
Performance Management of U.S. Job Training Programs: Lessons from the Job Training Partnership Act

Burt S. Barnow
Institute for Policy Studies
Johns Hopkins University

Jeffrey A. Smith
Department of Economics
University of Maryland

Abstract

Performance management systems that provide (relatively) quick feedback to program operators play an important role in employment and training programs around the world. This paper brings together evidence on the effects of such systems, drawn primarily from the widely studied U.S. Job Training Partnership Act (JTPA) program. The questions addressed include the effects of performance standards systems on who gets served, on what services they receive, on the technical efficiency of program operation, on the economic efficiency of program operation, and on strategic behavior by program managers to “game” the system. Though incomplete on important dimensions, the available evidence makes it clear that the JTPA standards did little to enhance the impact of the program on employment and earnings, while at the same time inducing attempts to “game” the system by actions such as terminating participants at a time that would maximize measured performance rather than when services are no longer provided. We conclude by drawing policy conclusions from the available evidence and making suggests for future research.

Introduction

This paper reviews the effects of performance management systems in federally sponsored employment and training programs. We focus on programs for the disadvantaged as these programs have the longest history, but the lessons generalize to other programs that

use performance management incentives to influence behavior and, somewhat less directly, to private sector incentive schemes such as those discussed in, e.g., (Baker, Gibbons and Murphy, 1994).

Although the Government Performance and Results Act (GPRA) of 1993 requires all government agencies to develop performance measures and standards, GPRA does not require the use of monetary rewards for exceeding standards and sanctions for failing to meet standards; thus, the inducements found for training programs, where rewards and sanctions are strong, may not be as strong for other programs. On the other hand, the incentives present in some high stakes testing regimes in public schools instituted under the No Child Left Behind Act may be stronger than those common in employment and training programs.

We find in our survey that most of the evidence on the effects of performance systems relates to their failure to motivate behavior in the direction of increasing the mean impact of program participation, and their success at inducing cream skimming and strategic responses on the part of program operators. At the same time, little or nothing is known about the effects of performance systems on assignment to service types or on the technical efficiency of program operation. We recommend further research to fill in gaps in our knowledge as well as policy changes to reflect the knowledge we already have.

The remainder of the paper proceeds as follows. The second section lays out the theory behind performance management systems in government programs. The third section provides the historical background on the use of performance management in U.S. employment and training programs, and the fourth section discusses the available evidence on incentive effects in employment and training programs. The final section provides conclusions and recommendations.

Theoretical Background

In this section, we explore why an incentive-based system might be useful in employment and training programs, and why existing performance management systems take the form they do. We draw primarily upon research on the Job Training Partnership Act

(JTPA). The JTPA program was the primary federal training program for the disadvantaged from 1982 through 1998, at which time the Workforce Investment Act (WIA) program replaced it. (1)

Why have performance management systems?

Consider the JTPA program (virtually the same issues arise in WIA). This program involved the federal, state, and local levels of government. The federal government funded the program and set its broad outlines. Administration was partly devolved to the state level, and operation was primarily the responsibility of local entities. The key problem with such an arrangement is that the state and local governments, and their contractors, may have goals different from those of the federal government. In the language of economics, such multi-level programs involve a principal-agent problem, in which the federal government (the principal) tries to get its agents (state and local governments and their contractors in JTPA and WIA) to further its program goals. (2) See (Prendergast, 1999) and (Dixit, 2002) for theoretical discussions of principal-agent problems and Marschke (2003) for a clear and non-technical discussion of the principal-agent problem in the specific context of JTPA.

A first step in solving principal-agent problems is for the principal to define its goals. As (Dixit, 2002) points out, ascertaining the goals of federal programs is not always a simple matter, and even when they are clear, there are often multiple, partially conflicting goals representing the aims of different stakeholders. In the case of JTPA, Section 141 of the statute states that opportunities for training are to be provided to “those who can benefit from, and are most in need of, such opportunities.” Furthermore, the statute states in Section 106, which describes the program’s performance management system, that training should be considered an investment and that “it is essential that criteria for measuring the return on this investment be developed and.... the basic measures of performance for adult training programs under Title II are the increases in employment and earnings and the reductions in welfare dependency resulting from the participation in the program.”

The statute clearly indicates both equity (serving the hard-to-serve) and efficiency (maximizing the net gain) goals for the program. As we discuss below, these goals may or may not conflict in practice.

For the moment, take them as given and consider the question of how the federal government gets the state and local players in JTPA to further its goals.

Under JTPA, the federal money for the program was first distributed to the states by formula and then further distributed to local areas known as “service delivery areas” (SDAs). (3) The SDAs then selected from one or more of the following options: (a) delivering services themselves, (b) contracting with for-profit organizations, (c) contracting with nonprofit organizations, typically community colleges or community based organizations (CBOs), and (d) making individual referrals to for-profit or nonprofit organizations.

States, SDAs, and the for-profit and non-profit service providers under contract to the SDAs may each have goals that differ in whole or in part from those of the federal government. States may wish to promote use of their community college systems, or economic development in specific regions. Local governments may reduce the training given to each participant below the optimal amount in order to provide services to a larger number of participants (voters), or they may allocate funds to activities based on popularity with voters rather than based on the present value of the earnings gains. (4) For-profit vendors want to maximize profits, so they will follow the incentives implicit in a performance standards system, whether or not those incentives promote program goals. Non-profit vendors may emphasize service to particular ethnic, religious, or target groups. They may also emphasize service to “hard to serve” clients.

The JTPA performance standards system sought, and the similar system under WIA seeks, to provide incentives for the lower level actors in the system to do the bidding of the federal government, instead of pursuing their own objectives. The system did so by setting out concrete performance measures related (it was hoped) to program goals, and by providing budgetary rewards and sanctions to SDAs based on their measured performance.

Why performance systems take the forms they do

A performance management system requires measures of performance, standards that indicate acceptable performance, and rewards and sanctions (which need not be monetary) for

organizations that exceed or fail to meet the standards. Although not the case originally for job training programs in the United States, performance management systems are typically linked to strategic planning efforts for agencies. (5) Performance-based contracting is a system where the vendor receives some or all of its compensation based on achieving certain performance goals. Both approaches attempt to align the interests of the agents with those of the principal, and performance-based contracting may be thought of as a special case of performance management.

Ideally, and in some cases in practice, the performance incentive system directly measures and rewards the government's goals. In our context, that means measuring and rewarding earnings impacts and service to the hard-to-serve. The latter is relatively straightforward, as it requires only measuring the characteristics of program participants. The former, however, is not straightforward. As is well known, measuring earnings impacts is not a trivial task because of the difficulty of estimating what labor market outcomes participants would have experienced, had they not participated. (6) Social experiments, the preferred way to estimate impacts, are expensive and time consuming, while non-experimental methods are controversial. Moreover, as shown in (Heckman and Smith, 1999), because of "Ashenfelter's (1978) dip," the observed phenomenon that the mean earnings of participants in employment and training programs decline in the period prior to participation, before-after estimates will not be a reliable guide to program impacts. Instead, a comparison group of non-participants must be utilized, an undertaking likely to greatly increase the cost of the system.

Furthermore, the real goal is long-run earnings impacts, but waiting around for the long run makes little sense in administrative terms. For administrative purposes, quick feedback is required, so that agents perceive a clear link between their actions and the rewards and punishments they receive under the incentive system. (7)

The difficulty in measuring program impacts, and the desire for quick response, leaves the federal government with three alternatives as it tries to get states and local agencies to advance its goals for the program. First, it can use fixed-price contracts. This leaves local governments, for-profit vendors, and nonprofit vendors to use the money they receive to pursue their own agendas, subject to

regulatory restrictions, such as program eligibility rules, and to professional norms. Second, the government can use cost-based reimbursement schemes. However, it is well documented in the health literature that such an approach can lead to overuse of resources, which is another way of saying that impacts net of costs will not be maximized and, therefore, the government's goals will not be served. Finally, the federal government can adopt a system of performance incentives based on short-term outcome levels, rather than on long-term program impacts. As we discuss in detail below, such a system provides local training programs and their service providers, regardless of type, with many perverse incentives, so that even a for-profit vendor with no agenda of its own may not end up pursuing the government's goals.

