Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness

Author(s): Carolyn J. Heinrich

Source: *Public Administration Review*, Vol. 62, No. 6 (Nov. - Dec., 2002), pp. 712-725
Published by: Wiley on behalf of the American Society for Public Administration
Stable URL: http://www.jstor.org/stable/3110329
Accessed: 20-12-2015 06:38 UTC

Carolyn J. Heinrich
University of North Carolina–Chapel Hill

Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness

Requirements for outcomes-based performance management are increasing performance-evaluation activities at all government levels. Research on public-sector performance management, however, points to problems in the design and management of these systems and questions their effectiveness as policy tools for increasing governmental accountability. In this article, I analyze experimental data and the performance-management experiences of federal job-training programs to estimate the influence of public management and system-design factors on program outcomes and impacts. I assess whether relying on administrative data to measure program outcomes (rather than impacts) produces information that might misdirect program managers in their performance-management activities. While the results of empirical analyses confirm that the use of administrative data in performance management is unlikely to produce accurate estimates of true program impacts, they also suggest these data can still generate useful information for public managers about policy levers that can be manipulated to improve organizational performance.

Introduction

The plethora of idioms and acronyms for performance-management initiatives—planning, programming and budgeting, performance-based budgeting, pay-for-performance, performance planning, total organizational performance system, management by objectives, and more—impede a facile understanding of how and why we measure public-sector performance. Yet as conceptions, designs, and methodologies for performance measurement continue to evolve, a single, central purpose of these initiatives has been unchanging: to improve public management and program outcomes.

The tools of performance management—and public expectations for their usefulness—have been growing in sophistication. Early performance-measurement efforts, including planning, programming, and budgeting in the 1960s, the Nixon administration's management by objectives, and zero-based budgeting in the 1970s were more internal and process focused. Romzek (1998) describes these early approaches to measurement as *hierarchical accountability for inputs* (administrative rules guiding routine tasks and budgetary allocations) and *legal accountability for processes* (audits, site visits, and other monitoring tasks). Appraisals of government performance were concerned primarily with assessing the relationship of inputs to costs and the value of cost-reduction activities in these systems, adapting techniques from the larger field of management science. Hollings (1996, 15) characterizes these types of activities as "performance auditing," aimed at pointing out breakdowns in operational controls and the implementation of functional responsibilities and areas for cost reduction and operating improvements.

The expansion of block grant programs in the late 1970s, followed by the Reagan administration's New Federalism in the early 1980s, shifted considerable responsibility and discretion in the management of public

Carolyn J. Heinrich *is an assistant professor in the Department of Public Policy at the University of North Carolina–Chapel Hill. She holds a doctorate in public policy studies from the University of Chicago. Her research interests traverse the fields of public management, labor economics, and social program evaluation. Recent publications include* Improving Governance: A New Logic for Empirical Research *(Georgetown University Press) with Laurence E. Lynn, Jr. and Carolyn J. Hill, and articles in the* Journal of Public Administration Research and Theory *and* Journal of Policy Analysis and Management. ***Email: cheinric@email.unc.edu.***

programs to the state and local levels of government. Included among these programs was the Job Training Partnership Act (JTPA) program, a $5 billion federal employment and training program established in 1982, with a highly decentralized administrative structure and formal participation of private-sector representatives in local program administration. At the same time, concerns about how local officials, distant from federal auditors, were exercising discretion in the distribution and management of program monies provided the impetus for the introduction of new mechanisms for performance accountability. As DeNisi (2000, 131) explains, all performance-management efforts ultimately are aimed at influencing performance at the highest level of organization; yet, as the distance from the individual level to the highest level increases, the links between individual and organizational behavior and performance grow more complex.

The JTPA performance standards system, described by Barnow (2000, 119) as one of the "pioneers in performance management," was a prominent example of performance accountability in this new era of decentralization. This system was distinctive from earlier systems in that it (1) centered performance measures on program *outcomes* (for example, the number of trainees placed in jobs) rather than on inputs or outputs (the number of persons trained); (2) linked measures of program performance across multiple levels of government; and (3) included financial (budgetary) incentives for program managers based on evaluations of organizational *outcomes*. More generally, the system was designed to focus management attention (at all levels) on the central organizational objectives (or program outcomes) and lessen the government's need for costly process and compliance monitoring. Romzek (1998, 204) describes this type of approach as one that relies on *professional accountability,* deferring to the discretion of managers "as they work within broad parameters, rather than on close scrutiny to ensure compliance with detailed rules and organizational directives."

Performance management was also being transformed in the 1980s by advances in statistical techniques for measuring program performance. States experimenting with welfare reform, the National JTPA Study's experimental and nonexperimental evaluations, and other quantitative evaluations of government programs were growing in number as experimental and administrative data were being collected or made available to analyze public program outcomes and impacts.[1] The National JTPA Study, for example, compared the earnings outcomes of trainees with those of a randomized control group to identify the value added by the program—that is, the difference between what trainees earned and what they would have experienced without the program. The program *impact* findings identified shortcomings with the outcome measures used in the JTPA performance-management system and prompted additional discussion about how performance measures in public programs might be improved.

The National Performance Review reports, along with the Clinton administration's early pledge to find out "what works, and what doesn't work" among the multitude of federal programs with overlapping objectives and target populations, further encouraged this evaluative approach to performance management. The Government Performance and Results Act of 1993 made formal, outcomes-based performance evaluations mandatory for federal programs. It continued and broadened the federal government's efforts to realign the focus of government accountability and performance analysis away from activities and process measures and toward results or outcomes. Specifically, the act requires federal agencies to develop (1) a strategic plan that specifies agency goals and how they will be achieved; (2) an annual performance plan that specifies quantitatively measurable goals and performance indicators, as well as levels of performance to be achieved; and (3) an annual program performance report that compares actual performance with performance goals. The requirement of program performance reports was intended to provide *political accountability* for results and the opportunity for increased responsiveness to program stakeholders and constituencies.

An important question that arises for public managers and researchers is, are outcomes-based performance-management systems more effective than traditional approaches to bureaucratic control (that is, accountability for inputs and processes)? In other words, do the quantitative evaluations of program outcomes induce agencies to change program priorities and primary work processes in ways that improve or positively influence program results? This implies that the information obtained through performance measurement is used to inform program managers at multiple levels of organization (from federal administration to primary work). As DeNisi (2000, 131) implores, "regardless of the level at which we want to influence performance, we must do so by influencing the behavior of individuals."

While it is too early to come to a decisive judgment regarding the effectiveness of Government Performance and Results Act as a policy tool for public accountability, some have harshly criticized its design and implementation. Radin (2000, 133), for example, argues that the act's "use of administrative rhetoric has caused it to collide with institutional, functional, and policy/political constraints that are part of the American decision-making system. The resulting collisions have bred a sense of cynicism and a compliant attitude within the federal government." Radin's analysis of the act's implementation suggests that requirements for specific performance goals, plans, and results have increased administrative constraints, elevated conflict

among multiple levels of program management, and ignored the political complexities that have complicated the act's implementation. Mintzberg (1996) identifies some of the same impediments to administrative reform through outcomes-based performance management, citing the absence of strong leadership or coalitions supporting a results orientation, measures that constrain flexibility and are not well linked to goals or consequences, mutual distrust between agencies and legislators about gaming of measures, and employee concerns that their responsibility is not commensurate with their authority.

