and when to count must be defined very precisely for performance results to have meaning. This paper assumes that the reader is familiar with WIA and the WIA performance measures.

The design and implementation of effective performance management involve many conceptual and operational issues. This analysis tries to briefly touch on many of them to illustrate how involved the process is and to alert program managers to the areas that they need to address. Each of these issues requires more extensive discussion than the scope of this paper allows. Performance management raises interesting and significant questions about organizational and human motivation, the dynamics of state-federal political power sharing, and the management of government programs. Policymakers tend to underestimate the challenges they face and sometimes lack the commitment necessary to make performance management processes as effective as they should be. The Employment & Training Administration (ETA) has corrected some of the problems that occurred early on, but there are still some operational aspects that need improvement. The pursuit of effective and fair performance management inevitably encounters challenges for which there are no easy solutions.

Research on employment and training programs focuses primarily on evaluations of the impact of public investment in job training services. But there are other factors to consider when analyzing the WIA performance management system. There is a clear dichotomy between the program management objectives of performance management and its evaluative objectives. This analysis argues that some form of performance tracking and data validation is necessary for managing a complex national system of job training programs even if the outcome data were not used to determine funding. Despite the great value of effective performance management, there are limits to using performance management data to drive funding decisions.

It is also important to look beyond WIA and take a comprehensive approach to assessing performance management of job training services by examining the programs that serve special

populations. Policymakers need to consider how to provide efficient and effective service to everyone, but especially people with disabilities, veterans, youth, and older workers, since the costs to serve them greatly exceed those of serving job seekers in general. This broader perspective also helps inform the debate about consolidating services under a universal program like WIA and provides the most useful information for the European Commission as it looks at performance management and service delivery alternatives. Choices must be made about whether to manage services under a more unified governance structure or as independent governance structures. In the United States, there is a somewhat confusing mix of approaches, with WIA and the Employment Service at the core and considerable fragmentation and overlap beyond that.

This analysis will draw broadly on lessons learned from implementing performance measurement systems for WIA, the Employment Service, the Senior Community Service Employment Program (SCSEP), and the Vocational Rehabilitation (VR) program at the Department of Education, among others.¹

Section I presents a conceptual framework for analyzing performance management issues. This includes discussion of the goals of performance systems, the limitations on measuring government program performance, and how measures are designed and defined. These concepts form the building blocks for designing a performance system.

Section II discusses the distinction between using informal processes to manage performance and effective performance management. It covers the importance of implementing rigorous standardization, validation and monitoring processes for effective performance management and how ETA has made great progress in this area despite continuing problems.

Section III discusses the challenges and benefits of involving stakeholders in the design and implementation of the performance measures. This section discusses the problems that occur when stakeholders are more concerned about meeting their goals than improving their results, as well as their somewhat exaggerated reaction to the burdens imposed by performance systems.

¹ The Performance Management Group at Mathematica Policy Research has been involved in designing and implementing performance management and data validation systems for WIA, the Trade Adjustment Assistance Act, the Labor Exchange (or Employment Service), the National Farmworker Jobs Program, the Senior Community Service Employment Program, the Unemployment Insurance Program and the Vocational Rehabilitation Program. The Group also works on Temporary Assistance for Needy Families (TANF) and Supplemental Nutrition Assistance Program (SNAP) performance reporting and on assessing performance for Medicaid and Education Department grants.

Section IV discusses key aspects of the WIA performance management experience to date, including how the measures have evolved and the use of wage records to measure outcomes.

I. A CONCEPTUAL FRAMEWORK FOR ANALYZING PERFORMANCE MANAGEMENT ISSUES

A. Performance Management Versus Program Evaluation

As Barnow and Smith (2004) point out, program evaluation and performance management derive from different sources and motives and have deeply committed adherents.² This analysis takes the position that managing very large-scale and far-flung programs involving thousands of staff, millions of customers, and billions of dollars requires comprehensive management information systems. In other words, tracking and measuring customer flow, services, and outcomes is inherently desirable and even necessary to managing any modern organization. Therefore, the question is not whether we should track customer flow and services and measure performance, but whether and how we should use the data to determine funding, incentives, and sanctions.

Some in the evaluation community argue that there are risks in drawing conclusions from administrative performance data and this concern is supported by a detailed understanding of data quality and measurement validity issues.³ ETA's experience in implementing performance management systems over the 10 years since the passage of WIA has shown that it is difficult to measure performance well, and that using inaccurate performance data to drive policy and incentives leads to misallocated resources. Putting more emphasis on using results to reward and sanction states than on ensuring that the results are valid and meaningful also leads to understandable yet often undesirable behavior by program operators.

Performance management systems and research evaluation methods both have their strengths and weaknesses. Performance data are much more efficient, systematic, comprehensive (they are produced for all customers), and timely, but they are somewhat crude and imprecise tools for measuring program impacts. Effective performance management systems, however, are essential to good evaluation, particularly since performance management is the main reason that reliable data are available on programs. Some research efforts are abandoned because of incomplete and flawed data sets, while other research projects draw erroneous conclusions because of bad administrative data. There is an increasing tendency to leverage the efficiency of analyzing administrative data versus more expensive experimental designs. In fact, both are needed. Even selecting stratified samples of participants for randomization requires clean and complete participant data sets with accurate enrollment and exit dates and customer characteristics.

² Barnow, G., and Jeffrey Smith. "Performance Management of U.S. Job Training Programs: Lessons from the Job Training Partnership Act." *Public Finance and Management*, vol. 4, no. 3, 2004, pp. 247-287.

³ Ibid.

B. Underlying Premises of Performance Management

First, we need to define precisely what the goals of a government program performance management system are and what constitutes a performance measure. We must also examine the motives and roles of the various actors in such politically and technically complex systems.

Two premises underlie the increasing emphasis on accountability in government performance. The first is that public funds must be spent wisely and produce a return on taxpayer investment. The second is that measuring the effectiveness of a business process is critical to managing a modern organization.

Behind the first premise is the implicit assumption that government agencies must be under pressure to perform just as in private enterprise—where profit and loss determine success, rewards, and even survival. This underlying “Darwinian” notion that competition is good and that programs should demonstrate results to justify their existence is accepted by both major political parties. The Bush administration greatly advanced this approach to federal management and reflected the program management approaches used in Texas and Florida, the most advanced states in using performance outcomes to drive funding decisions.