Despite these potential problems, JTPA made the third choice and rewarded short-term outcome levels. In the usual notation, the JTPA system based rewards on short-term values of Y_1 , the labor market outcome levels achieved by participants. In contrast, the program's goal is to maximize long-term values of $\Delta = Y_1 - Y_0$, where Y_0 is the counterfactual labor market outcome participants would experience if they did not participate and, as a result, Δ is the impact of participation. (For WIA, some of the rewards are based on the *change* in the outcome of interest rather than the level. Although this could be an improvement over the JTPA approach, it is still not capturing the impact of the program for the reasons noted above.)

Historical Background

Performance management in employment and training programs began during the Comprehensive Employment and Training Act (CETA) period in the 1970s. It was formally incorporated into the Job Training Partnership Act (JTPA) in the early 1980s. Unlike many of the performance management systems established in the wake of the Government Performance and Results Act of 1993 (GPRA), the JTPA system was designed primarily by economists who wanted to maximize the employment and earnings gains of participants. Many of the other systems devised in response to GPRA focus on management issues and program outputs rather than on program impacts. Another important distinction between the JTPA performance management system and the GPRA requirements

is that GPRA does not require the use of monetary sanctions and rewards. While the lack of strong incentives under GPRA may reduce many of the gaming problems described below, it also may be a major reason why some observers, such as Radin (1998, 2000), have concluded that GPRA has not affected management practices as much as some analysts had predicted.

The JTPA program included a performance management system that provided rankings of the local Service Delivery Areas (SDAs). There were about 620 SDAs in the program, each with a geographic monopoly. Six percent of the budget of JTPA was set aside for performance awards to SDAs that performed well relative to performance standards based on the labor market outcome levels (not impacts) achieved by their participants in a given program year and technical assistance for SDAs that failed to meet their standards. (8)

The JTPA performance standards system evolved considerably over its life from 1982 to 2000, when WIA replaced JTPA. (9) The short but controversial history of the JTPA performance standards system illustrates how successive attempts to develop an effective system led to new reforms, which in turn led to new concerns. (10)(11)

Originally, JTPA had four performance measures for Title II-A: the entered employment rate, the average wage at placement, the cost per entered employment, and the entered employment rate for welfare recipients. (12) Although the statute called for measures to use gains in employment and earnings from before the program, the U.S. Department of Labor (DOL) believed (incorrectly) that virtually all participants were unemployed prior to entry, so that the post-program outcomes would represent before-after changes. Although the statute also called for post-program measures, it was widely believed at that time that post-program standards should only be implemented after data were collected for several years, in order to provide a foundation for setting reasonable standards on such measures.

A desire to hold local programs harmless for variations in local economic conditions and the characteristics of their participants, combined with the fact that the people responsible for developing the system were mostly economists, led to the use of regression models to adjust the level of satisfactory performance for differences in local

conditions and participant characteristics. (13) To implement these models, the system included the collection of data on participant characteristics and local economic conditions.

Governors had a great deal of discretion in the JTPA system. They could use the national standards without making any adjustments, they could use the DOL regression model for their SDAs, they could use the regression model and make further adjustments to take account of unique features in their state, and they could develop their own adjustment procedure. Governors also could decide how to weight the various measures and could (and did) add additional measures. Governors also determined the “award function” that mapped SDA performance into budgetary rewards. These functions varied widely among states at a point in time and over time within states; see (Courty and Marschke, 2004a) for a detailed description.

When 13-week post-program employment and earnings data became available, four additional standards were added in Program Year (PY) 1988. (14) At this point, the employment and training community felt that there were too many performance measures, so as of PY 1990 all the measures at the time of termination were dropped and effectively replaced by the four measures based on the 13-week follow-up data. Another important development for PY 1990 was that the cost standard was dropped. This was done because it was widely believed (especially by providers of expensive long-term classroom training) that the cost standard was leading SDAs to focus too much on “quick fix” job search activities. (15) Although most states used the DOL regression model to adjust standards for local economic conditions and the characteristics of participants, some states did not do so. To encourage them to do so, DOL required states not using the DOL model to use an alternative adjustment procedure that met criteria established by DOL.

When the Workforce Investment Act became operational in July 2000, the performance management system was modified in several significant ways. (16) Standards are now set for states as well as local areas, and the standards are “negotiated” rather than set by a regression model. (17) No automatic adjustments are made to take account of economic conditions or participant characteristics, but states may petition to the Department of Labor if circumstances have changed. The lack of a regression adjustment model is not based on

statutory language. Indeed, while not requiring the use of a regression model, the statute states that the state-level standards are supposed to be set “taking into account factors including differences in economic conditions, the characteristics of participants when the participants entered the program, and the services to be provided.” Like JTPA, WIA called for the use of before-after earnings change performance measures, although under JTPA this requirement was ignored and the earnings performance measures were based on levels of post-program earnings. (18)

There are a total of 17 core performance measures for WIA. For adults, dislocated workers, and youth ages 19 through 21, the core measures are defined as:

- The entered employment rate;
- Retention in employment six months after entry into employment;
- The earnings change from the six months prior to entry to the six months after exit; and
- The obtained credential rate for participants who enter unsubsidized employment or, in the case of older youth, enter postsecondary education, advanced training, or unsubsidized employment.

For youth between the ages of 14 and 18, the core performance measures are:

- Attainment of basic skills and, as appropriate, work readiness or occupational skills;
- Attainment of a high school diploma or its equivalent; and
- Placement and retention in postsecondary education and training, employment, or military service.

Finally, there are customer satisfaction measures for both participants and employers.

The changes to the performance management system from JTPA to WIA were significant, so we discussed the rationale for the changes with individuals involved in the process. (19) The WIA legislation did not require dropping the model-based performance management system used under JTPA, so the switch was based on preferences rather than necessity. Indeed, a workgroup of practitioners established to advise the Employment and Training Administration recommended that a model-based system be retained. See (Heinrich, 2003) for a detailed discussion of problems with the WIA performance management system.

There were several reasons for substituting a negotiated standards system for a model-based system. First, ETA wanted to signal that WIA was going to be different than JTPA, so change was considered good in its own right. Second, the group charged with developing the performance management system felt that under JTPA the system was “looking back,” and they believed that a negotiated standards system was prospective in nature rather than retrospective. Finally, a model-based system, by definition, requires that data for the regression models be collected. States indicated to ETA that they found the JTPA data collection requirements to be onerous, and they urged that the data collection be reduced under WIA. (We believe, however, that ETA may have taken the data collection burden too seriously. As noted by Barnow and King (2003), a number of states have added additional performance measures to the federally mandated measures.) Although reducing the amount of data collection would not require abandonment of a model-based system, it would support such a decision. Barnow and King (2003) report that the performance management system in WIA is widely reported to be the new program’s most unpopular feature with states and local programs.

Evidence on Incentive Effects

In this section, we examine the literature on the effects of performance incentives on the behavior of local training centers – e.g., SDAs in JTPA and Workforce Investment Boards (WIBs) in WIA – that provide employment and training services to the disadvantaged. We divide the possible effects into five categories, and then consider the evidence from the literature on each category in turn. The five categories flow out of (in part) the theoretical model presented in (Heckman, Heinrich, and Smith, 2002).

The first type of response consists of changes in the set of persons served. Performance incentives may induce programs to serve persons who will increase their likelihood of doing well relative to the outcomes measured by the incentives, rather than serving, say, the hard-to-serve. The second type of response consists of changes in the types of services provided conditional on who is served. Here the incentives may lead to changes that will maximize the short-term outcomes, such as employment at termination from the program (or shortly thereafter), emphasized in incentive systems, rather than long-term earnings gains.

