

Performance management in federal employment and training programs

Carolyn J. Heinrich

Carolyn J. Heinrich is Associate Professor of Public Affairs, University of Wisconsin–Madison, and an IRP affiliate.

If there is a single theme that characterizes the public sector in the 1990s, it is the demand for performance. A mantra has emerged . . . at all levels of government, that calls for documentation of performance and explicit outcomes of government action.

— Beryl Radin, *Beyond Machiavelli*¹

Performance measurement as a tool of public management has a long history. Its primary goal has always been accountability—to legislative bodies, taxpayers, and program stakeholders. But the direction and purpose of performance measurement have been changing. Formerly, it was most intensely focused on financial performance or efficiency. The initiatives encapsulated in “planning, programming, and budgeting,” “management by objectives,” “zero-based budgeting,” and other popular buzzwords of the 1960s and 1970s were concerned primarily with the relationship of inputs to costs and the value of cost-reduction activities. In the last two decades, performance measurement has entered a much wider arena. It has come to be seen as an important tool for improving the quality of an agency’s services and the consequences for those served.

One reason for the shift was increasing dissatisfaction with systems that required managers to narrowly define and measure progress toward financial, technical, and strategic performance goals. Management theorists such as W. Edwards Deming challenged the “narrow, simple-minded” focus of management by the numbers and urged managers instead to strive for and measure quality.² Organization and management theories were evolving toward more open, adaptive models that took into account the agency’s environment, including the informal norms, social contexts, and communication and status issues that influence staff performance. In addition, the expansion of block grant programs from the late 1970s on shifted considerable responsibility and discretion to state and local governments. As the distance between the individuals who received services and the officials at the highest funding level increased, so too did anxiety about how local officials were exercising discretion in distributing federal monies. One consequence was the development of new mechanisms for accountability.

In the United States, a central piece of legislation in this changing perspective was the Government Performance and Results Act (GPRA) of 1993, which requires federal agencies to set goals and objectives, measure performance, and report their accomplishments as part of movement toward a performance-based environment. These requirements were intended to provide political accountability for results and give agencies the opportunity to increase their responsiveness to program stakeholders and constituencies.

Even this partial shift in emphasis from “government that costs less” to “government that does a better job” raises a whole host of issues. One challenge to an agency that tries to move beyond purely financial, efficiency-driven targets is reaching consensus on clearly defined and verifiable public objectives. Fragmented programs, multiple goals, and the deficiencies and inconsistencies of authorizing legislation can make it very difficult for staff and stakeholders to think about how their diverse activities are related to a common outcome. Some agencies, as a result, may opt for goals that are vague, uncontroversial, inconsequential, or easily attainable.³ Some prescribed program goals may be mutually inconsistent. The U.S. Job Training Partnership Act (JTPA) of 1982, for example, stated that programs should serve “those who can benefit from and are most in need of” employment and training services. Research suggests that efforts to reach those most in need (effectively, the bottom 20 percent of the skill distribution) may lead to modest inefficiencies in the allocation of program resources.⁴

Still unanswered is whether performance management systems that hold agencies accountable for the outcomes of their activities are more effective than traditional bureaucratic controls that hold agencies responsible for inputs and processes. Some analysts believe not, and they have been harshly critical of the kind of outcomes-based public management codified in the GPRA. They argue that its requirements for specific performance goals, plans, and results have increased administrative constraints, elevated conflict within and among agencies, ignored political complexities, and bred a sense of cynicism and an attitude of formal compliance within the federal government. These failings are compounded, they note, when strong leadership or effective management are absent within the agency.⁵ Further, the reporting requirements promote emphasis on short-term goals that may be perversely related to longer-term outcomes, encouraging employees to “game the system” and “cream-skim” applicants. Caseworkers, for example, may seek to improve a

job-training program's short-term results by selecting as participants those who would have done well without the program.⁶ Likewise, performance requirements based on students' annual test scores may lead some teachers to "teach to the test," at the expense of students' longer-term educational success.