But the notion that measuring the performance of a government program can substitute for the competitive pressures of the marketplace has many limitations. In the extreme, this idea takes the form of performance-based budgeting, where funding is directly correlated to performance, and programs can be totally defunded based on measured performance outcomes. In theory this makes sense, if there are valid measures and accurate performance data. The reality, however, is that measures frequently do not accurately reflect underlying program performance, and even more often the data are inaccurate and inconsistent across operational entities. Accounting for all the factors affecting WIA performance is impossible to do with great precision. We must control for variations both in the type of barriers to employment in the population served and in the employment opportunities available in an area, and then we must implement rigorous data validation methods. Without such steps, cutting budgets or defunding programs or operational entities based on program performance would be irrational and unfair. Such extreme approaches to using performance outcomes would also encourage program operators to engage in *creaming*: developing too-easily-reached goals and under serving the target population by focusing on those most likely to be deemed a success, instead of serving those most in need of services yet less likely to succeed.

C. Monopolies, Competition and Privatization

There is another flaw in the application of the competitive approach: some programs enjoy natural monopolies. For example, one might conclude that if Ohio's program was ineffective and Michigan's effective, people in Ohio should seek services from Michigan or Michigan staff should replace Ohio staff. Obviously, the Darwinian "perform or die" theory breaks down in this application. So if we cannot put the Ohio program out of business, how do we act on our performance data? Do we provide increased funding to Michigan as a reward (presumably not needed), or do we cut Ohio's funding as a penalty (and probably damage their performance further)? We are left with the industrial quality control concept that we have used performance management to identify superior and unacceptable performance. Using the performance information, we can now intervene to provide technical assistance to Ohio and transplant best practices and methods from Michigan to Ohio. Finally, we must continue to track Ohio's improvement until their performance becomes acceptable. This is an appropriate application of performance management in a government setting.

There are two situations in which the monopoly problem does not interfere with applying market forces to government performance management. First, programs that provide similar services to the same population can compete against each other. Since there are many overlapping job training programs, this is possible and indeed has been discussed by the Office of Management and Budget (OMB), as we shall soon see. Every Bush administration budget since 2000 contained no funding for the National Farmworker Jobs Program (NFJP) and maintained that WIA should be the vehicle to provide these services.⁴ All the programs for special populations are mandatory one-stop partners, but there is a wide diversity in the integration of these programs into the one-stop setting. Most operators of programs for hard-to-serve populations believe that their clients would not be well served by WIA. This sentiment derives partly from the instinct for self-preservation and partly from the common belief among social workers that the population they serve is unique and cannot be well served by a more general program. It is true that a one-stop operator might choose to prioritize services to customers that are more likely to gain employment unless counter-balancing incentives are built into the system. The Vocational Rehabilitation program, managed by the Department of Education, operates very differently from the labor programs and has not adopted common employment measures or data validation activities.

Second, services can be privatized and the public entity defunded. Some states have done this with large programs, and it is a common feature at the local level.⁵ So the ultimate application of market principles is to privatize services and make all provider payments contingent on performance. This was

⁴ Congress always restored NFJP funding.

⁵ WIA requires the local boards to contract the operation of the one-stop centers, although public entities often hold the contracts.

used to some degree with performance-based contracting of training providers and is a feature of the Pathways to Work program in the United Kingdom. However, this approach puts tremendous pressure on the providers to manipulate their performance rates and puts unrealistic expectations on the ability of the oversight agency to act on high-quality data that fairly measure performance. If effective performance management methods are used, performance-based budgeting would be an effective incentive in some settings.

The second premise underlying performance management systems is that measuring the effectiveness of a business process (job training and placement services) is critical to managing a modern organization. Performance management techniques derive from industrial quality control techniques that measure the rate of defects in an industrial process (as popularized by W.E. Deming). The quality movement is based on the notion that processes that are measured work better than processes that are not measured.⁶ Performance data are a vital tool for program managers at all levels to identify successful processes and methods, determine what works, share best practices, identify areas in need of improvement through technical assistance, forecast future customer flows and costs, and so forth. It is common sense that program managers at all levels should have data on customer flow, services, and outcomes. The complexity and cost associated with collecting and analyzing high-quality program data, however, leads too many state and federal officials to avoid these challenges and instead put minimal effort into performance data.

Both objectives of prudent public investment and improved service provisions have implications for understanding the potential limitations of measuring the performance of government programs. We can place performance management objectives on a continuum ranging from tracking and performance data solely for better management to the other extreme of basing all funding decisions solely on performance outcomes. Finding the middle ground is appropriate. Although there are limits to how much a performance management system can tell decision-makers about program costs and benefits, there are compelling reasons to track enrollments, services, and outcomes carefully. Doing so provides valuable information to managers at all levels of the system, from Congress and OMB through the federal agency, and to the state and the local area or grantee. In large, diverse systems like WIA and the other programs under discussion, the forces of fragmentation and inconsistent data are so great that only a very strong and standardized performance management system can overcome or at least neutralize them. We can thus see

⁶ Blalock, Ann B., and Burt S. Barnow. "Is the New Obsession with Performance Management Masking the Truth About Social Programs?" In *Quicker, Better, Cheaper? Managing Performance in American Government*, edited by Dall Forsythe. Albany, NY: Rockefeller Institute Press, 2001, pp. 487-519.

that it is more useful to think of performance data primarily as a management tool and secondarily as an evaluation tool.

D. Defining a Performance Measure

To serve the second premise—identifying relatively good or bad performance and measuring improved or decreased performance—a measure must produce a rate of success and not simply a count of activities. Thus a measure can be used to distinguish better and worse performance in meeting program objectives of a single operating entity (one-stop, Workforce Investment Board (WIB), state, program) over time, and also compare performance between operating entities at all levels. Standards that identify minimally acceptable performance must be associated with measures. Failure to meet these standards would trigger remedial steps, such as technical assistance, and even punitive actions such as sanctions. Standards for superior performance could trigger rewards such as incentives and documentation of best practices. A performance measure that does not produce a rate of success cannot accomplish these essential functions.⁷

Programs should set standards for minimum acceptable performance by analyzing the range of outcomes across reporting entities (states, in the case of WIA). One simple axiom is that the minimum acceptable level is what 85% of states achieve; the theory being that if the bulk of states can achieve this performance, then it is a practical goal, and the trailing 15% should strive to improve. For example, the UI program sets performance goals based on the actual distribution of state performance rates. Another approach is to set different goals for different sets of customers based on their barriers to employment. Separate performance calculations should be produced anyway for significant customer groups such as low literacy, people with disabilities and those with poor work history. As we will discuss later, however, absolute performance outcomes should be adjusted to account for differences in customers and labor markets. This approach, however, makes data validation even more essential as program operators have incentives to exaggerate their customers barriers. ETA has adopted a negotiation approach to setting standards and has moved away from using national performance means and standardized adjustment mechanisms.