The third type of response consists of changes in the (technical) efficiency of service provision conditional on who is served and what services they receive. We have in mind here both the effect of incentives on on-the-job leisure as well as their effects on the effort devoted to the design of office procedures and the like. The fourth type of response has to do with subcontracting. Training programs may change the set of providers they contract with, and may pass along (perhaps in modified form) their performance incentives to their providers. The latter will, in turn, affect the actions of those providers. Finally, the fifth type of response consists of gaming, whereby training centers take actions to affect their measured performance that do not affect their actual performance, other than indirect effects due to the diversion of time and resources.

In the remainder of this section, we summarize what is known from the employment and training literature about each of these responses. Much of the available evidence comes from a major experimental evaluation of the JTPA program funded by DOL and conducted in the late 1980s. This evaluation, called the National JTPA Study (NJS), took place at a non-random sample of 16 JTPA SDAs around the United States. The evaluation included both adult programs and out-of-school youth programs. See (Orr et al., 1996) or (Bloom et al., 1997) for a full description of the study as well as experimental impact estimates.

Effects of Performance Incentives on Who Gets Served

The majority of the employment and training literature on performance incentives addresses the question of their effect on who gets served. Under JTPA, SDAs had strong incentives to serve

persons likely to have good labor market outcomes, regardless of whether those outcomes were due to JTPA. Similar incentives, with a minor exception in the form of the before-after performance measure, guide the WIA program. In fact, the absence of a regression model to adjust standards for serving individuals with labor market barriers should make these incentives stronger under WIA than they were under JTPA.

The literature divides this issue into two parts. First, do SDAs (WIBs under WIA) respond to these incentives by differentially serving persons likely to have good outcomes, whether or not those good outcomes result from the effects of the program? This is the literature on “cream skimming.” Second, if there is cream skimming, what are its efficiency effects? Taking the best among the eligible could be efficient if the types of services offered by these programs have their largest net impacts for this group. In what follows, we review the literature on each of these two questions in turn.

Do employment and training programs cream skim?

A few papers, all about the JTPA program, examine whether or not program staff cream skim in response to the incentives provided to do so by the JTPA performance system. The key issue in this literature is the counterfactual: to what group of non-participants should the participants be compared in order to determine whether or not cream skimming has occurred? In all cases, because the performance outcome – some variant of Y_1 in our notation – cannot be observed for the non-participants, the studies proceed by comparing observable characteristics correlated with Y_1 , such as education levels or participation in transfer programs such as Aid to Families with Dependent Children (AFDC) or Temporary Aid to Needy Families (TANF). A finding that participants have “better” characteristics in the form of higher mean years of schooling or lower average pre-program transfer receipt, is interpreted as evidence of cream skimming.

Anderson et al. (1992, 1993) compare the characteristics of JTPA enrollees in Tennessee in 1987 with the characteristics of a sample of JTPA eligibles in the same state constructed from the Current Population Survey. The literature suggests that less than five percent of the eligible population participated in JTPA in each year (see the discussion in Heckman and Smith, 1999), which allows wide scope for cream skimming. Both papers find modest evidence of

cream skimming. In particular, (Anderson et al., 1993)'s bivariate probit analysis of program participation and post-program job placement suggests that if eligible persons participated at random, the placement rate would have been 61.6 percent rather than 70.7 percent, a fall of 9.1 percentage points.

The problem with the (Anderson et al., 1992, 1993) papers is that they potentially conflate participant self-selection with cream skimming by program officials. As documented in (Devine and Heckman, 1996), the JTPA program eligibility rules cast a wide net. The eligible population included both many stably employed working poor persons and persons who were out of the labor force. Both groups had little reason to participate in JTPA.

Heckman and Smith (2004) address the issue of self-selection versus selection by program staff using data from the Survey of Income and Program Participation (SIPP) on JTPA eligibles combined with data from the National JTPA Study. They break the participation process for JTPA into a series of stages – eligibility, awareness, application and acceptance, and participation – and look at the observed determinants of going from each stage to the next. They find that some differences between program eligibles and participants result primarily from self-selection at stages of the participation process, such as awareness, over which program staff have little or no control. For example, for persons with fewer than 10 years of schooling, lack of awareness plays a critical role in deterring participation, although this group is differentially less likely to make all four transitions in the participation process than are persons with more years of schooling. The evidence in (Heckman and Smith, 2004) suggests that while cream skimming may be empirically relevant, comparing the eligible population as a whole to participants likely overstates its extent, and misses a lot of substantive and policy-relevant detail.

The paper by (Heckman, Smith, and Taber, 1996) presents a contrasting view. They use data from the Corpus Christi, Texas SDA, the only SDA in the National JTPA Study for which reliable data on all program applicants are available for the period during the experiment. In their empirical work, they examine whether those applicants who reach random assignment (i.e., were selected to participate in the program) differ from those who do not in terms of both predicted outcome levels (earnings in the 18 months after

random assignment) and predicted program impacts (projected into the future and discounted). Heckman, Smith, and Taber (1996) argue that it is this stage over which program staff have the greatest control, although even here applicants may wander off if they find employment elsewhere, get in trouble with the law, and so on. The authors find strong evidence of negative selection on levels combined with weak evidence for positive selection on impacts. They attribute the former to a strong “social worker mentality” toward helping the hard-to-serve among the eligible that was evident in interactions with program staff at the Corpus Christi site.

The Workforce Investment Act (WIA) program offers an interesting contrast to JTPA because the WIA performance standards are not adjusted by a regression model, and they therefore do not hold programs harmless for the characteristics of their participants. Because programs now have stronger incentives to enroll individuals with few barriers to employment, we would expect to observe enrollment shift toward this group. A recent internal (U.S. Department of Labor, 2002) study finds that this is precisely what appears to be occurring, at least in the area scrutinized:

A brief survey of States by our Chicago Regional Office indicated that WIA registrations were occurring at only half the level of enrollment achieved by JTPA. While some of this may be due to start up issues, there are indications that the reduced registration levels are due to a reluctance in local areas to officially register people in WIA because of concerns about their ability to meet performance goals, especially the “earnings gain” measure. It appears that local areas in these States are selective in whom they will be accountable for. Some local areas are basing their decisions to register a person on the likelihood of success, rather than on an individual’s need for services.

A recent study by the (U.S. General Accounting Office (GAO), 2002) confirms these problems. The GAO report, based on a survey of all 50 states, indicated that “many states reported that the need to meet performance levels may be the driving factor in deciding who receives WIA-funded services at the local level” (p. 14).

Overall, the literature provides modest evidence that program staff responded to the incentives provided by the JTPA performance

standards system to choose participants likely to improve their measured performance whether or not they benefited from program services, and studies of the implementation of WIA indicate that, if anything, the situation has been exacerbated by the new program. At the same time, the evidence from the Corpus Christi SDA indicates that staff concerns about serving the hard-to-serve could trump the performance incentives in some contexts. (20)

What are the efficiency implications of cream skimming?

A number of studies have examined the efficiency implications of cream skimming by estimating the correlation between performance measures and program impacts. In terms of our notation, they estimate the relationship between Y_1 , the outcome conditional on participation, and Δ , the impact of participation. If this relationship is positive, so that higher outcome levels predict higher impacts, then cream skimming is efficient, since it implies serving those with the higher impacts. In contrast, if this relationship is negative, then cream skimming is inefficient, as it means that services are provided to those who benefit less from them than those who would be served in the absence of the incentive system.

Table 1 summarizes the evidence from the seven studies that comprise this literature. (21) The seven papers examine a variety of different programs, ranging from the MDTA program of the 1960s to the Job Corps program of today. Most rely on experimental data for their impact estimates. With one exception, (Zornitsky et al., 1988), the findings are negative or mixed regarding the relationship between outcome-based performance measures of the type typically used in employment and training programs and program impacts. The (Zornitsky et al., 1988) findings refer to a program, the AFDC Homemaker-Home Health Aide Demonstration, which differs from programs such as JTPA and WIA in that it provided a homogeneous treatment to a relatively homogeneous population.

Taken together, the literature summarized in Table 1 clearly indicates that, in the context of employment and training programs, commonly used performance measures do *not* improve program efficiency by inducing service to those who will benefit most. At the same time, the literature indicates that cream skimming likely has neither much of an efficiency benefit nor much of an efficiency cost.