In his discussion of policy tools and public management, Peters (2000, 35) suggests that policy instruments (such as performance-management systems) "are rarely capable of being effective in implementing policy without adequate management." This motivates a second research question about the implementation of performance-management systems in the public sector: If imperfect information and political complexities are inevitable in outcomes-based performance-management systems, with adequate management, can they still provide effective guidance and a means of bureaucratic control that improves program results? While research on public-sector performance management describes shortcomings in both the design and management of these systems (Courty and Marschke forthcoming; GAO 1997a, 1997b, 1999; Barnow 2000; Heckman, Heinrich, and Smith 1999; Heinrich 1999; Radin 2000), understanding the crux of these problems is made more complex by the fact that administrative discretion at various levels of organizational operations sometimes blurs the distinction between system design and management responsibilities.

In this article, I address questions about the design and management of public-sector performance-management systems and their effectiveness as policy tools for increasing governmental accountability, drawing from the experience of federal job-training programs (in particular, the JTPA program). First, I discuss some of the challenges and trade-offs that policy makers and program administrators face in developing accurate performance measures that reflect legislative, agency, and other stakeholder goals. Next, I present data analyses, using data from the National JTPA Study experimental evaluation, which estimate the influence of performance standards system design and management and contextual factors on program outcomes and impacts. In these empirical analyses, I assess whether relying on administrative data to measure outcome *levels* (rather than program *impacts)* produces information that is likely to misdirect JTPA program managers away from the goal of generating lasting earnings gains for participants.

I find it is useful and important to assess the influence of organizational structure, policy choices and constraints, and service-delivery practices on public program performance. The relationships observed among these organization- and site-level variables and program outcomes were consistent with those estimated using experimental data in models of program impacts. The concluding section of this article considers the implications of these study findings for current performance-management systems and the design of future systems in government.

Outcomes-Based Performance Management Systems in the Public Sector

Research on performance management suggests that, in responding to the requirements of Government Performance and Results Act, federal agencies should choose performance measures that (1) are closely aligned with their stated goals; (2) approximate actual performance as closely as possible; (3) are relatively simple and inexpensive to administer; and (4) make it difficult for managers to increase their measured performance in ways other than increasing their actual performance (Baker 1992, forthcoming; GAO 1997b, 1999; Hart 1988; Holmstrom and Milgrom 1991; McAfee and McMillan 1988). The JTPA program experience shows, however, how difficult it is to develop and implement a performance-management system with these qualities. When multiple or conflicting goals motivate employees, when organizational goals and performance measures diverge, or when bureaucratic effort across government levels is not readily observed, problems in performance-management systems are likely to arise (Kravchuk and Schack 1996).

Accordingly, Murphy and Cleveland (1995) urge that, in addition to specifying accurate measures of performance, managers should give more attention to contextual factors that influence or interfere with performance measurement and may undercut objectives to improve accountability and organizational performance. The contextual factors they identified in a study of private-sector organizations include organizational complexity and coordination; organizational climate or culture and values; competition among functional units or within sectors; and general economic and political conditions. The importance of these factors may be magnified in public organizations by frequently changing political and administrative priorities, professional and partisan conflict within bureaucracies, and the sometimes precarious links across government levels and between formal and informal authorities in program implementation.

Therefore, while every public organization is likely to have a unique experience in implementing performance-management systems, the long tenure of performance standards systems in federal job-training programs and data-collection activities associated with the National JTPA

Study have generated some of the best empirical data for considering the implications of performance management in public programs. I begin by briefly describing the JTPA program and performance standards system and follow with a discussion of some of the primary problems and challenges confronted in developing an outcomes-based performance-management system in federal job-training programs.

Government Accountability in Federal Job-Training Programs

The JTPA program and its successor, the Workforce Investment Act, are administered by the U.S. Department of Labor, which combines federal funding and broad oversight with the delegation of most operational and monitoring responsibilities of job-training programs to state and local agencies.[2] Job search and placement assistance, counseling, job-readiness activities, case management and supportive services, and other more intensive services such as vocational training, on-the-job-training, and basic education are provided to increase earnings and reduce welfare dependency among low-income individuals. Federal funds are allocated to states in proportion to the size of their unemployed and economically disadvantaged resident populations.[3]

The performance standards system for federally funded employment and training programs (the JTPA program and the new Workforce Investment Act program) divides operational responsibilities between federal, state, and local levels. It is a multilevel system that allows states and localities considerable discretion to shape their own performance-incentive policies and training-service strategies to achieve performance goals (figure 1). At the federal level, the Labor Department determines a set of core performance measures that, in effect, establish the broad parameters of public expectations for government performance. States are allowed to modify and add to the federal performance standards and develop their own systems for monitoring and rewarding (or sanctioning) local job-training agencies following annual performance reviews. A fixed percentage of the federal government's appropriations to the states is designated for bonuses or budgetary awards, and states define how measured performance translates into budgetary increases for local job-training agencies. Courty and Marschke (forthcoming) analyze in considerable detail the various performance-incentive and award plans developed by states to maintain accountability for local agency performance. Their research suggests the different strategies have had an important effect on the marginal incentives generated for increasing performance above the standards and the extent to which program administrators can determine in advance benefits associated with achieving particular levels of performance.

Figure 1 JTPA Program Administration and Service Delivery

Federal government	**U.S. Department of Labor** • Distributes federal employment and training funds to states • Establishes and transmits performance standard requirements • Monitors state-level JTPA operations
State government	**State government employment and training bureaucracy** • Establishes state-level target population goals and performance standards based on federal guidelines • Develops models to measure JTPA program outcomes and maintains records on all JTPA program participants • Distributes federal funds and any additional state employment and training funds and incentive monies to local service-delivery areas • Translates and develops regulations on the expenditure of funds by local agencies and monitors local agency operations
Local service-delivery areas	**Local administrative agencies and/or private industry councils/workforce development boards** • Sets target population and performance goals in conjunction with state-level requirements • Develops training plans and distributes funds through contracts with service provider organizations to deliver program services and also provides services directly through local offices • Monitors service providers and evaluates program outcomes

The performance measures mandated by the Workforce Investment Act are similar to those used in JTPA programs and include (1) participant program-completion rates; (2) participant job-placement rates (unsubsidized employment); (3) participant wages at placement; (4) retention rates at six months after job placement; (5) wages received at six months after job placement; (6) licenses or certifications, attainment of academic degrees, and other skills-attainment measures; and (7) measures of participant program costs. An important difference in the performance standards used is that the new Workforce Investment Act focuses on participant outcomes measured six months following placement, whereas the JTPA performance standards system focused solely on shorter-term outcomes (measured at termination or 90 days after program completion).

At the local level, program managers maintain records to generate performance data based on the performance standards described above. These data are reported to the state agencies charged with program implementation. States may adjust the locally reported performance levels for local economic conditions and client characteristics. The states subsequently report this performance information to the federal government (Department of Labor), but the states are responsible for determining performance bonuses or sanctions for local agencies based on their performance.