E. Measuring Processes , Outputs and Outcomes

⁷ Counts are sometimes used as performance measures, for example, if there is no real process to measure, such as program outcomes after a spell of services, or if the designer of the measures just wants to demonstrate results by adding up events that are considered to have social value. In these situations, the counts are not really performance measures in a technical sense and should not be confused with actual performance data. Such counts can be converted to rates if they reflect underlying performance and not increases in funding.

Another aspect of performance management is whether to measure processes or outcomes. Process measures are very indirect and are therefore usually unreliable for assessing actual performance. Process measurement operates on the assumption that adhering to good processes will produce a better result, but execution of the process can be highly variable. Many of the process measurement approaches popular two decades ago have resulted in a “checklist” approach: items are “checked off” when manuals are written or staff are given various responsibilities. These continuous-improvement approaches are good means to an end, but they cannot substitute for measuring actual program results.

Measuring intermediate outcomes, sometimes referred to as program outputs can be useful and can resemble process measures. For example, measuring program attendance, grade advancement, test scores, customer satisfaction and the timeliness and quality of customer services provides prompt feedback to program managers and helps predict actual outcomes. These intermediate outcomes or program outputs should be secondary to actual program outcomes (such as long-term employment and earnings).

Another school of thought focuses on societal rather than individual outcomes. This approach would use poverty levels and measures of community well-being to evaluate program effectiveness. This is a laudable objective and should be a component of an overall evaluation strategy, but it does not fit well within the performance management paradigm. Performance management relies on reasonably direct feedback to program operators at all levels about the effectiveness of service delivery strategies on customer outcomes in order to improve management decisions. Societal outcomes result too broad a range of sources to provide direct feedback to management, but they should be taken into account when determining long-term policy direction.

F. The Impact of Performance Management on Customer Selection

The general intent of the programs is to channel scarce funds to those who need services most because they have the most or the highest barriers to employment. Performance outcomes, however, are based on success, which is least likely for those with the highest barriers. How does the program operator respond to this dilemma? Clearly it would be rational to choose to serve the people most likely to succeed. This could be considered a socially useful impact of performance management on program behavior if the operator is selecting between two people to serve: one with barriers but whose success is feasible and one with more barriers whose success is unlikely. Society may benefit more if the operator chooses to serve the person with the highest chance of success. In the Vocational Rehabilitation (VR) program for example, states are instructed not to serve people considered “too significantly disabled” to become employed. If the choice is between someone with barriers where success is feasible and someone

with relatively few barriers where the service would not be a significant factor in employment, then society would not benefit from the incentive to serve the least-needy customers.

The correct means to rectify this potentially bad incentive is to adjust performance outcomes to provide more credit for achieving success with a customer with higher barriers. This then leads us back to the conundrum that measuring barriers can be subjective and unreliable. It also raises the question of how we perform the adjustment. Computing performance separately for different classes of customers based on barriers provides the clearest information to program operators. Or we can adjust performance after the fact, based on regression models. Either approach, if done well, would produce the same results, though there are other adjustment factors to consider.

G. Adjusting Performance Outcomes

An effective performance management system must produce objective and systematic results. The system must account for the high degree of variability in both customers served and in labor markets. So the system must adjust performance results to provide credit for serving those with the most barriers and for variations in labor market conditions across geographic areas. Here some type of objective regression model is necessary. In 1998, ETA decided to abandon the Job Training Partnership Act (JTPA) regression model, whose complexity had made it unpopular at both the state and the federal level, in favor of a negotiation process. Negotiation provides flexibility, but it does not allow for systematic and consistent performance goals across states. Instead, the outcome of the process is a function more of the toughness of the negotiator than of a method for developing consistent, reasonable performance goals. In addition, states could also try to manipulate the negotiation process by using various approaches to developing JTPA baseline data that would produce the lowest performance, thus ensuring that they could easily meet WIA improvement targets. The weakness of the negotiation approach was illustrated when ETA chose to train regional staff on negotiation skills rather than on how to interpret state baseline performance estimates so that initial WIA goals were set more accurately.

II. EFFECTIVE PERFORMANCE MANAGEMENT AND ITS COSTS

A. Effective Performance Management Methods

There is a significant difference between collecting and calculating performance data using informal methods and using formal performance management methods to ensure that performance results are meaningful and usable. Federal performance management processes should emulate management information systems approaches used to manage large business enterprises. The software development industry has developed formal methods to ensure that systems function properly from the original source

of the data to the distribution of results to end users. Rarely are federal performance systems designed with understanding of the risks to data quality and the methods needed to overcome them. Each time a system is set up the same long process of finding out what does not work takes place over several years before usable performance data is obtained. Sixteen years after the passage of the Government Performance Results Act (GPRA), the technical state of federal performance management is still dismal.

The contrast between ineffective and effective performance management can best be illustrated by the Senior Community Service Employment Program (SCSEP). Attempts to draw samples for customer satisfaction surveys for SCSEP failed because there were few usable case management data on who the customers were. This, and other deficiencies in the performance reporting system, led the SCSEP program to develop a sophisticated national case management system. The SCSEP system contains real-time data on all customers, services, and outcomes and robust analytical and reporting functions. The availability of these detailed, individual-level performance management data enabled the SCSEP program to smoothly transfer 12,000 participants (as a result of the national grantee competition of 2006), to develop timely and comprehensive management reports for all levels of the system, and to report on participants funded by the American Recovery and Reinvestment Act in real time with virtually no additional effort. Prior to this investment, most SCSEP performance data reports were aggregated essentially by hand at local offices and then re-aggregated at the state level before being submitted to ETA. There was no audit trail and thus no way to determine or support the reliability of the data. Through the use of thorough and effective performance management methods, SCSEP has gone from having few reliable performance data at all to having among the best data of all federal programs.