The objectives that an agency chooses have implications for the quantitative measures of performance that it must develop. Vague or conflicting goals make the task very much harder. All agencies process vast amounts of paperwork (and, increasingly, amass electronic files), and most issue annual program reports that compare the agency's performance with a set of predetermined targets. But many of these targets are concerned with processes—how many people came through the agency doors, how many checks were issued, or contracts let. Others reflect an emphasis on financial efficiency that is in some tension with the goals of social service agencies. The Health Administration of the U.S. Department of Veterans Affairs, for instance, defined its goal as improving the health status of veterans, but then, in glaring disparity with this aim, identified cost reductions per patient and number of patients served as measures of progress.⁷

The kind of information needed to help managers understand why performance is at the level it is or how they can effect change is much more complex than the information needed to monitor where the money goes. It encompasses resources and staff, workload and job complexity, outputs and outcomes in relation to intermediate and long-range goals, and effects (impacts) on service users. Ideally, the full range of this information would be used by public managers in a logical flow that moved from monitoring the agency's performance (its processes and their efficiency), to evaluating the program's outputs or effects, and then to management—the use of information on past performance to guide program planning and improve future performance.

Acquiring such information was never going to be simple or cheap, but it has been made easier by advances in statistical techniques for measuring performance. The last three decades have also seen the development of large and varied bodies of experimental and administrative data, as state welfare programs underwent mandatory evaluation, national programs such as the JTPA were documented and closely studied, and federal and state data requirements multiplied.

The U.S. Government Accountability Office (GAO), investigating performance management in public agencies, has suggested that we might gain a more precise understanding of causal links, program effects, and the relationship between short- and long-term goals by supplementing performance data with impact evaluation studies.⁸ But such experimental evaluations are likely to be infrequent. They are usually expensive, may disrupt an agency's operations, and probably cannot generate the

timely, regular feedback that program managers need to make adjustments in their budget allocations and practices. If experimental evaluations are likely to be rare, for reasons both practical and political, then the most readily available information comes from the administrative and performance data that government agencies at all levels now acquire and report. But can these data meet the expectations that the federal government codified in the GPRA, and provide reliable information for improving agency performance?

In the research summarized in the remainder of this article I explore issues of public agency performance management in the context of federal job-training programs for low-income and unskilled workers. What made this research possible is the existence of a large body of comparable experimental and program data for job-training programs that has been gathered under the JTPA and its successor, the Workforce Investment Act (WIA) of 1998.⁹ Here I consider two questions in particular: whether the measures that were established for the JTPA produced reliable and useful information for program managers, and whether the changes introduced under WIA constitute an improvement in the measurement of performance in public agencies.

Performance management in federal job-training programs

In an era of decentralization, the performance standards system established under the JTPA was a pioneer. It centered measures on program outcomes (the number of trainees placed in jobs, and how much they earned, for example) rather than inputs or outputs (the number of persons trained); it linked measures of program performance across multiple levels of government; and it included financial (budgetary) incentives for program managers based on the evaluation of organizational outcomes. In general, the system was designed to focus the attention of management at all levels on the central objectives of the program and lessen the government's need to monitor an agency's processes and compliance with federal regulations.

JTPA agencies were also the subject of an experimental evaluation, the National JTPA Study. Using the data from this study, we can compare the reliability of the performance measures against outcomes measured in a random-assignment experiment. The three-year experiment, commissioned in 1986 by the Employment and Training Administration of the U.S. Department of Labor (DOL), was conducted by MDRC, Abt Associates, and their subcontractors. It involved about 20,000 individuals, randomly selected into treatment and control groups, in 16 JTPA programs. It has been described as the largest and most rigorous evaluation ever conducted of programs designed to increase the employment and earnings of disadvantaged adults and youth.¹⁰

Responsibility for programs under the JTPA and WIA is shared among multiple levels of government. Funding and broad oversight are an obligation of the DOL, in part through the establishment of performance standards. Job search and placement assistance, job-readiness activities, case management and supportive services, and other more intensive education and training are regulated and monitored by state governments and administered by local government agencies or private industry councils and boards. Local program managers maintain records based on the established performance standards and report data to the state agencies, which are responsible for determining performance bonuses or sanctions for the local agencies but also transmit the data to the DOL.