Heinrich (1999) describes how some local service-delivery areas also develop their own systems of performance accountability for their contracted providers and use performance-management information in their contracting decisions. Contrary to Kravchuk and Schack's (1996, 353) contention that performance measures in these types of multilayered, intergovernmental systems "diverge widely," she finds that the *basic* performance standards in JTPA programs were fairly consistent across government levels. However, as Courty and Marschke (forthcoming) and Heckman, Heinrich, and Smith (1997) observe, it was the *marginal* incentives for increasing performance, generated by administrators' decisions about *how to use the performance information in program management* (directing, rewarding, or sanctioning agencies and providers) that differed substantially across state and local levels.

Problems in Implementing Outcomes-Based Performance Management in Federal Job-Training Programs

In a recent review of federal agencies' performance plans, the General Accounting Office (1999, 6) states that "mission fragmentation" is widespread in the federal government, and that it is difficult to get stakeholders to think beyond their own program operations to how diverse activities are related to a common outcome. As in the JTPA program, sometimes the multiplicity or conflict and fragmentation among goals is inherent in the originating legislation. The 1982 JTPA legislation, for example, stated that the program should serve "those who can benefit from, and are most in need of" employment and training services, implying several different goals (Job Training Partnership Act, P.L. 97-300, Section 141(c), 1982). One goal of the government might be to maximize the impact or total net benefits from the program, pertaining to the part of the legislative directive "those who can benefit from...." Another distinct goal would be to target those "most in need" of services to achieve equity, where the government would target those with low mean earnings in the absence of the program for participation.

Even in this simple interpretation of the intended goals, it is apparent that conflict in implementing the legislative directive is possible, particularly if those with low mean earnings in the absence of the program are unlikely to achieve higher earnings with participation. In fact, Heckman and Smith (1995) confirm a trade-off between equity and efficiency goals in the JTPA program. They find that targeting the highly disadvantaged (the bottom 20 percent of the skill distribution) substantially decreased the social efficiency of the program (the achievement of larger value-added gains through program participation).

Furthermore, state and local agencies may be interested in pursuing some combination of these and other goals, including objectives related to the total number served, targeting services to particular groups, geographic distribution of services, and meeting political obligations (Dickinson et al. 1988; Kravchuk and Schack 1996). The purpose of an outcomes-based performance standards system, of course, is to focus management and staff attention on those ends deemed most important—for example, emphasizing earnings impacts over output goals, such as the total number served, or ancillary goals, such as fulfilling political commitments.

After determining primary performance goals, the subsequent challenge is to identify appropriate performance measures. Federal managers interviewed about this process report that it is both conceptually and practically one of the most difficult tasks in developing outcomes-based performance-management systems. In JTPA programs, for example, using a measure of the average earnings gain or impact of participation would require information about what would have happened to participants had they not received training, as generated in the National JTPA Study. In fact, at least one General Accounting Office report (1997b) suggests supplementing performance data with impact-evaluation studies to better account for external influences on program results and to obtain a more accurate picture of program effects. The costs of experimental evaluations, however, prohibit the collection of these types of data on a regular basis.

In addition, it is important that performance standards systems provide timely feedback to state and local program managers—such as the annual program performance reports under the Government Performance and Results Act—to allow for adjustments (in budget allocations, service contracts, management practices, training strategies, etc.) to improve performance. For this reason, performance standards systems commonly rely on short-run rather than long-run measures. Federal agencies engaged in performance management have found it especially difficult to translate their long-term missions or strategic goals into annual performance goals and to predict the level of results that might be achieved over a shorter term (GAO 1997a). In the JTPA and Workforce Investment Act programs, performance measures are earnings *levels* in the year following participation, generated from routinely collected administrative data, rather than outcomes or *impacts* over a longer period.

Heckman and Smith (1995) point out that to be effective, performance measures should be strongly correlated with program goals in order to provide the right marginal incentives to program managers and staff. In their JTPA research, Heckman and Smith (1995) and Barnow (2000) show that performance measured according to the short-term standards used in the JTPA (participant employment rates and earnings levels 13 weeks following discharge from

the program) is, at best, weakly (and sometimes negatively) related to longer-term employment and earnings impacts estimated for participants using experimental data. Further exploration of the relationship between some alternative measures aimed at strengthening the link between measured performance and program impacts, including earnings levels measured at six months following program participation, suggests that improvements are unlikely, as long as empirical measures of performance continue to be based on earnings *levels* of participants rather than on earnings *gains* (Gay and Borus 1980; Heckman, Heinrich, and Smith 1999). Heckman, Heinrich, and Smith (1997) suggest the divergence between empirical performance measures and program impacts in the JTPA might unintendedly encourage a focus that actually undercuts progress toward the socially desirable goals of generating lasting earnings gains for participants and reducing welfare dependency.

Courty and Marschke (1997) generate empirical evidence that shows how some local program managers chose to "game" the performance standards system to increase their agency's measured performance in ways independent of its actual performance. Since the JTPA performance standards measured employment and earnings shortly after a participant's discharge from the program, by strategically timing the date that termination was recorded, program managers were sometimes able to increase the total number of job placements. The authors estimate that by managing participant terminations in this way, JTPA agencies were able to improve their measured performance by as much as 11 percentage points; they also found, however, smaller training impacts for participants who received training during the end of the year when these types of gaming activities were likely to occur.

Program managers' desire to generate incentives or manipulate a set of performance measures in ways that will enhance their *measured* performance is understandable, if not logical, given the limited or frequently indirect influence the government has in determining program results. This is particularly problematic for programs that attempt to affect highly complex systems or social phenomena, such as the employment trajectories of welfare recipients and other economically disadvantaged adults, which are largely outside of government control. In addition, this problem is exacerbated by the challenges of separating the efforts and effects of multiple layers or hierarchical levels of government policy, particularly in programs such as the JTPA, where state and local management decisions shape performance-management policies and service-delivery practices appreciably. In light of these limitations, the most useful feedback that program managers might receive from a performance-management system would be that which increases their understanding of how their own policy and programmatic decisions are linked to program outcomes. The next section presents a discussion and empirical analysis of the potential for using performance-management information specifically for this purpose.

Governance and Performance Management with Limited Information

Over nearly two decades of operations, considerable differentiation in the governance and management of federal job-training programs has evolved at the local level, reflecting factors such as state-level administrative policies, the size of the local service-delivery area or jurisdiction, the size and composition of the eligible population, urban versus rural location, and local political preferences. Heinrich and Lynn (2000) study the different management structures and performance-incentive policies in service-delivery areas that participated in the National JTPA Study, focusing on how the structure and form of JTPA program administration relate to management choices about the use of performance-incentive policies. Some of the key areas of administrative discretion (table 1) include minimum performance requirements and performance bonus award schemes, weights accorded to different performance goals, and service-delivery strategies and contracting practices.

Table 1 Key Areas of Administrative Discretion in the JTPA Performance Standards System

- *Minimum performance requirements and performance bonus award schemes*: The stringency of performance requirements instituted by states differs in a number of ways, such as the minimum number of performance standards a local agency must meet to qualify for incentive bonuses (or to be classified as "failed to meet"), the *level* of performance at or above state performance standards that it must attain to qualify for incentives, and other policies affecting the size of performance bonuses. Many states encourage competition among local service-delivery (or workforce investment) areas by making performance bonuses received by agencies contingent on their performance relative to other areas.
- *Weights accorded to different performance goals*: In addition to the actual performance levels required, the weights accorded by states to the core (federal) performance standards in determining bonuses also varies. Many states and localities establish additional performance standards for services to disadvantaged groups.
- *Service-delivery strategies and contracting*: Administrative entities vary in the extent to which they provide training services directly to participants or contract or form partnerships with local service providers. Some local agencies using performance-based contracts develop their own performance-incentive systems, including competitive bidding and performance-review processes to promote competition.