TABLE 1
EVIDENCE ON THE CORRELATION BETWEEN Y_1 AND Δ

Study: Gay and Borus (1980)

Program: Manpower Development and Training Act (MDTA), Job Opportunities in the Business Sector (JOBS), Neighborhood Youth Corps Out-of-School Program (NYC/OS) and the Job Corps.

Data: Randomly selected program participants entering programs from December 1968 to June 1970 and matched (on age, race, city and sometimes neighborhood) comparison sample of eligible non-participants.

Measure of impact: Impact on social security earnings in 1973 (from 18 to 36 months after program exit).

Impact estimator: Non-experimental "kitchen sink" Tobit model.

Performance measures: Employment in quarter after program, before-after (four quarters before to one quarter after) changes in weeks worked, weeks not in the labor force, wage rate, hours worked, income, amount of unemployment insurance received and amount of public assistance received.

Findings: No measure has a consistent, positive and statistically significant relationship to estimated impact across subgroups and programs. The before-after measures, particularly weeks worked and wages, do much better than employment in the quarter after the program.

Study: Zornitsky, et al. (1988)

Program: AFDC Homemaker-Home Health Aid Demonstration

Data: Volunteers in the seven states in which the demonstration projects were conducted. To be eligible, volunteers had to have been on AFDC continuously for at least 90 days.

Measure of impact: Mean monthly earnings in the 32 months after random assignment and mean monthly combined AFDC and food stamp benefits in the 29 months after random assignment.

Impact estimator: Experimental impact estimates.

Performance measures: Employment and wages at termination. Employment and welfare receipt three and six months after termination. Mean weekly earnings and welfare benefits in the three and six month periods after termination. These measures are examined both adjusted and not adjusted for observable factors including trainee demographics and welfare and employment histories and local labor markets.

TABLE 1 (CONTINUED)
 EVIDENCE ON THE CORRELATION BETWEEN Y_1 AND Δ

Findings: All measures have the correct sign on their correlation with earnings impacts, whether adjusted or not. The employment and earnings measures are all statistically significant (or close to it). The welfare measures are correctly correlated with welfare impacts but the employment measures are not unless adjusted. The measures at three and six months do better than those at termination, but there is little gain from going from three to six.

Study: Friedlander (1988)

Program: Mandatory welfare-to-work programs in San Diego, Baltimore, Virginia, Arkansas, and Cook County.

Data: Applicants and recipients of AFDC (varies across programs). Data collected as part of MDRC's experimental evaluations of these programs.

Measure of impact: Post random assignment earnings (from UI earnings records) and welfare receipt (from administrative data).

Impact estimator: Experimental impact estimates.

Performance measures: Employment (non-zero quarterly earnings) in quarters 2 and 3 (short-term) or quarters 4 to 6 (long term) after random assignment. Welfare receipt in quarter 3 (short-term) or quarter 6 (long-term) after random assignment.

Findings: Employment measure is positively correlated with earnings gains but not welfare savings for most programs. Welfare indicator is always positively correlated with earnings impacts, but rarely significantly so. It is not related to welfare savings. Long-term performance measures do little better (and sometimes worse) than short-term measures.

Study: Cragg (1997)

Program: JTPA (1983-87)

Data: NLSY

Measure of impact: Before-after change in participant earnings

Impact estimator: Generalized bivariate Tobit model of pre-program and post-program annual earnings.

Performance measures: Fraction of time spent working since leaving school in the pre-program period. This variable is strongly correlated with post-program employment levels.

TABLE 1 (CONTINUED)
EVIDENCE ON THE CORRELATION BETWEEN Y_1 AND Δ

Findings: Negative relationship between work experience and before-after earnings changes.

Study: Barnow (1999)

Program: JTPA (1987-89)

Data: National JTPA Study

Measure of impact: Earnings and hours worked in month 10 after random assignment.

Impact estimator: Experimental impact estimates.

Performance measures: Regression-adjusted levels of earnings and hours worked in month 10 after random assignment.

Findings: At best a weak relationship between performance measures and program impacts.

Study: Burghardt and Schochet (2001)

Program: Job Corps

Data: Experimental data from the National Job Corps Study

Measure of impact: The outcome measures include receipt of education or training, weeks of education or training, hours per week of education or training, receipt of a high school diploma or GED, receipt of a vocational certificate, earnings and being arrested. All are measured over the 48 months following random assignment.

Impact estimator: Experimental impact estimates.

Performance measures: Job Corps centers divided into three groups: high-performers, medium-performers and low-performers based on their overall performance rankings in Program Years 1994, 1995 and 1996. High and low centers were in the top and bottom third nationally in all three years, respectively.

Findings: No systematic relationship between the performance groups and the experimental impact estimates.

Study: Heckman, Heinrich and Smith (2002)

Program: JTPA (1987-1989)

Data: National JTPA Study

Measure of impact: Post-random assignment earnings and employment

Impact estimator: Experimental impact estimates

TABLE 1 (CONTINUED)
EVIDENCE ON THE CORRELATION BETWEEN Y_1 AND Δ

Performance measures: JTPA performance measures including employment at termination, employment 13 weeks after termination, wage at termination and earnings 13 weeks after termination.

Findings: No relationship between performance measures and experimental impact estimates.

Effects of Performance Incentives on Services Provided

Two papers we know of examine the effect of performance incentives on the types and duration of services offered in an employment and training program, holding constant the characteristics of persons served. Marschke (2002)'s novel analysis uses the variation in performance incentives facing the training centers in the National JTPA Study to identify the effects of performance incentives on the types of services received by JTPA participants. This variation includes both time series variation within states and cross-sectional variation across states during the period of the study. In his analysis, he estimates, for each member of the experimental treatment group, a predicted outcome on each performance measure. These are then entered into a service type choice model along with other factors, such as predicted impacts from each service type and measures of whether or not the participant is "hard to serve." Both the predicted impacts and the hard-to-serve measures are intended to capture any caseworker efforts to act in the interest of the participants (and the long-suffering taxpayer) or to follow their hearts by providing the most expensive services to the worst off.

Marschke (2002) finds evidence that changes in the performance measures employed in JTPA led SDAs to alter the mix of services provided in ways that would improve their performance relative to the altered incentives they faced. In some cases, these changes led to increases in efficiency, but in others they did not. Marschke (2002) interprets his evidence as indicating that SDAs' service choices are responsive at the margin, but that existing performance measures do a poor job of capturing program goals such as maximizing the (net) impacts of the services provided.

More recently, (Courty and Marschke, 2004b) demonstrate that the JTPA performance management system affects the duration of training for some participants because program managers manipulate the duration of services for some participants in order to be able to count them on their performance measures for a specific program year. Courty and Marschke (2004b) find that these manipulations reduce the overall mean impact of the employment and training services provided by JTPA.

Effects of Incentives on the Technical Efficiency of Service Provision

Performance incentives may affect how hard training center employees work and how smart they work, conditional on their choices about whom to serve and how to serve them. Indeed, traditional incentive systems in industry such as piece rates, which are intended to increase the price of on-the-job leisure, aim to produce just such effects.

The only evidence on this type of behavioral response in the literature on employment and training programs that we have located is that for the United Kingdom presented in (Burgess, et al., 2003). They look at office-based incentives for job placements in “Jobcentre Plus” offices. As all participants in their analysis receive the same service, namely a job placement, and because the scheme they examine provides more points for hard-to-serve participants, the effects they find suggest increases in technical efficiency (though this depends on how well the different points reflect differences in the difficulty of placement). Burgess, et al. (2003) find that the incentives work best in smaller offices where, presumably, morale and peer monitoring overcome the free-rider problems associated with office-level (rather than individual-level) performance measures.

Effects of performance systems on technical efficiency are unlikely to get picked up by the sort of regression models employed in the studies summarized in Table 1. To see why, consider the following example. Suppose that establishing performance incentives leads training program workers to work harder, which in turn raises the expected impact of the program for every participant by \$10. In this case, the regressions described above would see their intercepts increase by \$10, but the coefficient on the performance measures

would not increase at all.