Over nearly two decades of operating in a highly decentralized environment, the local and state agencies administering federal jobs programs have evolved quite different governance and management systems. This variability is not random. In work with Laurence Lynn, Jr., I found strong associations between the administrative structures chosen in different service-delivery areas and the types of policies and incentives adopted by states and local agencies to motivate performance. For example, when the local administrative agency was a private-sector body, it tended to emphasize measured performance and to adopt administrative practices such as performance-based contracting. In those areas where local public officials played a larger role than the private sector, performance-based contracts were less likely and there were more explicit incentives to focus on “hard-to-serve” groups.¹¹

State administrative policies, the size and the population characteristics of the local service-delivery area, and the wide discretion afforded local units have also fostered some local variation in the performance measures themselves (see Table 1). Research does suggest that the basic performance *standards* in JTPA programs have been fairly consistent across government levels. But divergences appear when administrators begin to make decisions about how to use the performance information in managing programs—in directing, rewarding, or sanctioning agencies and providers.¹² Disentangling the effects of different measures or actions is likely to be very challenging, because administrative decisions and management actions at different organizational levels may influence not only how well participants do but also the types of performance management policies that are adopted. Program managers aiming to maximize their agency’s measured performance may, without intending to, damage the long-run earnings prospects of participants by too close a focus on short-term gains.

The usefulness of JTPA administrative data for performance management

In circumstances as complex as those I described, can assessments based on limited administrative data provide

Table 1
Administrative Discretion in the JTPA
Performance Standards System

Minimum performance bonus and performance bonus award schemes. The stringency of state requirements differed in a number of ways, such as the minimum number of performance standards a local agency had to meet in order to qualify for a bonus or to be penalized, or the level of performance at or above state standards that generated incentive payments. Many states encouraged competition among local service-delivery areas by making bonus levels contingent upon performance relative to other areas.

Weights accorded to different performance goals. States could not only set different performance levels, but in determining bonuses they could attach different degrees of importance to the core federal performance standards. States and localities could also establish additional performance standards for services to disadvantaged groups.

Service-delivery strategies and contracting. Some agencies provided training services directly to participants, others contracted them out or formed partnerships with local providers. Some local service-delivery areas developed their own systems of competitive bidding and performance accountability for their private providers and used these data in their contracting decisions.

managers with useful information? In the empirical analyses summarized here, I make use of multilevel modeling to test hypotheses about how factors measured at one level of an administrative hierarchy (the state or local job-training agency) interact with variables at another level (the individual client). The National JTPA Study includes data on the demographic characteristics, earnings histories, and program participation of treatment and control group members; information on the administrative structures, performance incentive policies, service delivery, and contracting strategies of the 16 programs; and the unemployment rate and regional indicators of the service delivery areas over the years of the study. Major policy changes during the 1987–89 study years led to significant variation from year to year; this variation expands the number of observations and strengthens the analysis.

In the models, I compared the earnings outcomes of individual participants in the first year after they left the program—these are the data that would typically be available to program managers—with the impacts estimated with data from the National JTPA Study. I found that the estimated effects for individuals were broadly in agreement with each other and with earlier JTPA research.¹³ Dollar amounts differ; one cannot expect that the earnings outcomes will accurately provide the kind of net value-added estimates for the program that might be possible with an impact analysis such as that provided by the National JTPA Study. Labor market outcomes constitute one example of the differences. It had been observed that those with less than a high school diploma generally had poorer labor market outcomes, even after they completed the jobs program, than those who completed high school. According to the administrative outcome data, workers without the diploma earned about \$1,700 less in

the first year, whereas the model using the experimental study estimated that the workers without diplomas in the experimental group received about \$1,100 less. But these models do agree on the nature and direction of the policy and administration effects.