Heinrich and Lynn find strong associations between the choices of administrative structures in service-delivery areas and the types of policies and incentives adopted at both state and local levels to motivate performance. For example, when private-sector representatives assumed more formal authority through the role of private industry councils as administrative entities and recipients of federal job-training funds, they appeared to emphasize measured performance and adopted administrative prac-

tices (such as performance-based contracting and contracts with for-profit organizations) that also demanded accountability for results. In addition, service-delivery areas with a larger administrative role for local public officials and less control by private-sector representatives delivered more JTPA program services directly (versus contracting out) and were less likely to use performance-based contracts. These administrative entities also had more explicit incentives to focus on "hard-to-serve" groups and were less likely be in fervent competition with other agencies for performance bonuses.

Case studies of the JTPA program also describe how differing managerial approaches have affected the implementation and influence of performance-incentive policies. Heckman, Heinrich, and Smith (1997), for example, compare their findings from two studies of JTPA service-delivery areas that adopted contrasting administrative approaches. In one agency, the goal of exceeding performance standards was emphasized in all aspects of program operations, including contracting, participation selection, and service delivery. In the other, performance standards were given minimal attention and case workers' social service norms guided decision making in these key program processes. Although comparable data were not available from both sites to establish a conclusive link between performance-management practices and outcomes, the agency with a strong emphasis on performance standards consistently exceeded its performance requirements, while the other site failed to meet minimum performance requirements in some years and was at risk for federal reorganization (the strongest penalty for poor performance).

Empirical Models of JTPA Program Outcomes and Impacts

Because administrative decisions and management actions *at different organizational levels* have the potential to influence not only program outcomes, but also the specific types of performance-management policies adopted, disentangling the effects of particular policy or performance incentive is analytically challenging. In addition, understanding the effects of performance-management policies on program outcomes (JTPA participants' earnings *levels* following participation) may be less useful if, in aiming to maximize *measured* performance, program managers have a negative influence on participants' earnings *impacts*.

This section brings empirical evidence to bear on the question of whether performance evaluations based on limited data, such as those available through government administrative records, can generate enough information to effectively guide program administrators and policy decisions. I use information about administrative structures and performance standards policies (table 1), along with the JTPA experimental data, to examine the relationships between program administration and performance-management policies and the outcomes and impacts of JTPA programs. In addition, I investigate how the availability and use of information about program *impacts* in performance management affects what program managers and policy makers learn from performance-management systems. In other words, by relying on administrative data to assess outcome *levels* rather than *impacts,* does the JTPA performance-management system produce information that is likely to misguide program managers, or direct them away from the goal of generating lasting earnings gains for participants?

Because performance management involves activities and interactions that span multiple levels of organization or systemic structures, I use multilevel modeling[4] in the empirical analyses. Multilevel modeling allows one to test hypotheses about how factors or variables 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 data include *individual-level* data (approximately 20,000 total sample observations) on JTPA treatment and control group member demographic characteristics and earnings histories, and *site-level* data that describe the administrative structures, performance-incentive policies, service-delivery and contracting strategies, and the unemployment rate and regional indicators in the 16 service-delivery areas across the three study years. (The data and data sources are further described in appendix A). Although there are only 16 sites from 16 states in the National JTPA Study (NJS), there is significant year-by-year variation in the state- and site-level variables, including the performance-incentive policies, as a result of major policy changes during the 1987–89 study years. This variation facilitates analyses by both site and year and triples the number of observations at this level (n = 48).

Although less than 5 percent of the *total* variation in one-year earnings outcomes and impacts is between NJS sites, unconditional models of JTPA program outcomes and impacts confirm this variation is substantive and statistically significant.[5] A model controlling for individual characteristics of treatment and control group members (without site-level controls) also verified a statistically significant, *adjusted* average treatment impact of $1,465.

In the multilevel models of individuals' earnings outcomes and impacts in the first post-program year, each level in the hierarchical data structure is formally represented by its own submodel, and each submodel specifies the structural relations occurring at that level. The *outcome* model (shown in equations 1 and 2) is estimated using data only for those who received JTPA program services (the NJS treatment group), that is, data typically available to program managers.

$$(1)\ Y_{ij} = \beta_{0j} + \beta_{1j}X_{1j} + \ldots + \beta_{nj}X_{nj} + r_{ij} \qquad \text{(level 1 submodel)}$$

$$(2)\ \beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \ldots + \gamma_{0n}W_{nj} + \lambda_{01}\Psi_{1j} + \ldots + \lambda_{0n}\Psi_{nj} + u_j \qquad \text{(level 2 submodel)}$$

Y_{ij} measures earnings in the first post-program year; β_{0j} is the average earnings outcome,[6] adjusted for the observable differences in individual characteristics (X_{1j} to X_{nj}); β_{1j} to β_{nj} are the average effects of individual characteristics on earnings; W_j is the site-level variable describing administrative structures, performance-incentive policies, and contracting practices; Ψ_{1j} to Ψ_{nj} control for regional and economic conditions at the sites; γ_{00} is the adjusted average outcome across sites and years); and γ_{01} to γ_{0n} and λ_{01} to λ_{0n} are the effects of site-level variables on the adjusted average earnings outcome. Although one may allow the effects of individual characteristics to vary across sites and years, in this specification, these effects are assumed to be fixed across sites and years, $\beta_{1j} = \gamma_{10} \ldots \beta_{nj} = \gamma_{n0}$. The submodels are estimated simultaneously using maximum-likelihood methods. (Appendix B includes additional discussion of the methodology and models.)

In modeling the JTPA program impact on earnings in the first post-program year, data on both the NJS treatment and control group members is used, and there are some important differences in these submodel specifications of the multilevel model (equations 3 and 4).

$$Y_{ij} = \beta_{0j} + \beta_{1j}Dij + \beta_{kj}X_{2j} + \ldots + \beta_{nj}X_{nj} + \delta_{kj}X_jDij + \ldots + \delta_{nj}X_{nj}Dij + r_{ij} \qquad \text{(level 1 submodel)}$$

$$(3)\ \beta_{0j} = \lambda_{00} + \lambda_{01}\Psi_{1j} + \ldots + \lambda_{0n}\Psi_{nj} + u_j \qquad \text{(random intercept)}$$

$$(4)\ \beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + \ldots + \gamma_{1n}W_{nj} + v_{0j} \qquad \text{(random coefficient)}$$

The level 1 submodel includes an indicator variable, Dij, that equals 1 if an individual is a member of the treatment group; the coefficient on this indicator, β_{1j}, is the conditional treatment impact at site j for the average treatment group member. Because randomization was flawed in the NJS experimental evaluation, an additional set of covariates interacted with the treatment group indicator, X_jDij to $X_{nj}Dij$, are included to measure the differential effects of individual characteristics (δ_{kj} to δ_{nj}) on treatment group impacts. At level 2, the random intercept model measures the effects of regional and economic conditions (λ_{01} to λ_{0n}) on the adjusted average outcome in site j, and the random coefficient model measures the effects of administrative structures, performance incentive policies, and contracting practices (γ_{11} to γ_{1nj}) on conditional treatment impacts. As in the outcome model, the relationships between the other level 1 covariates and site-level explanatory variables are assumed to be fixed.