In principle, cross-state variation in performance incentive intensity, such as that employed by (Cragg, 1997), in combination with data on outputs (number of persons served, etc.) and number of workers could be used to answer this question in the United States.

In the absence of such evidence, it remains to refer to the evidence from the UK noted above and to the broader economic literature on this question, which is summarized in (Prendergast, 1999). He reports that this “literature points to considerable effects of compensation on performance.” How well his conclusion generalizes to government programs where the rewards consist of additional budgetary allocations, rather than higher earnings for individual workers, remains an open question.

Effects of Performance Incentives on Contracts and Vendor Behavior

In many, if not most, employment and training programs that have operated in the United States, vendors operating under contracts to the local programs that are the grant recipients have played an important role in service delivery. In this sub-section, we consider the evidence on how performance incentives alter the contracts that agencies make with their vendors, and how performance-based contracts affect the performance of providers.

As elsewhere, the literature we survey draws primarily on the experience of the JTPA program. Performance-based contracting had an interesting history under JTPA. (22) Initially, SDAs that entered into performance-based contracts for training were able to exceed the 15 percent limit on administrative expenses in JTPA if the contract met certain provisions. By the late 1980s, a number of concerns surfaced about the use of performance-based contracting. As enumerated in (Spaulding, 2001), DOL was concerned that states were not effectively monitoring their performance-based contracts (PBCs), that total costs billed under PBCs were not “reasonable,” that SDAs were using PBCs for activities that contained little if any training, that contracts did not include the performance measures required by law, that payment schedules either eliminated contractor risk or built in high profit levels, and that profits were sometimes used to establish economic development loan funds, which was

prohibited. The Department of Labor issued a series of guidance letters in the late 1980s intended to reduce the use of PBCs.

In a series of papers, (Heinrich 1995, 1999, 2003) examines the contracting behavior of a JTPA SDA in Cook County, Illinois. (23) She finds that this site passed along its performance incentives to its service providers through performance-based contracts. These contracts often included performance levels in excess of those facing the SDA itself, apparently as a form of insurance. Even if some contractors failed to meet the (inflated) standards in their contracts, most would, and so the SDA would meet its own, overall standards despite a few subcontractor failures. Heinrich (1995, 2004) found that at this SDA, which had technical resources that most other SDAs did not, caseworkers and managers were keenly aware of how they and their subcontracts were doing relative to their performance standards throughout the program year. This was particularly true of the cost-per-placement standard. Heinrich (1999) shows that contractor performance in one program year relative to the cost-per-placement standards in their contract affected whether or not they were awarded a contract in the next year.

Now consider the studies that examine the effects of performance based contracting on vendor behavior. Dickinson et al. (1988) performed some analyses looking at how the use of performance-based contracting affected the mix of participants in JTPA. They found that, contrary to their expectations, the use of performance-based contracting was associated with a statistically significant increase in services to minority groups and had no effects on services to welfare recipients, females, older workers, or individuals with other barriers to employment. Dickinson et al. (1988) also analyzed the impact of higher wage at placement provisions on participants served, and found that they led to a reduction in services to welfare recipients; estimated effects on other hard to serve groups were also negative but not statistically significant.

Heinrich (2000) focuses primarily on the relationship between organizational form (for-profit or nonprofit) and performance, but she also explores the effects of having performance incentives in provider contracts. She finds, for her study of an Illinois SDA, that inclusion of performance incentives has a very strong positive effect on realized wages and employment at termination and up to four quarters

following termination. Similarly, (Spaulding, 2001) analyzed the effect of performance-based contracting in JTPA programs on the performance of SDAs in program year 1998. Her results indicate that the use of performance-based contracting is generally associated with higher outcomes.

Overall, the literature makes two things clear. First, local training programs sometimes pass along the performance incentives they face to their subcontractors, perhaps with something added on as insurance. Second, performance-based contracts yield higher performance on the rewarded dimension.

Strategic Responses to Performance Incentives

In addition to the substantive responses to performance incentives considered above, in which training centers changed what they actually did, local training programs can also attempt to change their measured performance without changing their actual performance. We refer to this as a strategic response, or as “gaming” the performance system. Regardless of their differing goals, all types of organizations have an incentive to respond strategically to performance incentives, provided the cost is low, as doing so yields additional resources to further their own goals. The literature provides clear evidence of such gaming behavior under JTPA.

One important form of strategic behavior under JTPA was the manipulation of whether or not participants were formally enrolled. Under the JTPA incentive system, only persons formally enrolled counted towards site performance. In addition, for the first decade of JTPA’s existence, training centers had substantial flexibility in regard to when someone became formally enrolled. Clever SDAs improved their performance by basing enrollments on job placements rather than the initiation of services. For example, some SDAs boosted performance by providing job search assistance without formally enrolling those receiving it in the program. Then, if an individual found a job, the person would be enrolled, counted as a placement, and terminated, all in quick succession. Similarly, SDAs would send potential trainees to employers to see if the employer would approve them for an on-the-job training slot; enrollment would not take place until a willing employer was found.

There are several pieces of evidence regarding the empirical importance of this phenomenon. The first is indirect, and consists of the fact that DOL found it enough of a problem to change the regulations. Specifically, in 1992 the Department of Labor required that individuals become enrolled once they received objective assessment and that they count as a participant for performance standards purposes once they received any substantive service, including job search assistance. (24)

Other evidence comes from the National JTPA Study. As part of their process analysis of the treatments provided at the 16 SDAs in the study, (Kemple, Doolittle, and Wallace, 1993) conducted interviews of non-enrolled members of the experimental treatment group at 12 of the 16 sites. These results, reported in their Table 3.2, show that 53 percent of non-enrolled treatment group members received services, most often referrals to employers for possible on-the-job training (36 percent of all non-enrollees) and job search assistance (20 percent of all non-enrollees). They report that "... most of the study sites enrolled individuals in classroom training when they attended their first class or in OJT when they worked their first day ..."

There is also evidence that this type of behavior has continued under WIA. The (U.S. General Accounting Office, 2002: p. 14) notes that "All the states we visited told us that local areas are not registering many WIA participants, largely attributing the low number of WIA participants to concerns by local staff about meeting performance levels".

The flexibility of JTPA also allowed strategic manipulation of the termination decision. Because performance standards in JTPA were based on terminees, SDAs had no incentive to terminate individuals from the program who were not successfully placed in a job. By keeping them on the rolls, the person's lack of success would never be recognized and used against the SDA in measuring its performance. As the Department of Labor explains in one of its guidance letters, "Without some policy on termination, performance standards create strong incentives for local programs to avoid terminating failures even when individuals no longer have any contact with the program." (25)

Problems with local programs retaining participants on the

rolls long after they stopped receiving services go back to the days of JTPA's predecessor, the Comprehensive Employment and Training Act (CETA). In one of their guidance letters, the Department of Labor observed that "monitors and auditors found that some participants continued to be carried in an 'active' or 'inactive' status for two or three years after last contact with these programs."¹ For Title II-A of JTPA, DOL limited the period of inactivity to 90 days, although some commentators suggested periods of 180 days or more.

Courty and Marschke (1996, 1997, 2004b) provide additional evidence on the strategic manipulation of termination dates using data from the National JTPA Study. The first type of evidence consists of the timing of termination relative to the end of services as a function of the employment status of the trainee as of the end of services. Assuming that the timing of termination responds mainly to the employment at termination standard in place during the time their data were collected (rather than the wage rate or cost standards, which would be more difficult to game), they argue that sites should immediately terminate participants who are employed when their services end. In contrast, they should not terminate participants who are not employed at the end of their services; instead, they should wait and see if they later become employed, at which point they should then terminate them from the program. Not surprisingly, (Courty and Marschke, 1996, 1997, 2004b) find that the sites in the National JTPA Study did exactly this with, for example, Figure 1 of (Courty and Marschke, 1997) revealing a spike in terminations at the end of services for employed participants, and a spike in terminations at the end of the mandatory 90 days after the end of services for participants not employed at the end of services. Their analysis likely understates the full extent of sites' strategic behavior, as it takes the date of the end of services as given, when in fact sites had some control over this as well. For example, a participant without a job at the end of classroom training could be assigned to a job club in the hope that employment would soon follow.