When I included the effects of structural, policy, and management factors on earnings for the year after the workers left the program, both models were substantially in agreement. In both, these controls, which included regional and economic conditions, explained over 90 percent of the variation in outcomes and impacts across the sites. Notably, the relative size and statistical significance of the coefficients in both models were fairly consistent. Using either set of results, policymakers could have determined, for example, that:

- when private-sector representatives assume more formal management responsibilities, participants realize significantly higher earnings levels and the experimental impacts on the treated group are higher;
- participants also earn more when private industry councils *do not* share management responsibilities as equal partners with local elected authorities;
- the weight given to the “entered-employment” rate—one of the most important performance standards in the JTPA program—is positively and significantly related to the earnings outcomes of participants and to the experimental impacts.

In general, then, managers who rely upon the administrative data for guidance in improving performance are not likely to be led astray. They may obtain a better understanding of the effects of policy decisions concerning those factors that are within their control.

That said, over 95 percent of the total variation in earnings outcomes and impacts is within sites, at the level of the *individual*. Moreover, the proportion of the variation in the earnings of JTPA participants that we can explain at this level is low: only 14 percent of earnings outcomes in the administrative data and 6 percent of the experimental impacts can be attributed to the individual-level variables included in the models. This is not surprising, when one considers the many factors that affect an individual’s labor market success, particularly a year after he or she has left the job-training program. Relationships with employers, the acquisition of additional education or job skills, and other environmental influences can all make a difference, and variation among individuals cannot always be explained by observed characteristics. These policy findings are nonetheless important. Organizations cannot change clients’ backgrounds, but they do have control over what have been called “fundamental levers for influencing client outcomes,” such as the availability of services and administrative structures that shape service delivery.¹⁴

If these findings are encouraging, there are still formidable challenges in identifying the influences on organizational performance and linking them to their origins in what managers do or in external factors outside a manager’s control. For example, should the system focus managers’ attention on a single impact indicator or should it allow for multiple goals (equity *and* efficiency)? In performance management systems, public managers have to confront inherent tensions between simple, verifiable goals and more complex measures, and between the capacity and adaptability of the measurement system. The promise of continuous improvement held out by the advocates of performance management might be better served by developing systems that focus on effective policy tools for guiding program management (“getting the question right”) rather than on precise measurement of government performance (“getting the numbers right”). And how should the system identify the influence of diverse administrative priorities and goals? How can it account for the tendency to “game the system” that sometimes develops when goals and priorities diverge?

Improving performance management under WIA

JTPA’s successor, WIA, sought to deal with some of these complexities by changing the way performance measures were established in the first place. WIA retained the basic structure and organizational components of the JTPA program, but made major changes in eligibility, in the types of services offered, and in the processes for performance accountability.

Under WIA, a range of core services such as labor market information or job search assistance is available to the general public, not solely to those with low income. Individuals can access more intensive services, such as comprehensive assessment and case management, if they fail to get a job after receiving the core services. These services are typically provided through one-stop employment centers that also include programs from other human service agencies.

The emphasis on performance management is greater under WIA than under the JTPA. Prominent in the WIA provisions are the measurement and analysis of results, continuous improvement in performance, shared accountability, and a “marketplace” focus on service to “customers” instead of “clients.” The federal government now negotiates performance targets and annual adjustments with the states, abandoning the regression model approach used under JTPA. New measures of “customer satisfaction” (the “customer” being both the program participant and the employer) were introduced. WIA also added credential rates that measure education, training, and skill certification completed by adults, dislocated workers, and older youth, for a total of 17 performance