However, this drastic progress in SCSEP (and to a lesser extent in other ETA programs using less extensive methods) has come with significant expense at the federal level. Federal managers at all levels find it difficult to justify the costs of high quality data systems for several reasons. First, they view performance management narrowly as a reporting function and not a performance enhancement process. Second, they tend to focus only on the costs to the federal agency budget rather than the overall cost of the program to the taxpayers. This is a funding allocation issue that Congress should address. Finally, it is hard to convey the complex technical risks and complexities of collecting data from such a large and highly fragmented system, where there are incentives to interpret data rules in such a way as to optimize program performance outcomes. Investments in standardized data processing technology are the only means to develop high quality data sets and result in considerable cost savings overall.

Performance management is a much simpler and more efficient process for federally run programs than for state-run programs because there are fewer operational entities. Therefore, there are

considerable costs involved in allowing states to administer their own programs- even in some cases make their own rules (such as UI, Medicaid, and SNAP)- and in trying to achieve usable national data. It is possible to map or translate state variations to a federal data template to make the data more consistent but this requires significant effort and expense. Given the high turnover in state staff, getting all states to understand and operate using consistent data rules is a never-ending task. This task would seem to be even more challenging for Europe.

The large number of Congressional committees that have control of sources of employment program funding cause the overall system to be highly fragmented, with considerable overlapping services and more costly reporting processes. The “one-stop” system is intended to be seamless to the customer but certainly not to the program managers, accountants, or performance and reporting staff. The fragmented funding streams result in higher implementation costs because one-stops have to collect data to report to many programs and agencies with varying and even conflicting definitions of customer characteristics (such as multiple definitions of *veteran*). That is why it is necessary to step back and look at the whole range of programs serving the population needing employment supports while acknowledging that specialized programs may be more effective in serving difficult populations.

Having consistent and reliable data across all states and local workforce areas is essential to using the data to manage programs. Without reliable and consistent data, the entire performance process is at best a waste of effort and at worst a source of bad policy (rewarding inferior and punishing superior performance). Most program operators at the state and local levels are diligent and honest, but there are some who see performance as a game, not a management tool, and find clever ways to manipulate their performance outcomes. The most fundamental challenges to obtaining reliable and consistent performance data are lack of precision and clarity in data requirements and lack of standardized and sophisticated data processing and calculation tools.

B. Defining Data Elements

The risks to reliable and consistent data are two-fold: (1) case workers will interpret and thus enter information into the case management systems very differently, and (2) the data will be processed very differently by software developed separately by every state or grantee. Some of the WIA measures were ill defined early on, but in 2001 ETA did launch an ambitious data validation effort that has resulted in data that are considerably more reliable.

The first step in obtaining reliable data is to write clear, objective definitions and to define precise source documentation to verify the validity of a data element. This is much more difficult than one would

think. When asked to validate data, we have responded by asking policymakers for the criteria to use to determine valid from invalid data and what source documentation we can use to make the determination. Policymakers are often stumped by these very difficult questions. Measures and their component data elements should not be used if they cannot be defined and validated.

There were some definitional problems in the original WIA youth measures for credential and skill attainment. The skill attainment measure was intended to give credit for youth that advanced in a skill area over a one-year period. The first operational step was to limit the number of possible skills attained to a maximum of three per youth per year. This put a cap on local areas setting large numbers of easily attained goals for a single customer. The next step was to define a skill attainment. Some felt that this was too difficult because of the wide variation in capabilities of the youth being served. An easy skill for one youth might be a huge challenge for another. This is obviously true, so ETA decided to provide flexible parameters for what constituted a skill attainment. Case managers used this flexibility to decide subjectively, on a case-by-case basis, what constituted a skill attainment and, in so doing, created inconsistent definitions of skill attainments across states and WIBs. Thus from the first day it was difficult to compare the skill attainment results across reporting entities. Considerable effort was made to program the calculations, to train the states and local areas, and to collect all the required data and discuss what it all meant. In fact, such vaguely specified measures end up costing more than clearly defined ones, because there is never any closure to the discussions on how to calculate the measures and what to make of the results. This is an example of how effort and resources can be wasted if performance measures are vaguely defined or performance data are inconsistent and unreliable.

The credential attainment measure met a similar fate. The first problem was that some decision-makers believed they needed to show strong results early in the implementation of WIA in order to demonstrate the effectiveness of the program. This view led to a loose definition of credentials, which encouraged states to define them so as to ensure that most customers would attain them. One state manager said it was “like a license to print our own money.” Needless to say, the measure produced unreliable data.

Fortunately, the advent of common-measures discussions by OMB in 2003 allowed ETA to correct these definitional problems. Partly based upon the lessons of the skill attainment and credential rates, OMB and ETA decided to develop new measures that would overcome some of the deficiencies of the original ones. They defined *credential* more strictly by eliminating work readiness credentials and focusing more on credentials and certificates that reward improvement in occupational skills. They also merged the credential rate with the diploma rate, which led to the new attainment of a degree or certificate

rate. In addition, they replaced the skill attainment rate with a literacy and numeracy gains measure that required that states use U.S. Department of Education–approved standardized tests to determine whether or not an individual youth had improved his/her skills. This change created a well-defined measure but presented a complex challenge to write detailed specifications for calculating the measure accurately, given the almost infinite number of possible sequences of test scores and exceptions. Once the programming was done, testing the accuracy of the calculations consumed hundreds of hours of staff time.

C. Manipulating Performance

Performance outcomes can be manipulated during the enrollment and exit processes. A casual observer would not see how difficult it is to define enrollment date and exit date, which drive all performance calculations. Some states' first reaction to the launch of WIA was to impose more restrictive criteria on enrollment. They did not want to be held accountable for outcomes for customers who received very inexpensive services. The lower enrollment did not reflect the number of people being served, just the number for which the state was accountable in the performance system. This was done by redefining "staff-assisted services." Because WIA and Employment Service (ES) are universal-access programs with a broad range of services, from self-service only (e.g., using the job-seeking aids on websites or at one-stops without assistance) to staff-assisted training, there was significant discussion early on about at what point in the continuum of services a customer should be formally included in the performance system. ETA instructed states to include customers when they received significant staff-assisted services. Some states, however, defined staff-assisted services very broadly, while others defined them very narrowly. Enrollment numbers fell sharply in some areas. It is not clear whether particular types of customers (such as incumbent workers, where earnings gains would be most difficult to achieve) were more likely to be left out of the reporting system or not. Measuring performance outcomes for customers who received little or no staff assisted service and may have never visited a one-stop is problematic. On the other hand, there has been very significant public investment in self-service facilities so it is appropriate to determine whether that investment has led to better outcomes. Self-service utilization measures might be good complements to outcome measures for this customer group.