Courty and Marschke (1997) interviewed 11 of the 16 sites in the NJS regarding their responses to the switch from measuring employment at termination to measuring it 90 days after termination. They report that

"[m]ost administrators indicated that ... case managers began tracking terminees until the follow-up period expired. To

increase the chances that an employment match lasted until the third month, some SDAs reported that they offered special services between termination and follow-up, such as child-care, transportation and clothing allowances. Case managers also attempted to influence employers to keep their clients until the third month.”

Moreover, “training administrators reported that after the third month, they did not contact the client again.” While these follow-up services may add value, their sudden termination at 90 days, and their sudden use after the change in performance standards, suggests motives other than impact maximization.

The second type of evidence from the National JTPA Study reported in (Courty and Marschke, 1996, 1997, 2004) concerns the timing of terminations relative to the end of the program year. In JTPA, performance was measured over the program year from July 1 to June 30. For SDAs in states where there were no marginal rewards for performance above the standard, this leads to an incentive to wait on termination until the end of the program year when possible, and then to strategically terminate each participant in the program year in which his or her marginal value is highest. Consider a site that comes into June well above its performance standard. It should then terminate non-employed participants who have finished their services until its measured performance is just above the standard. It thereby gets its reward in the current year, while starting the next year with as small a stock of poorly performing enrollees as possible.

Courty and Marschke (2004b) builds on the analyses of (Courty and Marschke, 1996, 1997) by embedding them in an econometric framework and by examining whether the manipulation of the termination dates is merely an accounting phenomenon or whether it has efficiency costs. To do this, they look at non-experimental differences in mean impacts between persons terminated at training centers that appear to engage in more gaming (based on measures of the average waiting time to termination after the conclusion of training), at differences in mean impacts for trainees terminated in June (at the end of the program year) relative to other trainees, and at whether or not trainees are more likely to have their training truncated at the end of the program year. The impacts at the end of the training year are also interacted with how close the center is to

its performance standards for the year. All of their analyses indicate an apparent (and surprisingly large) efficiency cost to the gaming behavior.

Conclusions and Recommendations

The literature on the behavioral effects of performance management systems in employment and training programs is a small one. From it, we draw the following conclusions, which are also summarized in Table 2. First, there is modest evidence of cream skimming in JTPA, which had such a system. Because the performance management system in WIA does not adjust the standards for sites that serve more disadvantaged groups, WIA provides even stronger incentives to cream skim than did JTPA. There is no evidence, however, that cream skimming behavior would not have occurred even in the absence of the federal performance standards system, perhaps in response to local political incentives. Second, there is fairly strong evidence in the literature that the performance measures typically used in these systems, which focus on short-term outcome levels of participants, have little or no relationship to long-run impacts on employment or earnings. As such, to the extent that program administrators devote time and effort to including persons in the program who will do well on the performance measures, they are not promoting efficiency. Third, there is not much empirical evidence about the effect of performance standards systems on the types and duration of services provided. The two available papers suggest that SDAs under JTPA allocated services, and set the duration of services, to increase their measured performance; effects on efficiency are mixed on service type and negative for service duration.

Fourth, there is very little direct evidence on the important question of the effects of performance management on the technical efficiency of service delivery. Fifth, performance management at the level of the SDA or WIB leads to changes in the relationship between the SDA or WIB and their vendors in some instances. The nature of the contracts changes, as local programs seek to insure their aggregate (across contractors) performance, and vendors respond by changing their own behavior to please the local program. Sixth, and

finally, there is strong evidence that local programs devote time and resources to gaming performance management systems by increasing their measured performance in ways that do not affect their actual performance. These strategic responses represent a cost of having a performance management system.

In light of these findings, we make two main recommendations regarding performance management in U.S. employment and training programs. The first is that the Department of Labor commission additional research on the performance of performance management systems and that it devote resources to providing the data necessary to support such research. The Department of Labor has spent large sums evaluating its employment and training programs, but much less on evaluating its performance management systems. It is clear to us that marginal benefits have not been equated on these two lines of research.

TABLE 2
SUMMARY OF THE EVIDENCE ON
THE INCENTIVE EFFECTS OF PERFORMANCE STANDARDS
IN EMPLOYMENT AND TRAINING PROGRAMS

Do performance standards lead to cream skimming?

The evidence indicates cream skimming in JTPA and WIA, but less than some critics have suggested. The extent of cream skimming likely varies across sites, with some sites with a strong social work orientation focusing their attention on the hard to serve.

Does cream skimming increase or decrease mean impacts?

The evidence from all but one of several studies on a number of different programs suggests no relationship or even a negative relationship between program impacts and labor market outcomes typically used as performance measures. Thus, if there is cream skimming, at best it has no effect on mean impacts.

Do performance standards affect the mix of services provided?

A single study finds that in the JTPA program the service mix responds to changes in the performance standards. The changes in services provided have a mixed effect on program impacts.

Do performance standards affect the technical efficiency of service provision?

There is no available evidence on this important question in the context of employment and training programs.

Do performance standards lead to strategic behavior by program staff?
Yes, and lots of it in the case of JTPA.

Several types of research would serve to improve the performance of performance management systems. These include, but are not limited to, the following:

- The search should continue for short-term outcome measures that are reliably correlated with long-run program impacts and cannot be gamed by local programs.
- Additional research on the effects of performance management on the types of services offered, on the match between participant characteristics and service type, and on the technical efficiency of service provision would provide a fuller understanding of what the current types of standards actually do.
- Research on the effects of other types of performance measures sometimes adopted at the state level, such as measures designed to encourage service to particular subgroups among the eligible, would inform decisions about whether or not to introduce such measures at the national level.
- Finally, research on the response of WIBs to alternative reward functions at the state level would provide useful information about how to design such functions in the future. Key aspects here include the extent of competition among WIBs, as in tournament systems, variation in the number of standards a WIB must pass to receive any budgetary reward, and the effects of marginal incentives for performance above the standard.

The data required to support the proposed research effort includes a panel data set, with the WIB as the unit of observation, containing for each program year the negotiated standards for each

WIB, the actual performance of the WIB, characteristics of the economic environment and eligible and participant populations for the WIB, and the details of the relevant state policies including any additional standards and related outcomes, and the reward function linking WIB outcomes to budgetary rewards. Had such data been collected under JTPA, the knowledge base for redesigning the WIA system would be much more solid. Even the limited information for the National JTPA Study experimental sites described in (Courty and Marschke, 2004a) yielded useful insights. These data should be collected, maintained, and distributed to the research community, presumably by a private research firm under contract to DOL.

Our second recommendation is that the Department of Labor should redesign the WIA performance management system to reflect the current base of evidence on the performance of these systems. As we show in this paper, the systemic changes from the JTPA performance management system to the WIA system ignored the literature and, overall, took the system farther away from the evidence than it was before. In the absence of a redesign along the lines suggested here, we view the present system as a step backward that should either be scrapped or have its effects reduced by limiting the amount of incentive payments based upon it, pending further research.

We envision four possible scenarios for such a redesign effort, which we list in order of what we see as their desirability. The first redesign scenario represents adoption of an “ideal” performance system. In an ideal system, randomization would be directly incorporated in the normal operations of the WIA program. Such randomization need not exclude persons from any services, but only assign a modest fraction to low-intensity services, i.e., the core services under WIA. It could be incorporated directly into a system similar in spirit to the Frontline Decision Support System (see Eberts et al., 2002 for a description) and so made invisible to line workers. The randomization would then be used, in conjunction with outcome data already collected, to produce experimental impact estimates that would serve as the performance measures. For sample size reasons, randomization might be viable in practice only for state-level performance incentives or only when applied to performance measures consisting of moving averages over several program years.