Table 2
Negotiated Performance Standards under WIA^a

Negotiated Performance Standard	Minimum Level Set across States	Maximum Level Set across States	States Meeting or Exceeding Their Negotiated Performance Target		
			PY 2000	PY 2001	PY 2002
Adult entered-employment rate	45.0%	78.0%	56.7%	66.5%	61.5%
Adult employment-retention rate	60.0	88.0	54.0	60.7	57.7
Adult earnings change	\$674	\$4,638	49.3	64.6	48.1
Adult credential rate	30.0%	71.0%	36.7	45.6	46.2
Dislocated worker entered-employment rate	61.0	84.4	52.7	65.5	55.8
Dislocated worker employment-retention rate	59.0	93.2	42.0	58.7	51.9
Dislocated worker earning-replacement rate	80.0	106.0	54.7	74.8	61.5
Dislocated worker credential rate	27.0	72.0	36.7	58.7	55.8
Older youth entered-employment rate	50.0	75.0	58.7	63.6	42.3
Older youth employment-retention rate	59.0	83.6	52.0	61.2	48.1
Older youth earnings change	\$517	\$4,075	52.7	64.6	59.6
Older youth credential rate	21.0%	55.0%	29.3	31.6	23.1
Younger youth retention rate	35.0	74.0	38.0	59.2	57.7
Younger youth skill attainment rate	50.0	90.0	72.0	69.4	53.9
Younger youth diploma rate	25.0	66.0	25.3	45.6	50.0
Employer satisfaction	60.0	78.0	45.3	75.7	69.2
Participant satisfaction	63.0	78.0	51.3	78.6	76.9

^aFor 50 states, the District of Columbia, and Puerto Rico.

measures (see Table 2). Agencies are expected to develop “five-year plans” in which the negotiated performance standards are revisited every year.

Experience with the JTPA suggests that we should pay very close attention to the way in which performance standards are developed and the incentives they create for program managers and staff. To improve their measured performance, for example, some JTPA program administrators and caseworkers were reported to be limiting the access of more disadvantaged applicants to program services—in other words, “cream-skimming.” Some also strategically organized their “trainee inventories” and timed participants’ exits from the program to improve their year-end performance levels.¹⁵

Early studies of the WIA performance management system have suggested that it is working poorly.¹⁶ One problem is data. States have struggled to meet the new DOL requirements for management information systems; some were able to modify existing JTPA systems, but others had to begin afresh, and experienced the expected delays and complications that attend the development of new software systems. Data lags did not help matters. States constructed their standards for 2000–02 using baseline data drawn from 1994–99. Using baseline data that were two to three years old to project performance targets one to three years ahead might always be less than ideal, but the recession that began in 2000 created some fundamental problems for the state efforts to meet the measures.

In light of these criticisms, how effectively did the WIA performance management system gauge program performance? Did it create better incentives to improve outcomes for participants? To answer such questions, I be-

gan with a qualitative analysis of how states determined their performance goals and standards, and how they made adjustments to those standards. An examination of the variation in and relationships among the negotiated standards and state performance levels was followed by an empirical exploration of the difference between the negotiated standards and actual state performance. The research examined the first three years of experience under the new management system (2000–02). This examination suggests that, rather than increasing the comparability of performance across states, the WIA system added new sources of arbitrary decision making, compromising the effectiveness of the measures as a tool for performance evaluation and improvement.

Some new measures under WIA have simply failed so far to prove useful. For example, the “soft” measures of customer satisfaction, intended to make program administrators more accountable to the primary customers of WIA services, proved disappointingly uninformative—in part because the questions were vaguely phrased. Nor did analysis find a consistent or significant relationship between these new measures and the objective measures of labor market outcomes.

How are performance measures being set?

The performance management system under WIA has been aptly described as a “high stakes game” with strong financial incentives. In order to be eligible for substantial bonuses, for example, states must achieve at least 80 percent of the negotiated performance level for each of the required measures. States that do not meet their performance goals two years in a row may be penalized. So far, there is no performance measure on which all states

have met 80 percent of their target (Table 2, columns 3–5). In PY 2000, only four states met their minimum requirements for all 17 measures. Furthermore, many states are at risk of sanctions: 38 states failed to achieve the 80 percent level for all measures for two consecutive years (not shown in Table 2). These results alone are sufficient to explain the great dissatisfaction that administrators have expressed with the new performance management system.