Another more direct way of distorting outcomes was to manipulate exit dates. One dilemma states faced when they converted from JTPA to WIA was what to do with hundreds of thousands of JTPA customers who had never exited from JTPA but were no longer receiving services. The records of these customers had gone to the "data graveyard," never to be included in performance outcomes. States were instructed to purge these unmeasured customers from the system to allow WIA to start with a clean slate,

and the concept of “soft exit” was developed to prevent a reoccurrence of the problem. States were instructed to generate an exit date for any customer who had not received a service for 90 days. There was much discussion about whether the exit date would be the last date of service or 90 days later, and also about how to avoid exiting customers in long-term training programs. The obvious means of manipulating performance is to avoid exiting customers until they have been placed in jobs. It is impossible to enforce rigorous standards or consistency across states for these issues because there is no way to tell from the case management files whether there were real continuing services provided or if the customer was being "held" in the system until job placement.

D. Defining Employment and Earnings

Even defining employment and retention and earnings are not straightforward. Traditionally programs using manual follow-up methods have used fairly rigorous definitions. For example, SCSEP required exited customers to be employed for 60 of the first 90 days after exit to be counted as placed. Other programs set minimum levels of hours per week to exclude very partial employment or looked at average hourly wage and even whether the employment was related to the training provided. The universal use of wage record data under WIA raised a new set of issues. Wage records were quite thorough (all wages reported from multiple employers could be easily captured and aggregated), but did not provide details on employment and only report quarterly totals. The number of hours worked, the hourly wage and the occupation were generally not available and it was not known if someone worked one day in the quarter or 90 days. Therefore, it was decided that the total earnings in the quarter after exit would define “entered employment.” The threshold for total dollars required was discussed, and finally ETA determined that any amount would qualify making a very low barrier for placement.

The wage record system does not operate to serve employment program research or performance assessment, but to determine employer Unemployment Insurance (UI) tax rates. It would be useful but difficult to obtain more detail on employment from employers, but given the high degree of automation of payroll systems, especially for larger employers, it may be feasible at some point in the future.

The earnings gain measure raised a host of additional definitional and technical problems. States had considerable concern about how enrolling laid-off, high-wage manufacturing sector workers would produce sharp earnings decreases after services when they were placed in lower paying service industries. Some initial analyses of dislocated worker earnings replacement rates, however, showed earnings gains of over 300%. These spectacular results derived from customers who had already received services for a year or more prior to enrollment in WIA and thus had zero pre-program earnings. Defining the dates to use to calculate pre-program earnings, determining the actual dislocation date, and then collecting the correct

quarters of wage record data proved to be very problematic. States approached these issues in various ways and with various levels of success leading to inconsistent results.

A related issue involved measuring earnings gain from the first quarter after exit to the third quarter after exit. This did not seem to be a meaningful measure because it only measured earnings increases over a very short period when raises would not likely be provided. This measure was supported by the Temporary Assistance for Needy Families (TANF) program where there were by definition no pre-program earnings, however. Ultimately ETA abandoned pre-program and post-program earnings measures and now reports only average post-program earnings.

The lessons from the early implementation of WIA are clear: do not attempt to measure something you cannot define or validate, and make sure the calculations are reliable and well tested.

E. Reporting and Validating Performance Data

ETA has been in the forefront of federal performance management and data validation efforts since the 1970s for two reasons. First, the U.S. Supreme Court ruled in 1969 that the Department of Labor (DOL) had to ensure that UI claimants received payments on time, and this required measuring the timeliness of UI activities. Second, it was discovered that the allocation of administrative funds to state UI agencies was inequitable because of inconsistencies in how states counted their activities. UI conducted a significant upgrade to its performance management and data validation systems in the 1990s and set the model for the rest of the federal government to follow, which generated a healthy culture of data quality and standardization in the state workforce agencies.

With the passage of WIA, ETA sought to bring standardized reporting and data validation to the workforce programs, and such systems were gradually put in place starting in 2002. The data validation process asks whether the data used to calculate performance are correct and whether the performance measures were calculated correctly. This process led to much more rigorous definitions of data elements, as well as the development of standardized reporting software that states could use to edit files, perform reporting and performance calculations, and receive immediate feedback on data problems and performance. It also served to enhance the quality of program data and greatly speed the availability of performance data to ETA, OMB, Congress and the research community. In addition, the reporting and data validation software provided basic analytical functionality so that states could explain in detail the changes in performance over time. Some states made extensive use of the analytical capabilities of the software to educate their local areas about how the measures worked.

Performance calculations are highly complex, and extensive testing is required to ensure their accuracy.⁸ Before the use of standardized data calculations at a federal level, each state calculated its own performance at considerable expense and with inconsistent methods and results. Initially, ETA was committed to the standardized reporting and data validation methods and processes, especially in light of DOL Office of Inspector General (OIG) reports on deficiencies in ETA data.⁹ The standardized software required continued investment to maintain because of the large number of changes in the performance measures and because the software had to be enhanced to meet growing state analytical and diagnostic needs. States embraced the concept of accurate and consistent performance calculations and data edits and liked the immediate feedback they received on their data quality and performance.¹⁰ Further, the total cost of using standardized software was far less overall than the cost of having each state program its own calculations, not to mention that the separately calculated performance data would be unreliable.¹¹

In addition to editing files, calculating performance, and providing basic analytical functionality for states, the data validation software samples customer records for validation of data elements. In the

⁸ The data reporting and validation software calculates over 1,600 individual cells on various WIA and ES reports, as well as tens of thousands of additional calculations needed for other reporting, validation, and analytical functions. These calculations, reports, and functionality are documented in more than 500 pages of specifications and high-level requirements. The software also applies roughly 300 edit checks to the data. Extensive testing is done to ensure that the calculated results are correct for every state regardless of numerous variations in data files submitted by the states.