The second reform scenario takes a different direction. It acknowledges that short-term outcome levels have little or no correlation with program impacts and so changes the system to focus on the program's goals other than efficiency. Such a system could focus, for example, on measures related to who gets served and measures of customer (participants and employers) satisfaction. The customer satisfaction measures would focus on aspects of the program such as waiting times and courtesy of staff, about which the customer is the best judge, and not on value-added, of which the customer is likely to be a poor evaluator (as shown empirically for JTPA in Heckman and Smith, 1998). Somewhat surprisingly, the present system does not do a very good job of guiding behavior along these dimensions, though it easily could. The timeliness standards employed in the Unemployment Insurance system provide an example of a successful system along these lines; see the discussion in (West and Hildebrand, 1997).

The third reform scenario downplays or scraps the current system until additional research identifies measures based on short-term outcomes that correlate with long-term program impacts, or provides convincing evidence that the current system has beneficial effects on dimensions, such as the efficiency of time use by program staff, for which little or no evidence presently exists. In this scenario, the negotiated performance standards could be taken over at the national level and used in a systematic manner to generate knowledge about WIB responses to particular performance measures and to the general toughness of the standards.

The fourth and final reform scenario simply modifies the WIA system to look more like the JTPA system. In practice, this scenario might represent a baseline to which elements of the other scenarios could be added. The heart of this scenario consists of replacing the negotiated standards with a model-based system similar to that in JTPA. Within the context of such a model-based system, a number of suggestions for marginal changes become relevant. First, the model should not be re-estimated every year using a single year's data. Doing so caused a lot of volatility in the standards, and in the effects of particular variables, but did not produce any corresponding benefit. Barnow (1996) discusses how this variability applied to persons with disabilities. Second, given the current focus on return on investment within WIA, a cost standard might be reintroduced, designed in a way to get around problems with WIBs that mix funds

from a variety of sources, but that encourages local programs to focus more on the return on investment. Third, the literature surveyed here has some lessons for the optimal length of the follow-up period for the outcome-based performance measures. In particular, the literature suggest that longer is not always better in terms of correlation with program impacts, above and beyond the problem that longer follow-up periods interfere with the system's ability to provide reasonably quick feedback.

Although a general review of performance management in contexts other than employment and training programs less beyond the scope of this paper, our conclusions have important general implications for the design and implementation of such systems. First, such systems are far from a panacea for reasons made plain by Wilson (1989) many years ago but still often ignored by performance management enthusiasts. Second, expect strategic behavior by the individuals or administrative units whose performance a system seeks to manage. To a large extent, the work of a performance management system designer consists of attempting to anticipate, and counteract, such behavior. In particular, the JTPA experience suggests the importance of clear and careful definitions of all quantities used to measure performance. Third, the JTPA experience suggests that system designers will not anticipate all possible forms of strategic behavior. As a result, performance management systems need to include mechanisms for feedback (from inside and outside the system) for and incremental reform in response to that feedback. Fourth and finally, designers and implementers of performance management systems should read the literature, so that they avoid the fate of the WIA designers who took a problematic JTPA system and made it worse rather than better. While there is much more to learn about the design and implementation of performance management systems in public and private organizations, this fact does not excuse a failure to exploit the knowledge we do have.

Notes

(1). See (D'Amico et al., 2002) and (D'Amico, 2002) for more information on the implementation of WIA and its relationship to JTPA.

(2). A principal-agent problem is not a necessary condition for a performance management system. An individual or firm might adopt such a system as a way of quantifying and rewarding progress toward distant or difficult to measure goals.

(3). In some instances, the state government assumed responsibility for some or all of the service delivery in the state.

(4). See (Barnow, 1979) for illustrative models of this type of behavior.

(5). See (Radin, 2000) for a discussion of the requirements of the Government Performance and Results Act (GPRA), which requires federal agencies to develop performance measures and standards in conjunction with strategic plans and annual reports.

(6). See, for example, the discussion in (Heckman, LaLonde and Smith, 1999).

(7). See (Blalock and Barnow, 2001).

(8). See (Blalock and Barnow, 2001) and (Barnow, 2000) for information regarding the origin of the JTPA performance management system.

(9). WIA became law in 1998 but was not implemented until 2000.

(10). We have more to say about this back-and-forth pattern in a later section.

(11). A referee raised the question of whether the changes in the JTPA performance management system were driven by ideology or politics. Perhaps surprisingly, the answer is no. Because the system involved major risks to the local program and significant responsibilities for the states, an advisory committee was established with representation from major direct stakeholders (but excluding other groups of interest, such as taxpayers), and the system was established on a consensus basis. This was not the case when major changes were made to the performance management system for WIA. In that instance, Department of Labor officials who believed that making changes in the system was important imposed the changes.

(12). For a description of the performance management system in the early days of JTPA, see (Barnow and Constantine, 1988).

(13). It might make more sense to hold constant the characteristics of the eligible population, which the sites do not choose, rather than the characteristics of the participants, which they do.

(14). JTPA operated on a program year rather than a fiscal year basis. For example, program year 1990 began July 1, 1990 and ran through June 30, 1991.

(15). See the discussion in (Dickinson et al., 1988).

(16). See Section 136 of the WIA statute for a description of the law's performance management requirements.

(17). States we have spoken with indicated that the negotiations are largely one-sided, with the Department of Labor Effectively establishing the standards.

(18). See (U.S. Department of Labor, 2000a, 2000b) for more details on the WIA performance standards system.

(19). Our discussion of the motivation for the WIA changes draws on discussions with several staff members who participated in the development of the WIA performance management system. As some of our informants requested anonymity (!) we simply thank all of them for sharing their views without mentioning their names.

(20). This is consistent with the findings from the earlier study by Dickinson et al., 1988).

(21). This table is a modified version of Table 3 in (Heckman, Heinrich and Smith, 2002). See that paper for a more detailed survey of these results.

(22). This section is based on (Spaulding, 2001).

(23). This section draws on (Heckman, Heinrich and Smith, 2002).

(24). See Training and Employment Notice (TEIN) 31-92 for the formal description of the requirements regarding when a person

must be enrolled in the program. This document is available at http://wdr.doleta.gov/directives/corr_doc.asp?DOCN=299.

(25). See TEIN 5-93 available at http://wdr.doleta.gov/directives/corr_doc.asp?DOCN=770.

(26). See TEIN 3-92 available at http://wdr.doleta.gov/directives/corr_doc.asp?DOCN=282.

References

Anderson, Kathryn, Richard Burkhauser, Jennie Raymond, and Clifford Russell. 1992. "Mixed Signals in the Job Training Partnership Act." *Growth and Change*. 22(3): pp. 32-48.

Anderson, Kathryn, Richard Burkhauser, and Jennie Raymond. 1993. "The Effect of Creaming on Placement Rates under the Job Training Partnership Act." *Industrial and Labor Relations Review*. 46(4): pp. 613-624.

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics*. 6(1): pp. 47-57.

Barnow, Burt. 1979. "Theoretical Issues in the Estimation of Production Functions in Manpower Programs." In *Research in Labor Economics, Supplement 1*. Ed. Farrell Bloch. Greenwich, CT: JAI Press, pp. 295-338.

Barnow, Burt. 1996. "Policies for People with Disabilities in U.S. Employment and Training Programs." In *Disabilities, Cash Benefits, and Work*. Eds. Jerry Mashaw, Virginia Reno, Richard Burkhauser, and Monroe Berkowitz. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Barnow, Burt. 1999. "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): pp. 118-141.

Barnow, Burt and Jill Constantine. 1988. "Using Performance Management to Encourage Services to Hard-to-Serve Individuals in JTPA." National Commission for Employment Policy Research Report 88-04. Washington, DC: National Commission for Employment Policy.

Barnow, Burt and Christopher King. 2003. "The Workforce Investment Act in Eight States: Overview of Findings from a Field Network Study. Interim Report." U.S. Department of Labor, Employment and Training Administration Occasional Paper No. 2003-03.