The WIA performance measures drew upon varied sources. Historical data from the JTPA were used in about half the states, but states also made use of projected national averages provided by the DOL, unemployment insurance data, and their own performance baselines from previous years. The process was complex and sometimes incomplete. States were instructed to take into account differences in economic conditions, participant characteristics, and services provided; for most, these adjustments were made informally during the review process.

Wisconsin, for example, used program year (PY) 1997 data and the projected national averages in negotiations with local officials. When the PY 1997 baseline was above the projected national averages, the latter were used as the target; when Wisconsin's baseline numbers were below the national averages, the state's baseline numbers were used. Indiana reported that it used PY 1999 data to determine the performance standards but did not have time for consultations with local workforce development officials, and presented only first-year goals in its five-year plan. Only two states (Texas and Maryland) and the District of Columbia reported using statistical models to determine performance standards.

Although one goal of the WIA performance management system was to standardize the types of performance data collected, the negotiation of performance standards clearly introduced substantially greater discretion and variability (Table 2, columns 1–2). Some state targets were above the national standards, some below, likely reflecting risk-balancing strategies such as those used in Wisconsin. Three states (California, Rhode Island, and North Carolina) established performance standards that were all below the national goals. But the differences in standards did not appear to systematically adjust (or account) for differences among states.

Analyses suggest that there were only two consistent associations between negotiated performance standards and participant characteristics. First, states with higher percentages of Hispanic and limited-English-proficiency populations had significantly lower performance targets for all adult, dislocated-worker, and youth performance measures. There were, indeed, strong, statistically significant, negative relationships between the performance levels states achieved and their percentages of Hispanic and limited-English-proficiency participants. Second, states that in 1998 had higher levels of unemployment set

significantly lower standards for the entered-employment rates of adults, dislocated workers, and older youth and for the employment retention rates of younger youth. But although states had thus “hedged their bets,” the state unemployment rate was strongly linked to failure to meet the target for several of these measured standards (for example, the entered-employment rate of older youth).

This particular finding also directs attention to a serious failing in the performance system: the expectations of continuous improvement built into the annual ratcheting up of performance standards. In each of the three program years, both national goals and state standards set higher absolute levels of performance requirements for nearly all measures—this in a time of economic recession and rising unemployment, when downward adjustments, if anything, were needed to account for changing local conditions in all states. In the absence of an adequate process for adjusting these standards, program managers appear to have made undesirable accommodations after the fact, for example, by restricting access to participants they deemed likely to fail. “Cream-skimming” had been reported under the JTPA. A study of WIA suggests that history may be repeating itself.¹⁷

The difficulties described in this analysis of the WIA system are characteristic of the challenges that public managers face in trying to design and implement outcomes-based performance management. Particularly in social programs, it is difficult if not infeasible to attempt to distinguish the contributions of program services and management from the influence of other local factors. And when the stakes are raised, it is hardly surprising that public managers might turn to counterproductive means of achieving higher levels of measured performance at the expense of other program goals. Technological improvements—better computer software and data collection—are only part of the solution. As both this WIA study and the JTPA analyses summarized earlier in the article suggest, major questions of design and goal-setting remain. ■

¹B. Radin, *Beyond Machiavelli: Policy Analysis Comes of Age* (Washington, DC: Georgetown University Press, 2000), p. 168.

²W. Edwards Deming, *Out of the Crisis* (Cambridge, MA: MIT Institute for Advanced Engineering Study, 1986).

³R. Behn, *Rethinking Democratic Accountability* (Washington, DC: Brookings Institution, 2001); C. Lindblom, “The Science of Muddling Through,” *Public Administration Review* 19, no.1 (1959): 79–88.