Model Findings and Implications for Performance Management

The results of the multilevel models of JTPA program participant earnings outcomes and impacts are shown in table 2. The direction of the effects of individual-level predictors in these models are all consistent with prior JTPA research (Orr et al. 1994, 1997).[7] For example, a congruent finding in the literature is that those with less than a high school degree have a poorer post-program labor market experience. This relationship is confirmed in both models, showing that those with less than a high school degree earn more than $1,700 less in the first post-program year, and the treatment impact is more than $1,100 less for this subgroup. In the impact model, only the *differential* (not the average) effects of individual characteristics on treatment group impacts are shown. (See appendix B for further discussion of the interpretation and comparison of the level 1 model coefficients.)[8]

The comparison of estimated effects of structural, policy, and management factors on 12-month post-program earnings outcomes and impacts (also shown in table 2) is of primary interest in this study. These site-level controls (including those for regional and economic conditions) explain over 90 percent (95 percent and 91 percent, respectively) of the site-level variation in outcomes and impacts. The adjusted average treatment impact is reduced to $673 with the inclusion of site-level controls.

Most notably, however, are the fairly consistent relative sizes and statistical significance of the coefficients on these variables in outcome and impact models. For example, with either set of results, policy makers should reach the conclusion that when private-sector representatives assume more formal authority through the role of the private industry councils as administrative entities of federal job-training funds, significantly higher earning levels and impacts are realized, on average, by participants. They also earn more when private industry councils do *not* share these responsibilities as an equal partner with local elected officials. In addition, the weight given to the entered employment rate standard—one of the most prominent performance standards in the JTPA program—is positively and significantly related to both participants' earnings outcomes and impacts (1 percent more weight on this goal increases adjusted average earnings about $180 and adjusted average impacts by about $94 per year). Likewise, Bloom, Hill, and Riccio (2001) find that greater emphasis on quick job entry in welfare-to-work programs has a large, statistically significant effect on program earnings impacts. Policy directives that require administrative entities to use performance bonuses to serve the highly disadvantaged, more direct service provision by the administrative entity, and a higher proportion of performance-based contracts are all negatively related to both adjusted average earnings outcomes and impacts.

In general, the findings suggest that although there are some differences in the estimates of structural, policy, and management effects on participant earnings based on the

Table 2 Hierarchical Linear Models of JTPA Program Earnings Outcomes and Impacts

Predictors: Level 1 (individual level)	Earnings in first post-program year	Earnings in first post-program year (impact model) differential effects for treatment group[a] and earnings impact[b]
Intercept	1595.76 (1.04)	5601.25 (9.15)*
Gender (1=male)	2165.91 (7.81)*	61.97 (0.18)
Age 22–29 years	1055.41 (3.16)*	1189.90 (2.75)*
Age 30–39 years	406.56 (1.07)	495.09 (1.01)
Age 40 and over	–228.27 (–0.48)	672.23 (1.15)
Black	–1117.12 (–3.03)*	–459.68 (–1.07)
Hispanic	-741.96 (-1.74)	–526.63 (–0.94)
Other race	–1661.77 (-2.28)*	–1347.71 (–1.32)
Married	1185.11 (3.35)*	1211.61 (2.69)*
Divorced, widowed, or separated	734.66 (2.18)*	764.39 (1.75)
No high school degree	–1728.94 (–6.00)*	–1126.18 (–3.07)*
Some post–high school education	1218.72 (3.58)*	37.22 (0.09)
Welfare recipient at time of application	–1027.09 (3.74)*	217.27 (0.63)
Children under age 6	222.29 (0.74)	–18.93 (–0.05)
Employment—unemployment transition in year before enrollment	–856.75 (–3.28)*	–471.06 (–1.39)
Earnings in year before enrollment	0.32 (10.47)*	–0.10 (–2.52)*
Treatment group indicator	n.a. 673.50	(1.49)
Level 2 (site level) fixed effects		
Southern region	1726.65 (1.84)	971.75 (2.64)*
Midwestern region	2345.37 (3.45)*	1180.56 (3.21)*
Western region	3696.25 (2.89)*	647.37 (1.49)
Unemployment rate	602.51 (3.57)*	–1900.49 (–0.27)
Level 2 (site level) structural, policy, and management effects[c]		
Private industry council is the administrative entity	2063.25 (5.13)*	1549.12 (3.67)*
Private industry council and local/chief elected official are equal partners	–2151.11 (–2.67)*	–414.41 (–0.96)
Percentage of services provided directly by administrative entity	–4460.69 (–3.20)*	–621.67 (–0.47)
Percentage of performance-based contracts	–2458.07 (–2.28)*	–1542.19 (–1.95)
Weight accorded to employment rate standard	180.85 (4.11)*	94.49 (2.31)*
Weight accorded to cost per entered employment rate	–26.06 (–1.27)	7.55 (0.42)
Minimum number of standards that sites must meet to qualify for performance bonuses	8.40 (0.14)	–44.34 (–0.62)
Requirement to use performance bonuses to serve highly disadvantaged	–750.69 (–1.88)	–1196.69 (–3.41)*
Model predicting power—percentage of variation explained by model	14 percent individual level 95 percent between sites	6 percent individual level 91 percent between sites

[a] Coefficients measure differential effects of covariates on first post-program year earnings for treatment group.

[b] Coefficient on treatment group indicator shows the program earnings impact for treatment group members.

[c] Fixed effects in the outcome model and random coefficients (treatment group indicator) in the impact model.

Coefficient value followed by t-ratio in parentheses; n.a. = not applicable

*Statistically significant at α=0.05.

type of data used—administrative (outcome) versus experimental (impact)—relying on administrative data to generate information about how to improve program performance would not likely misdirect managers away from the goal of increasing program impacts. Statistical tests show that no statistically significant variation in outcomes or impacts *between sites and years* remained to be explained in these models. This suggests that policy makers might use the model findings without serious concern that some other mediating, site-level variable (unaccounted for in this model) might redirect the influence of policy actions developed in light of the results.

It is also important to reiterate, however, that over 95 percent of the *total* variation in earnings outcomes and impacts is within sites (at the *individual* level), and the proportion of variation in JTPA participant earnings outcomes and impacts explained at this level is comparatively low (14 percent and 6 percent, respectively). Because many factors affect individuals' labor market success following participation in job-training programs (particularly as long as a year after program termination)—employer–employee relationships, the acquisition of additional education and job-related skills, and other environmental influences—it is not surprising that more variation occurs at the individual level and that less of it is explained by observed characteristics. Studies of welfare-to-work program impacts on earnings have found similar proportions of variation (unexplained and explained) at the individual level (Bloom, Hill, and Riccio 2001; Jennings and Ewalt 2000). Hill (2001, 4), nonetheless, argues convincingly for the importance of the policy findings of these models, noting that organizations cannot change client backgrounds, but they do have control over "fundamental levers for influencing client outcomes," such as the availability of services, administrative structures that shape service-delivery processes, and other performance-management policies that influence program outcomes.