⁹ See “Workforce Investment Act Performance Outcomes Reporting Oversight.” Office of Inspector General, U.S. Department of Labor, Report No. 06-02-006-03-309, September 30, 2002. Also useful are “Audit of the Unemployment Insurance Data Validation Pilot Program.” Office of Inspector General, U.S. Department of Labor, Report No. 22-02-005-03-315, March 29, 2002; “Job Training Partnership Act Title II-C Out-of-School Youth Performance Audit.” Office of Inspector General, U.S. Department of Labor, Report No. 06-01-001-03-340, March 19, 2001; “Improving Trade Act Programs.” Office of Inspector General, U.S. Department of Labor, Report No. 04-01-009-330, September 26, 2001; “Workforce Investment Act Evaluation of Youth Program Enrollments, Services, and Recorded Outcomes.” Office of Inspector General, U.S. Department of Labor, Report No. 06-03-006-03-390, September 30, 2003; “GPRA Data Validation Review Trade Adjustment Assistance Program.” Office of Inspector General, U.S. Department of Labor, Report No. 22-05-007-03-330, September 15, 2005; and “Unemployment Insurance Data Validation Pilot Program.” Office of Inspector General, U.S. Department of Labor, Report No. 22-02-005-03-315, March 29, 2002.

¹⁰ The feedback was immediate when states loaded their customer files (e.g., the WIASRD) into the software, but was still constrained by data lags associated with wage records.

¹¹ Many states have invested in performance software and use the federal validation software for testing and to validate their performance reports. Even if the state calculations are determined to be incorrect by the data validation software, ETA uses data from the state calculations and not the validated calculations in its report to Congress. 20-25 states use the federal software to generate their performance reports. Given budget cuts and the focus on other priorities, ETA reduced the funding for the maintenance of the reporting and data validation software for WIA and ES. The functionality for states has diminished since 2005, and many of the suggestions states made for enhancing the software have not been implemented. The Office of Inspector General concluded that “with the lack of software upgrades, the effectiveness and efficiency of using the data validation software as a tool to improve the accuracy and reliability of WIA performance data has been compromised.” See the Audit of Workforce Investment Act Data Validation For the Adult and Dislocated Worker Programs.” Office of Inspector General, U.S. Department of Labor, Report No. 03-090003-03-390, September 30, 2009, pp. 3, 11.

validation process, state monitoring staff review a small sample of records against supporting documentation maintained at the local area. The software contains sampling algorithms that make the state monitoring process as efficient as possible for creating estimates of errors for each data element by state. State staff generally found the data validation process to be very helpful and efficient for monitoring data quality at the local level. Unfortunately, the software does not yet compute the standard error rate for each data element, so ETA cannot set or enforce data accuracy standards.¹²

F. Federal Performance Monitoring

ETA's regional offices have always monitored state programs. There are two basic forms of monitoring: (1) *process monitoring* consists of reviews of required state functions to ensure that they meet federal standards, while (2) *data monitoring* involves reviewing samples of records against source documentation and is thus identical to the data validation effort conducted by states. In 2006, ETA designed a process by which federal regional monitoring staff would review a subsample of the records reviewed by the state staff to ensure that states were applying the data validation rules consistently. The federal staff would enter their results for the sub-sample into the data validation software, which would then generate a report to ETA on the state's accuracy. This monitoring process has not yet been implemented. The recent OIG report concluded that "without an effective monitoring process, ETA has no assurance that data validation is operating as designed so that the data can be relied upon for accurately reporting performance results."¹³

Some ETA regional office staff developed their own data-monitoring processes, but they are implemented inconsistently, and not all regions monitor data systematically. Therefore, there is no systematic check on whether the states are performing the data element validation consistently and correctly. With no data accuracy standards, no precise calculations of state error rates, and no check that states are performing the validation correctly, the reliability of WIA data is still not clear. This is unfortunate because states still incur the full burden to perform annual data validation. This annual validation exercise does allow state staff to conduct effective data monitoring of local areas and thus facilitates the detection of data problems and discussion of remedies.

III. THE CHALLENGES AND BENEFITS OF STAKEHOLDER INVOLVEMENT

¹² The UI data validation program does have data accuracy standards and computes reliable estimates of error taking sampling error into account.

¹³ Audit of Workforce Investment Act Data Validation For the Adult and Dislocated Worker Programs, Office of Inspector General, U.S. Department of Labor, Report No. 03-090003-03-390, September 30, 2009, p. 11.

The process of designing performance measures starts with the funding legislation. Congress requires that recipients of funding submit certain performance outcomes to justify continued funding. Statutes also normally contain some language about remedial or punitive steps that will be taken if programs fall short of performance goals. Of course, Congress does not operate in a vacuum, and the legislation reflects input from the executive branch (OMB and agencies) as well as from lobbyists for the state agencies and population-specific interest groups.¹⁴ Statutory language about measures is usually very general, so agencies must add further levels of detail to “operationalize” the measures, including the specific data elements that must be reported and how the measures are to be calculated. Effective performance measurement requires strong leadership from the federal government both in defining the measures and objectives and in providing the definitional structure and necessary performance management tools. In some programs with immature performance management processes, program operators are given the latitude to define or choose their own measures. This approach may be politically popular, but it rarely produces any usable results and does not lead to program improvement. States and grantees look to the federal partner for leadership and structure but still want input on the operational details.

A. Soliciting Feedback and Consulting Program Operators

Both aspects of the dual rationale for measuring performance—accountability and program improvement—make it desirable that program operators “buy in” to the system. Obviously, the program management and program improvement rationale for measuring performance is advanced when program operators find the results meaningful and helpful. Even the program accountability rationale works best when operators find the measures to be legitimate. Therefore, the programs discussed in this paper have sought extensive consultation from program operators (states, local areas, and grantees) during the design process and during the phase of the process when the measures are being operationalized.

At the end of 1998, ETA produced an initial draft of the approach to measuring WIA, but it did not contain clear and well-defined measures, and was thus not well received by the states. Therefore, during the spring of 1999 there ensued a series of consultative meetings attended by federal and state staff. Six early adaptor states launched WIA on July 1, 1999, and representatives of these states met with federal staff over a series of months to hammer out the details. The first complete set of technical performance specifications was published in March 2000, before the other states implemented WIA on July 1, 2000. Input from state staff was very helpful in operationalizing the measures because of their

¹⁴ The National Association of State Workforce Agencies (NASWA) represents the state agencies that administer WIA and most related programs. The VR program is represented by the Council of State Administrators of Vocational Rehabilitation (CSAVR).

rich knowledge of program operations and workforce data. Feedback on the technical aspects of the measures continued to be received during conferences and meetings for the next two years. Other programs had similar if less extensive consultations. All the ETA programs relating to WIA established performance workgroups to seek input from state and grantee staff.¹⁵

B. Fear of Performance Management

The performance measures were seen by some states, local areas, and grantees primarily as a threat rather than as a management tool. This perception greatly influenced their input on how the measures should be designed. It is logical that those at risk of sanctions from measured poor performance would become defensive and try to reduce the effectiveness of the measurement system itself. This defensive impulse leads to actions to evade the implications of the measures and to resist measurement in a number of ways. As mentioned in the data validation discussion above, state WIA staff play a dual role in the performance system and are sometimes defensive because they are being measured as states and sometimes supportive of effective performance methods because they are oversee local areas.