Blalock, Ann and Burt Barnow. 2001. "Is the New Obsession with 'Performance Management' Masking the Truth About Social Programs?" In *Quicker, Better, Cheaper: Managing Performance in American Government*. Ed. Dall Forsythe. Albany, NY: Rockefeller Institute Press, pp. 485-517.

Bloom, Howard, Larry Orr, Stephen Bell, George Cave, Fred Doolittle, Winston Lin and Johannes Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Findings from the National Job Training Partnership Act Study." *Journal of Human Resources*. 32(3): pp. 549-576.

Burgess, Simon, Carol Propper, Mariso Ratto and Emma Tominey. 2003. "Incentives in the Public Sector: Some Preliminary Evidence from a UK Government Agency." CMPO Working Paper Series No. 03/080.

Burghardt, John and Peter Schochet. 2001. *National Job Corps Study: Impacts by Center Characteristics*. Princeton, NJ: Mathematica Policy Research.

Courty, Pascal, and Gerald Marschke. 1996. "Moral Hazard under Incentive Systems: The Case of a Federal Bureaucracy." In *Advances in the Study of Entrepreneurship, Innovation and Economic Growth*, Volume 7. Ed. Gary Libecap. Greenwich, CT: JAI Press, pp. 157-190.

Courty, Pascal, and Gerald Marschke. 1997. "Measuring Government Performance: Lessons from a Federal Job-Training Program." *American Economic Review*. 87(2): pp. 383-388.

Courty, Pascal, and Gerald Marschke. 2004a. "The JTPA Incentive System." In *Performance Standards in a Government Bureaucracy*. Ed. James Heckman. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, forthcoming.

Courty, Pascal, and Gerald Marschke. 2004b. "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." *Journal of Labor Economics*. 22(1): pp. 22-56.

Cragg, Michael. 1997. "Performance Incentives in the Public Sector: Evidence from the Job Training Partnership Act." *Journal of Law, Economics and Organization*. 13(1): pp. 147-168.

D'Amico, Ronald. 2002. "Training Services for Adults and Dislocated Workers Under the Workforce Investment Act." Unpublished manuscript.

D'Amico, Ronald, Alexandria Martinez, Jeffrey Salzman and Robin Wagner. 2002. *An Evaluation of the Individual Training Account/Eligible Training Provider Demonstration: Interim Report*. Report prepared by Social Policy Research Associates under contract to Mathematica Policy Research and the U.S. Department of Labor.

Devine, Theresa, and James Heckman. 1996. "The Structure and Consequences of Eligibility Rules for a Social Program." In *Research in Labor Economics, Volume 15*. Ed. Solomon Polachek. Greenwich, CT: JAI Press, pp. 111-170.

Dickinson, Katherine, Richard West, Deborah Kogan, David Drury, Marlene 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 88-16. Washington, DC: National Commission for Employment Policy.

Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretive Review." *Journal of Human Resources*. 37(4): pp. 696-727.

Eberts, Randall, Christopher O'Leary, and Kelly Durango. 2002. "A Frontline Decision Support System for One-Stop Centers." In *Targeting Employment Services*. Eds. Randall Eberts, Christopher O'Leary and Stephen Wandner. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, pp. 337-380.

Friedlander, Daniel. 1988. *Subgroup Impacts and Performance Indicators for Selected Welfare Employment Programs*. New York, NY: Manpower Demonstration Research Corporation.

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

Heckman, James, Carolyn Heinrich and Jeffrey Smith. 2002. "The Performance of Performance Standards." *Journal of Human Resources*. 37(4): pp. 778-811.

Heckman, James, Robert LaLonde and Jeffrey Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics, Volume 3A*. Eds. Orley Ashenfelter and David Cards. Amsterdam: North-Holland, pp. 1865-2097.

Heckman, James, and Jeffrey Smith. 1998. "Evaluating the Welfare State." In *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*. Ed. Steiner Strom. New York, NY: Cambridge University Press for Econometric Society Monograph Series, pp. 241-318.

Heckman, James and Jeffrey Smith. 1999. "The Pre-Programme Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies." *Economic Journal*. 109(457): pp. 313-348.

Heckman, James, and Jeffrey Smith. 2004. "The Determinants of Participation in a Social Program: Evidence from JTPA." *Journal of Labor Economics*. 22(2): pp. 243-298.

Heckman, James, Jeffrey 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, Volume 7. Ed. Gary Libecap. Greenwich, CT: JAI Press, pp. 191-218.

Heinrich, Carolyn. 1995. Public Policy and Methodological Issues in the Design and Evaluation of Employment and Training Programs at the Service Delivery Area Level. Ph.D. dissertation, Irving B. Harris Graduate School of Public Policy Studies, University of Chicago, Chicago, IL.

Heinrich, Carolyn. 1999. "Do Bureaucrats Make Effective Use of Performance Management Information?" *Journal of Public Administration Research and Theory*. 9(3): pp. 363-393.

Heinrich, Carolyn. 2000. "Organizational Form and Performance: An Empirical Investigation of Nonprofit and For-Profit Job Training Service Providers." *Journal of Policy Analysis and Management*. 19(2): pp. 233-261.

Heinrich, Carolyn. 2003. "Improving Performance Management in the Workforce Investment Act Programs." Working Paper, University of North Carolina at Chapel Hill.

Heinrich, Carolyn. 2004. "The Role of Performance Standards in JTPA Program Administration and Service Delivery at the Local Level." In *Performance Standards in a Government Bureaucracy*. Ed. James Heckman. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, forthcoming.

Kemple, James, Fred Doolittle, and John Wallace. 1993. *The National JTPA Study: Site Characteristics and Participation Patterns*. New York, NY: Manpower Demonstration Research Corporation.

Marschke, Gerald. 2002. "Performance Incentives and Bureaucratic Behavior: Evidence from a Federal Bureaucracy." Working paper, Department of Economics, SUNY Albany, Albany, NY.

Marschke, Gerald. 2003. "The Economics of Performance Incentives in Government with Evidence from a Federal Job Training Program." In *Quicker, Better, Cheaper: Managing Performance in American*

Government. Ed. Dall Forsythe. Albany, NY: Rockefeller Institute Press, pp. 61-97.

Orr, Larry, Howard Bloom, Stephen Bell, Fred Doolittle, Winston Lin, and George Cave. 1996. *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*. Washington, DC: Urban Institute Press.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature*. 37(1): pp. 7-63.

Radin, Beryl A. "The Government Performance and Results Act (GPRA): Hydra-Headed Monster or Flexible Management Tool." *Public Administration Review*. 58(4): pp. 307-316.

Radin, Beryl A. "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): pp. 111-135.

Spaulding, Shayne. 2001. *Performance-Based Contracting under the Job Training Partnership Act*. Master's Thesis, John's Hopkins University, Baltimore, MD.

U.S. Department of Labor. 2000a. "Core and Customer Satisfaction Performance Measures for the Workforce Investment System." Training and Employment Guidance Letter No. 799. Washington, DC: Employment and Training Administration.

U.S. Department of Labor. 2000b. "Negotiating Performance Goals; and Incentives and Sanctions Process under Title I of the Workforce Investment Act." Training and Employment Guidance Letter No. 899. Washington, DC: Employment and Training Administration.

U.S. Department of Labor. 2002. "Summary Report on WIA Implementation." Washington, DC: Employment and Training Administration.

U.S. General Accounting Office 2002. *Workforce Investment Act: Improvements Needed in Performance Measures to Provide a More*

Accurate Picture of WIA's Effectiveness. Washington, DC: U.S. General Accounting Office.

West, Thomas and Gerard Hildebrand. 1997. "Federal-State Relations." In *Unemployment Insurance in the United States*. Eds. Christopher O'Leary and Stephen Wandner. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, pp. 545-598.

Wilson, James. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York, NY: Basic Books.

Zornitsky, Jeffrey, Mary Rubin, Stephen Bell, and William Martin. 1988. "Establishing a Performance Management System for Targeted Welfare Programs." National Commission for Employment Policy Research Report 88-14. Washington, DC: National Commission for Employment Policy.