Discussion and Conclusion

The substantive empirical results of this study illustrate the importance of weighing the influence of organizational structure and complexity, policy choices and constraints, and service-delivery practices in assessing program performance. One of the stronger multilevel-model results suggests that greater centralization of formal authority and control by private industry councils in their role as administrative entities not only encouraged a greater emphasis on measured performance, but also improved participants'

earnings outcomes and impacts. Considered in the context of incentive-compatibility theory, this result should not be surprising, given the private industry councils' tendencies toward a more market-oriented, less politicized interpretation of program goals, which is consistent with a system in which performance incentives are tied to labor market outcomes (Holmstrom and Milgrom 1991; Miller 1992).

The multilevel models also facilitated more precise appraisals of structure and management effects, distinguishing, for example, between the effects of a more centralized administrative structure and those of more centralized (or direct) service provision. The model findings suggest it may be more important that the administrative entity that receives the training grant also has primary control over the *selection* of service providers (contract administration) than over service provision directly.

More generally, public executives and mid-level managers engaged in performance-management activities should view the empirical findings of this study as encouraging to their efforts to obtain and use information about the performance of their organizations. The coefficients on the structure and management variables in the models of program performance were similar in relative size and sign in both the earnings outcome and impact models. In their research on relationships between program administration and service-delivery practices and program outcomes and impacts in welfare-to-work programs, Bloom, Hill, and Riccio (2001) and Hill (2001, 6) also find that "it is not crucial to measure performance using experimental impacts" to understand "how front-line, managerial, organizational, or institutional factors are associated with site performance."

The importance of these results is elevated in the context of the General Accounting Office's 1999 finding that a continuing lack of confidence in the credibility of performance information was a major concern for agencies implementing the requirements of the Government Performance and Results Act. In other words, the good news is that imperfect data can still generate information that might effectively guide program managers in improving agency performance. To realize these benefits, however, federal managers and local program administrators may have to change how they use the performance-management information once it is collected.

Radin's (2000, 123) criticism of the Government Performance and Results Act's focus on performance outcomes "to the exclusion of processes and outputs" is valid in this regard. Acknowledging that working with limited data is inevitable, and that performance measures will be *indicators,* at best, and not highly accurate gauges of actual performance, it seems short-sighted to focus annual program performance reports primarily on performance comparisons (empirical performance indicators compared with performance goals defined relative to abstract performance measures or standards). Kravchuk and Schack (1996, 357) convey an analogous concern about the emergence of a "cybernetic-decision" mode, in which managers become insensitive to information unless it comes through the highly structured channels of the performance-management system. They criticize managers for "using measures as substitutes for expert knowledge about, or direct management of, programs." In this mode, managers might use the system to "buffer themselves from the overwhelming complexity of the internal and external environments," particularly when the level at which performance is evaluated is remote from the level of primary work.

In this study, I have argued that by simultaneously monitoring program processes and documenting management and program changes in diverse settings and across multiple levels of government, federal managers might use the performance data more effectively to evaluate the effects of different policies and approaches to managing and delivering government services. Empirically, this might be accomplished in the same way that federal job-training programs already use statistical methodologies (such as multiple regression) to control for local population characteristics and economic conditions in evaluating performance by adding controls in their models for state and local program structures, policies, and management practices.

The review of the JTPA program experience in this study also highlights, however, some of the challenges of identifying internal, external, and cross-level influences on organizational performance and linking them to responsible levels of public management (or to external factors outside of managers' control). For example, how should the system identify and account for the influence of diverse administrative priorities and goals across government levels or the performance "gaming" activities that sometimes arise in response to divergent incentives? Should the system allow for multiple goals (equity and efficiency), or should it focus managers' attention on a single impact indicator? As Kravchuk and Schack (1996) submit, in performance-management systems that aim to account for internal and external complexities, public managers will have to confront inherent tensions between simple, verifiable goals and more complex performance measures, and between the capacity and adaptability of the measurement systems. Radin (2000, 117) argues that the Government Performance and Results Act already has become "another very specific layer of internal management controls," rather than a system based on professional accountability as intended.

If Radin is not overstating these problems, why not simply rely on more traditional systems of accountability: *hierarchical accountability for inputs* (administrative rules guiding routine tasks and budgetary allocations) and *legal*

accountability for processes (audits, site visits, and other monitoring tasks)? A simple answer is that, as a public, we have not been satisfied with the status quo, internally focused approach to government accountability. As Lynn (1987, 6) argues, the study and practice of public management has been hampered by a lack of "widely recognized standards for judging the performance of public executives in policymaking positions. Often 'success' means little more than approval of the elite reference group from which the appointment was made, the cultivation of a good image or reputation with the press or key legislators, or maintenance of good standing with the chief executive's staff because of 'team play.'" The same insights apply to mid-level managers, who, in practice, are typically accountable only to those in their own organization or another government organization (such as the General Accounting Office) that performs the monitoring and oversight tasks.

Although the management and measurement challenges to effective implementation of outcomes-based performance-management systems in government programs are considerable, they are not uncommon to administrative reforms that attempt to change not only the behavior of individuals, but also that of organizational relationships, networks of service providers, and the organizational incentives that sustain them (Weiss 1981). In other words, these early challenges and setbacks confronted in implementing outcomes-based performance management in government programs should not discourage efforts to improve government performance and to make performance evaluation a more public process. As a General Accounting Office report (1999, 8) notes, "Congress understood that effectively implementing management changes of the magnitude envisioned under the Act would take several years." The performance standards system in federal job-training programs has been operating for almost two decades now, and it continues to evolve with successive legislative acts and amendments. Public managers should use the knowledge we have gained through the JTPA performance standards experience, as well as information generated by research and internal reviews, to help the performance-management systems instituted under the Government Performance and Results Act develop into more effective policy tools for guiding program management and organizational functioning, *with less emphasis on the objective of precisely measuring government performance.*

Acknowledgments

I thank the Pew Charitable Trusts for funding this research, Annie Zhang for computer programming assistance, and Laurence E. Lynn, Jr., Carolyn J. Hill, Patricia Ingraham, and anonymous reviewers for helpful comments.

Notes

1. The National JTPA Study was a three-year, randomized experimental evaluation involving 16 local service-delivery areas in 16 states (conducted over the period 1987–89). The Manpower Demonstration Research Corporation was one of the co-investigators in the National JTPA Study and has led numerous other experimental evaluations in the last two decades.
2. In July 2000, the JTPA legislation was repealed and superseded by the Workforce Investment Act of 1998, although some aspects of the original JTPA program structure and operations remain the same.
3. An important change under Workforce Investment Act is a new "universal access" approach to service delivery, in which all adults (not solely economically disadvantaged adults), are eligible for "core" workforce development services such as job search and placement assistance, career counseling, and labor market information.
4. I use the SAS proc mixed procedure to estimate the multilevel models. Using hierarchical linear modeling software produces the same results.
5. The unconditional models of earnings in the first post-program year do not include any covariates in the level 1 or level 2 submodels. The random intercept estimate indicates whether there is statistically significant variation in the dependent variable between level 2 units.
6. The level 1 explanatory variables are centered on their grand mean values, so that the intercept is interpreted as the adjusted average outcome. See Bryk and Raudenbush (1992) for additional discussion of centering options in multilevel models.
7. The complete set of model results are available from the author upon request.
8. The multilevel models presented in this article include both male and female cases, which is similar to the way performance-management data is typically evaluated by program managers. The same models were estimated separately for males and females. The direction and relative magnitude of the site-level variable coefficients did not differ from the results presented here for either males or females. These results are also available from the author.