⁴J. Heckman, C. Heinrich, and J. Smith, “The Performance of Performance Standards,” *Journal of Human Resources* 37, no. 4 (Fall 2002): 778–811.

⁵Radin, *Beyond Machiavelli*; H. Mintzberg, “Managing Government, Governing Management,” *Harvard Business Review* 74 (May/June 1996): 75–83; B. Peters, “Policy Instruments and Public Management: Bridging the Gaps,” *Journal of Public Administration Research and Theory* 10, no. 1 (2000): 35–48.

⁶On cream-skimming, see especially Heckman, Heinrich, and Smith, “The Performance of Performance Standards.”

⁷U.S. Government Accountability Office (formerly U.S. General Accounting Office), *Managing for Results: Opportunities for Continued Improvements in Agencies' Performance Plans*, GAO/GGD/AIMD99-215, 1999.

⁸U.S. Government Accountability Office (formerly U.S. General Accounting Office) has published three reports: *Managing for Results: Prospects for Effective Implementation of the Government Performance and Results Act*, GAO/T-GGD-97-113, 1997; *Managing for Results: Analytic Challenges in Measuring Performance*, GAO/HEHS/GGD-97-138, 1997; *Managing for Results: Opportunities for Continued Improvements*.

⁹The research that I discuss here has been reported at length in C. Heinrich, "Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness," *Public Administration Review* 62, no. 6 (November/December 2002): 712–25 (on JTPA; IRP thanks *Public Administration Review* for permission to use these materials); "Improving Public Sector Performance Management: One Step Forward, Two Steps Back?" forthcoming, *Public Finance and Management* (on WIA). A more general article discussing these issues is C. Heinrich, "Measuring Public Sector Performance and Effectiveness," in *Handbook of Public Administration*, ed. B. Peters and J. Pierre (Thousand Oaks, CA: Sage Publications, 2003).

¹⁰H. Bloom, L. Orr, G. Cave, S. Bell, and F. Doolittle, *The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months* (Bethesda, MD: Abt Associates, 2003).

¹¹C. Heinrich and L. Lynn, Jr. "Governance and Performance: The Influence of Program Structure and Management on Job Training Partnership Act (JTPA) Program Outcomes," in *Governance and Per-*

formance: New Perspectives, ed. C. Heinrich and L. Lynn, Jr. (Washington, DC: Georgetown University Press, 2000).

¹²C. Heinrich, "Do Government Bureaucrats Make Effective Use of Performance Management Information?" *Journal of Public Administration Research and Theory* 9, no. 3 (1999): 363–93; J. Heckman, C. Heinrich, and J. Smith, "Assessing the Performance of Performance Standards in Public Bureaucracies," *American Economic Review* 7, no. 2 (1997): 389–96; R. Kravchuk and S. Schack, "Designing Effective Performance Measurement Systems under the Government Performance and Results Act of 1993," *Public Administration Review* 56, no. 4 (1996): 348–58.

¹³For example, L. Orr, H. Bloom, S. Bell, W. Lin, G. Cave, and F. Doolittle, *The National JTPA Study: Impacts, Benefits, and Costs of Title IIA* (Bethesda, Md: Abt Associates, 1994).

¹⁴C. Hill, "Impacts, Outcomes, and Management in Welfare-to-Work Programs," Ph.D. dissertation, University of Chicago, 2001 (p. 4).

¹⁵See, e.g., P. Courty and G. Marschke, "An Empirical Investigation of Gaming Responses to Performance Incentives," *Journal of Labor Economics* 22, no. 1 (January 2004): 23–56.

¹⁶U.S. Government Accountability Office (formerly U.S. General Accounting Office), *Workforce Investment Act: Improvements Needed in Performance Measures to Provide a More Accurate Picture of WIA's Effectiveness*, GAO Report 02-275, 2002.

¹⁷U.S. GAO, *Workforce Investment Act: Improvements Needed*.