The essence of the resistance to effective federal performance methods was documented in a Government Accountability Office (GAO) report on WIA, which said that “collecting uniform performance data at the national level [and] giving states and localities the flexibility they need to implement programs” are “competing objectives.”¹⁶ This is based on a misunderstanding actively pushed by people who resist performance processes to muddy the true role of performance management techniques in improving government services.

We must distinguish clearly between service delivery and program management. Performance management systems track common events such as enrollment date, customer characteristics, limited service dates, exit date, and outcomes. Performance management systems do not specify how services are delivered. Therefore, there is no inherent conflict between (1) allowing program operators creativity and flexibility in customer outreach and providing services on the one hand, and (2) tracking customer characteristics and flow through the system and measuring outcomes on the other. Local program operators and grantees are fond of saying that they can either serve people or enter case management data, but they cannot do both. This is all too often a defensive reaction to fear of being measured and a reflection of inadequate management capacity. That is why it is so important to focus initially on building

¹⁵ SCSEP, ES, NFJP, and the Division of Indian and Native American Programs (DINAP) all convened performance workgroups in 2000 and 2001.

¹⁶ “Workforce Investment Act: Labor and States Have Taken Action to Improve Data Quality, but Additional Steps Are Needed.” November 14, 2005, General Government Accountability Office, GAO-06-82, p. 1.

strong data capacity through effective performance management tools and methods rather than on the punitive aspects of performance management.

C. The Relative Burden of Federal Performance Requirements

Despite the complaints about the burden of federal data requirements, many states collect far more detailed performance data and invest in more sophisticated performance management systems than anything imposed by ETA. There is wide variation among states, grantees, and local program operators in their level of sophistication and the level of case management data they collect. The goal of the federal performance management system, including the key data validation component, should be to raise every state and grantee to a minimum acceptable level of data management and data reliability. There will always be states with more sophisticated performance systems than are practical for the federal partner to develop.

Tracking participants, services, and outcomes is essential to any effective program management at all levels of the system and would be done for the most part in the absence of any federal performance and data validation initiatives. From having overseen the development of the Paperwork Reduction Act requests for performance and data validation for many years, it has become clear that little information is collected solely for performance purposes and that none of it constrains program operators from employing innovative and diverse service delivery methods.

Once we accept that program operators must know whom they are serving and what services they are providing, the only aspect of performance management that is a true burden is collecting outcome data. It is less important to the basic management of the program to track extended outcomes as required by WIA retention rates than to track customers and services. But if those long-term outcome data can be collected efficiently through wage records or some other form of already existing high-quality administrative data, then measuring even long-term outcomes becomes very cost-effective. There is also a strong argument to be made that long-term follow-up services are an expensive but essential part of an effective service delivery strategy, especially for the hardest-to-serve populations. The WIA youth program and SCSEP both require long-term follow-up to support customers after exit, but this is resisted in the VR program even though the need for long-term support is evident in that population.

Another source of resistance to performance management is the concern that the population served is too varied and complex to permit effective measurement of the actual performance of a program operator. This becomes a problem when the emphasis is on incentives and sanctions and not program management, because program operators do not trust that the measures are fair. This notion is reinforced by the "social worker" mindset that is especially pervasive in programs serving special hard-to-serve

populations. Many staff in these programs assert that all programs must be run well and must be effective because program staff are sympathetic to the population being served. There are thus three complementary threads to the resistance to effective performance management: (1) collecting data is a burden; (2) performance measures cannot accurately reflect the quality of services rendered; and (3) staff are well intentioned and therefore must be left alone to perform their work.

To be fair, local program operators and grantees often operate under stressful conditions. They serve very difficult populations with inadequate and declining funding levels and operate under the weight of threatening and somewhat crude performance measures. It is therefore critical that the performance system be sold primarily as a means to achieving better management and analysis capacity.

There is one critical area of performance management in which program operators are forced to bear a true burden for which no relief is likely to be found. That burden comes with requirements to collect data validation documentation from the most difficult-to-serve populations, such as homeless youth, people with disabilities, very low-income older workers, and non-English-speaking customers. Collecting such documentation is important to program integrity, not only because these are important program eligibility criteria, but also because programs are given extra credit for serving these people. SCSEP measures, for example, give credit to the grantees for specific categories of customers (disabled, homeless, low literacy skills, and frail), and in the VR program the category of “severely disabled” is critical to program intake and performance outcomes. In addition, programs allow performance outcome exemptions for medical conditions and even for medical conditions of family members; this is a major issue for the integrity of performance outcome data for SCSEP, which serves many people over age 70. It is very convenient to avoid a negative performance outcome by classifying the customer as excluded from performance.

Collecting documentation to show that customers meet criteria for extensive barriers to employment or exclusion from performance represents a true burden for case managers. Medical conditions, disability, homelessness, homeless youth, and family income are all very difficult areas to document. For example, how do you prove you are homeless? The only approach that we have found is to allow—in lieu of actual evidence—“self-attestation” in the form of documents the customers sign testifying as to their conditions. This will continue to be a challenge to effective performance management for the foreseeable future.

D. Measuring Accountability

Once we get past the “data are a burden” argument, we find a more subtle and valid tension between simplicity and clarity in measures and determining the program's actual accountability for outcomes. With a defensive mentality, program operators view performance measures as directly measuring their accountability or their effort in serving each individual customer. In fact, it is impossible to design measures that can account for all the factors bearing on success with a single customer. Performance management is a statistical process that assumes that measures of a sufficient number of outcomes can distinguish between more successful and less successful processes and methods.