References

Baker, George. 1992. Incentive Contracts and Performance Measurement. *Journal of Political Economy* 100(31): 598–614.

———. Forthcoming. Distortion and Risk in Optimal Incentive Contracts. *Journal of Human Resources*.

Barnow, Burt S. 2000. Exploring the Relationship between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act. *Journal of Policy Analysis and Management* 19(1): 118–41.

Bloom, Howard S., Carolyn J. Hill, and James Riccio. 2001. Modeling the Performance of Welfare-to-Work Programs: The Effects of Program Management and Services, Economic Environment, and Client Characteristics. Manpower Demonstration Research Corporation, Working Papers on Research Methodology.

Bryk, Anthony S., and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. London: Sage Publications.

Courty, Pascal, and Gerald R. Marschke. 1997. Empirical Investigation of Gaming Responses to Performance Incentives. University of Chicago, Working Paper.

———. Forthcoming. Do Incentives Motivate Organizations? An Empirical Test. In *Performance Standards in a Government Bureaucracy*, edited by James J. Heckman. Kalamazoo, MI: W.E. Upjohn Institute and University of Chicago Press.

DeNisi, Angelo S. 2000. Performance Appraisal and Performance Management. In *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions and New Directions*, edited by Katherine J. Klein and Steve W. J. Kozlowski, 121–56. San Francisco, CA: Jossey-Bass.

Dickinson, Katherine P., Richard W. West, Deborah J. Kogan, David A. Drury, Marlene S. Franks, Laura Schlichtmann, and Mary Vencill. 1988. *Evaluation of the Effects of JTPA Performance Standards on Clients, Services, and Costs*. National Commission for Employment Policy, Research Report no. 88-17.

Gay, Robert S., and Michael E. Borus. 1980. Validating Performance Indicators for Employment and Training Programs. *Journal of Human Resources* 15(1): 29–48.

Hart, Oliver. 1988. Incomplete Contracts and the Theory of the Firm. *Journal of Law, Economics, and Organization* 4(1): 119–39.

Heckman, James J., and Jeffrey A. Smith. 1995. The Performance of Performance Standards: The Effects of JTPA Performance Standards on Efficiency, Equity and Participant Outcomes. University of Chicago, Department of Economics, Working Paper.

Heckman, James J., Carolyn J. Heinrich, and Jeffrey A. Smith. 1997. Assessing the Performance of Performance Standards in Public Bureaucracies. *American Economic Review* 87(2): 389–96.

———. 1999. The Performance of Performance Standards. *Journal of Human Resources* 38(4): 778–811.

Heckman, James J., Jeffrey A. Smith, and Christopher Taber. 1996. What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program. In *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth, Vol 7: Reinventing Government and the Problem of Bureaucracy*, edited by Gary D. Libecap, 191–217. Greenwich, CT: JAI Press.

Heinrich, Carolyn J. 1999. Do Government Bureaucrats Make Effective Use of Performance Management Information? *Journal of Public Administration Research and Theory* 9(3): 363–93.

Heinrich, Carolyn J., and Laurence E. Lynn, Jr. 2000. Governance and Performance: The Influence of Program Structure and Management on Job Training Partnership Act (JTPA) Program Outcomes. In *Governance and Performance: New Perspectives*, edited by Carolyn J. Heinrich and Laurence E. Lynn, Jr., 75–123. Washington, DC: Georgetown University Press.

Hill, Carolyn J. 2001. Impacts, Outcomes, and Management in Welfare-to-Work Programs. Ph.D. diss., University of Chicago.

Hollings, Robert L. 1996. *Reinventing Government: An Analysis and Annotated Bibliography*. Commack, NY: Nova Science Publishers.

Holmstrom, Bengt, and Paul Milgrom. 1991. Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization* 7: 24–52.

Jennings, Edward T., and Jo Ann G. Ewalt. 2000. Driving Caseloads Down: Welfare Policy Choices and Administrative Action in the States. In *Governance and Performance: New Perspectives*, edited by Laurence E. Lynn, Jr. and Carolyn J. Heinrich, 109–39. Washington, DC: Georgetown University Press.

Kravchuk, Robert S., and Ronald W. Schack. 1996. Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993. *Public Administration Review* 56(4): 348–58.

Lynn, Laurence E., Jr. 1987. *Managing Public Policy*. Boston, MA: Little, Brown and Company.

McAfee, R. Preston, and John McMillan. 1988. *Incentives in Government Contracting*. Toronto: University of Toronto Press.

Miller, Gary. 1992. *Managerial Dilemmas: The Political Economy of Hierarchy*. New York: Cambridge University Press.

Mintzberg, Henry. 1996. Managing Government, Governing Management. *Harvard Business Review* 74(3): 75–83.

Murphy, Kevin R., and Jeanette N. Cleveland. 1995. *Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives*. Thousand Oaks, CA: Sage Publications.

Orr, Larry, Burt S. Barnow, Robert I. Lerman, and Erik Beecroft. 1997. *Follow-Up Analyses of the National JTPA Study Sample*. Report prepared for the U.S. Department of Labor. Washington, DC: Government Printing Office.

Orr, Larry, Howard Bloom, Stephen Bell, Winston Lin, George Cave, and Fred Doolittle. 1994. *The National JTPA Study: Impacts, Benefits, and Costs of Title IIA*. Bethesda, MD: Abt Associates.

Peters, B. Guy. 2000. Policy Instruments and Public Management: Bridging the Gaps. *Journal of Public Administration Research and Theory* 10(1): 35–48.

Radin, Beryl A. 2000. The Government Performance and Results Act and the Tradition of Federal Management Reform: Square Pegs in Round Holes? *Journal of Public Administration Research and Theory* 10(1): 111–35.

Romzek, Barbara S. 1998. Where the Buck Stops: Accountability in Reformed Public Organizations. In *Transforming Government: Lessons from the Reinvention Labs*, edited by Patricia W. Ingraham, James R. Thompson, and Ronald P. Sanders, 193–219. San Francisco, CA: Jossey-Bass.

U.S. General Accounting Office (GAO). 1997a. Managing for Results: Prospects for Effective Implementation of the Government Performance and Results Act. Washington, DC: Government Printing Office. GAO/AIMD-97-113.

———. 1997b. Managing for Results: Analytical Challenges in Measuring Performance. Washington, DC: Government Printing Office. GAO/AIMD-97-138.

———. 1999. Managing for Results: Opportunities for Continued Improvements in Agencies' Performance Plans. Washington, DC: Government Printing Office. GAO/AIMD-99-215.

Weiss, Janet A. 1981. Substance vs. Symbol in Administrative Reform: The Case of Human Services Coordination. *Policy Analysis* 7(1): 21–46.