Not understanding how performance management works, program operators seek direct measures of their accountability and thus want the measures to be designed to account for every exception. One state staff person argued that their state should not be held accountable when a customer failed to show up for services and had a negative outcome. I responded with two questions: Why would more people fail to show up for services in your state than in other states? And if customers did tend to show up less in your state than in other states, was that not a valid finding about the quality of your services? Performance goals are always set well below 100 percent so that the system accounts for such “failures” that cannot be directly attributable to a program operator’s deficiencies.

The impulse to design measures that account for individual customer circumstances leads to exponential increases in complexity. Each additional factor that a measure must consider to define success, such as excluding outcomes where customers became ill, doubles the number of possible outcomes. Some accountability factors are significant enough to incorporate into the measure design, and more sophisticated measures are practical if standard automated tools are used to perform data analysis and calculate measures. But ironically, once program operators have succeeded in adding factors to better measure what they are directly accountable for, they often complain that the measures have become too complex to understand or to explain to their local stakeholders. So there is a tricky balance between designing detailed measures of actual accountability and designing measures that are easy to understand and explain.

IV. THE WIA PERFORMANCE SYSTEM

A. The Evolution of WIA Measures

Stakeholder concerns had a direct and significant impact on the early WIA measures, where there was significant input from states and local areas. The initial WIA measures were very simple, but within a year they had become much more complex. One example of a change that added complexity but greatly strengthened the measure was in the treatment of youth who were placed in postsecondary education but not in employment. Originally, since it was considered that the Department of *Labor* could not reward an

educational placement, the postsecondary education placement without employment was classified as a negative outcome. In other words, placing a youth in Harvard was bad, but placing him at McDonald's was good. After further discussion it was decided that postsecondary placement without employment would become a "neutral outcome," where the record was excluded from the placement calculation completely.

In 2003, OMB launched a common-measures initiative for federal employment programs to try to standardize performance calculations across the many federal job training programs. This effort was prompted by the breakdown of the competitive approach to program funding. The Bush administration sought to use program outcome data to determine which programs were effective and which were ineffective and should be defunded or folded into more effective programs. This attempt was confounded by the lack of comparability of performance data across the data sets. For example, SCSEP defined a successful placement as 30 days of continuous employment within the first 90 days after exit, while WIA defined it as any earnings at all in the quarter after exit. The disparate definitions of success and performance goals across programs made it an even greater challenge to control for differences in the populations served.

ETA embraced the OMB initiative and launched a second round of state-federal discussions over how to implement the new common measures. Other programs within ETA, including VR and even SCSEP, resisted the common measures, arguing that their populations were special and that they could not be expected to achieve results comparable to those of the mainstream programs. SCSEP has since adopted the common measures, but VR has still not implemented them. The common measures are a good step toward effective performance management at the national level across training programs and do not constrain programs from using other performance management tools.

B. The Use of Wage Records

In the absence of a good source of data on post-program earnings like the UI wage record system, it would be very difficult to develop an efficient and effective performance management system, especially for such large-scale programs as WIA and ES. UI wage records are reasonably reliable because they are official tax records and are subject to some audit controls. They are not perfect, however, for a number of reasons, including uncovered employment, failure to report by employers, and errors in reporting that prevent matches of wages to participant records, but they do supply the vast majority of the data needed to measure outcomes. ETA must continue to allow states to collect "supplemental" earnings data collected directly from program customers to compensate for the gaps in wage record data. This is particularly important, because the need for supplemental data varies widely by region. Wage record data

are significantly less complete in agricultural areas; areas with a larger “underground economy” (such as Hawaii and Puerto Rico, where employment in tourist-related industries is more informal); and in areas with a high concentration of contract labor, such as the movie and software industries. Another critical issue is providing states with access to wage data collected by other states. Until recently, ETA had experienced mixed success in establishing such a system, but privacy and legal concerns have rendered interstate wage data useless for performance management purposes. States can send files of Social Security Numbers and get back aggregate rates of matches with a national wage file (including federal and military employment) to obtain higher and more accurate entered employment, retention, and earnings data; however, this data is not provided at the customer level and the information is useless for analyzing and improving performance. Many states have had bilateral wage-record-sharing agreements since WIA began and can continue to use these more detailed data to analyze their performance at the customer level.

Not all employment and training programs can access the state wage record file; this is either because some are nongovernmental entities or because it is too cumbersome to negotiate access with the UI agency. SCSEP, for example, still conducts manual follow-up with each exiter up to three times to obtain post-exit earnings data, which must be carefully documented for data validation. This additional burden can be seen as adding value because it allows grantees to provide follow-up employment support services. The Pathways to Work project in the U.K. planned to conduct extensive provider follow-up because there were no available earnings data equivalent to the UI data in the United States.

One of the major problems with reliance on wage data for performance management is that the files are not complete and available until about six to nine months after entry into the employment activity being measured. This prevents timely feedback to program operators, but it is still a far more cost-effective approach than expensive and unreliable informal follow-up data as gathered under JTPA. The six to nine month lag in the availability of complete employment outcome data is an unfortunate reality and does limit the benefits of the analytical feedback loop to program operators, which is a key aspect of an effective performance management system.

Although outcome data is the primary source for performance management, additional data is helpful for some programs where customers receive services over a long period of time. This is especially an issue in the WIA younger youth program where customers may be enrolled for five years or more and to some degree in SCSEP and VR as well. These programs would benefit from intermediary progress measures to provide more timely feedback to program operators on their performance. The literacy and

numeracy gain measure in the WIA youth program is an ideal measure not only because it is well-defined but also because it provides continuous feedback on youth progress to program operators.

V. CONCLUSION

This paper has begun to introduce some of the challenges of effective performance management. We can conclude that top priority should be placed on establishing a solid foundation of collecting and processing data consistently and accurately to help Congress, program managers, and local One-Stop administrators to understand who is being served and what their outcomes are. There are many technical aspects to developing this foundation, and this paper has only touched on them. An effective performance management system requires enlightened federal leadership with a sound understanding of the potential and limitations of performance system and a commitment to effective performance management. Federal staff must take the lead in promoting the value of performance management to the other levels of the system and firmly enforce performance objectives.

Only when the foundation for effective performance management is securely in place, should policymakers take punitive action on the findings. They should concentrate initially on identifying superior and inferior performers, analyzing which processes and methods produce the best results, and providing technical assistance to the poor performers. The least emphasis should be on rewards and sanctions. These motivational devices can be useful but are often rushed into play before the data are reliable or well understood and thus engender resistance to performance management and inappropriate behavior by program operators.