Appendix A Study Data Sources and Descriptive Statistics

1. National JTPA Study (NJS)

The U.S. Department of Labor commissioned the National JTPA Study in 1986, fulfilling a congressional mandate that required a study of the effectiveness of programs operating under the Job Training Partnership Act of 1982. Sixteen local service-delivery areas agreed to participate in this study, which included both experimental and nonexperimental evaluation components. Data were collected beginning in November 1987 and continuing through September 1989 through random assignment of JTPA applicants to "treatment" and "control" groups. Treatment group members were allowed access to JTPA program services, while control group members were prohibited from enrolling in the program for a period of 18 months after random assignment. Approximately 20,000 treatment and control group members who applied to Title IIA adult or youth programs were involved in this study, with 9,621 JTPA participants in the treatment group. The Manpower Demonstration Research Corporation, Abt Associates, Inc., and their subcontractors, including the National Opinion Research Center, conducted the evaluation.

2. Training service plans of the 16 NJS sites for program years 1987–89

All JTPA administrative entities are required to prepare annual training service plans that document (1) the composition of the private industry council; (2) the structure of the administrative entity and its service-delivery arrangements; (3) the types of program services that will be offered and anticipated wage and employment outcomes of the training recipients; (4) the types of supportive services that will be made available; (5) coordination agreements with other government, nonprofit, and for-profit organizations; (6) performance-monitoring plans; and (7) basic information on budget allocations. The administrative structure of the service-delivery area is documented through formal, signed agreements that specify the relationships between the private industry council and the CEO or local elected official, the designated administrative entity, and the grant recipient. The information contained in these agreements was used to create variables that describe the organizational structure of the 16 NJS service-delivery areas.

3. Performance-incentive policy documentation from the 16 NJS sites

Detailed information was collected directly from the 16 NJS sites on the performance-incentive policies adopted in the states in which they operated, as well as the performance-incentive policies they implemented at the local level. State-level performance-incentive policy information was verified by comparing it with data in the JTPA Annual Status Report data system. The local-level performance-incentive policy information was also confirmed through copies of policy documentation and interviews with program administrators at the sites.

4. Local area unemployment statistics data

The U.S. Department of Labor and Bureau of Labor Statistics data on annual local area unemployment rates were matched to geographical areas of the National JTPA Study sites.

Description of variables included in empirical models

Variables[a]	Frequency or mean	Standard deviation
Demographic Characteristics		
Gender (male)	46.2%	n.a.
Age 22-29 years	31.6%	n.a.
Age 30-39 years	26.2%	n.a.
Over age 40 years	15.2%	n.a.
Black	31.1%	n.a.
Hispanic	11.7%	n.a.
Other race/ethnicity	3.0%	
Married	20.4%	
Divorced, widowed, or separated	25.5%	n.a.
No high school degree	34.0%	n.a.
Post high school education	14.1%	n.a.
Employment -> unemployment transition in pre-program year	65.5%	n.a.
Welfare recipient	46.5%	n.a.
Parent of child(ren) under age 6	21.5%	n.a.
Earnings History		
Gross earnings in pre-program year	$2,883.55	$3,981.67
Earnings Outcome		
Earnings in first post-program year	$6,067.49	$7,280.03
Regional and Economic Indicators		
South	26.7%	n.a.
West	20.5%	n.a.
Midwest	31.5%	n.a.
Unemployment rate	6.2%	n.a.
Service Delivery Area Structure/Management Policies		
Private industry council is the administrative entity	56.0%	n.a.
Private industry council and CEO/LEO share administrative authority equally	45.8%	n.a.
Percent of services provided directly by the administrative entity	28.1%	n.a.
Percent of performance-based contracts	24.1%	n.a.
Minimum number of standards an SDA must meet to qualify for performance bonuses	4.7	2.7
Administrative entity required to use performance bonuses in service to "hard-to-serve" groups	19.3%	n.a.
Weight accorded entered employment rate standard	11.9	4.8
Weight accorded to cost per entered employment rate standard	9.1	10.5

[a]Descriptive statistics for treatment group members.

Appendix B Further Discussion of the Multilevel Model Specifications

Multilevel models are estimated to avoid two common problems in empirical analyses with hierarchical data: (1) the attempt to draw individual inferences from aggregate data; and (2) inferences about organizational relationships based on individual-level data. In the multilevel models, each of these levels of analysis in the hierarchical data structure is formally represented by its own submodel. Equations 1 and 2 below represent the submodels of the earnings outcome model (described in the text of this article). In simultaneously estimating this model using maximum likelihood methods, equation 2 is substituted into equation 1.

$$(1)\quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{1j} + \ldots + \beta_{nj}X_{nj} + r_{ij} \qquad \text{(level 1 submodel)}$$

$$(2)\quad \beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \ldots + \gamma_{0n}W_{nj} + \lambda_{01}\Psi_{1j} + \ldots + \lambda_{0n}\Psi_{nj} + u_j \qquad \text{(level 2 submodel)}$$

The subscript j denotes the site and allows each site to have a unique intercept and slope for each of the level 1 predictors, (X_{1j} to X_{nj}). The residual, r_{ij}, is assumed to be normally distributed with homogeneous variance across sites. In the outcome model estimated in this article, a unique slope is not estimated for each of the level 1 predictors for each of the sites (across years); only a random (unique) intercept is specified. This specification, with fixed coefficient effects, is also known as a random intercept model. Since the level 1 covariates are grand-mean centered (measured as deviations from their mean for the full sample), the coefficients on the site-level variables measure the effects of these variables (administrative structures, performance incentive policies, and contracting practices) on *adjusted average earnings outcomes* for training recipients at the sites. See Bryk and Raudenbush (1992) for further discussion of centering and model specifications.

In the earnings impact model estimated in this article using data on both treatment and control group members, a random coefficient (on the treatment group indicator) is added to the level 2 submodel (shown below in equation 4). The addition of the random coefficient for the treatment group indicator (β_{1j}) changes the interpretation of some of the site-level effects; the estimated site level coefficients (γ_{11} to γ_{1nj}) now measure the effects of site-level variables on *conditional treatment impacts* rather than adjusted average earnings outcomes. The site-level variables controlling for regional and economic conditions (Ψ_{1j} to Ψ_{nj}) are still included in the random intercept model, and thus their interpretation is the same as in the outcome model.

$$Y_{ij} = \beta_{0j} + \beta_{1j}Dij + \beta_{kj}X_{2j} + \ldots + \beta_{nj}X_{nj} + \delta_{kj}X_iDij + \ldots + \delta_{nj}X_{nj}Dij + r_{ij} \qquad \text{(level 1 submodel)}$$

$$(3)\quad \beta_{0j} = \lambda_{00} + \lambda_{01}\Psi_{1j} + \ldots + \lambda_{0n}\Psi_{nj} + u_j \qquad \text{(random intercept)}$$

$$(4)\quad \beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + \ldots + \gamma_{1n}W_{nj} + v_{0j} \qquad \text{(random coefficient)}$$

In addition, as noted in the text, interaction terms between the treatment group indicator and the level 1 covariates are added to the level 1 model to control for differences between treatment and control group members (in light of the problems with random assignment in the National JTPA Study). Therefore, to see the total effect of a covariate on treatment group members' earnings, the β_{kj} (average) coefficient value must be added to the δ_{kj} (differential) coefficient value. Only the δ_{kj} coefficient values are shown in table 2. The full set of model results are available from the author.