

**Performance Funding in Federal Agencies:  
A Case Study of a Federal Job Training Program**

Pascal Courty and Gerald Marschke<sup>1</sup>

February 2003

ABSTRACT: Our case study highlights important details that enter into developing performance contingent budgeting schemes---details that do not emerge in more general discussions of the subject---and shows how the handling of these details can be crucial to these schemes' success. We study a federal job training program that gives state and local decision makers discretion over the program's operation, but through performance funding holds them accountable for achieving specific objectives. We find that states' modifications to the scheme's construction produced over time highly individualized performance funding schemes that likely varied in their effectiveness.

---

<sup>1</sup> Pascal Courty, assistant professor of economics, London Business School, Regent's Park, NW1 4SA, London, UK, [pcourty@london.edu](mailto:pcourty@london.edu). Gerald Marschke, assistant professor of public administration & policy and of economics, Department of Public Administration & Policy and Department of Economics, Milne Hall 308, University at Albany, State University of New York, Albany, NY 12222, [marschke@albany.edu](mailto:marschke@albany.edu).

## INTRODUCTION

Performance measurement, whether in performance budgeting or in performance funding schemes, is a cornerstone of the reinventing government movement.<sup>1</sup> Its advocates believe that by conveying goals, motivating government workers, and holding them accountable, performance measurement improves the efficiency of government agencies (see, e.g., Osborne and Gaebler).<sup>2</sup> This is the logic behind the Government Performance and Results Act (GPRA) of 1993, which now requires most federal bureaus to formulate and submit performance goals and measures of performance outcomes along with their budgetary requests to Congress.<sup>3</sup>

Many studies treat the subject of performance measurement at a highly general or abstract level.<sup>4</sup> Here we study the performance measurement and funding regime in place in a real-world government agency. Our case study is the bureaucracy created by the Job Training Partnership Act (JTPA) of 1982. From the early 1980s through the end of the 1990s, JTPA was one of the largest federally funded job training programs for the economically disadvantaged.<sup>5</sup> JTPA was also one of the early large-scale experiments with financially backed performance incentives. JTPA was highly decentralized: training activities were managed by approximately 640 local, independent job training centers who wielded substantial discretion over who was enrolled and how they were trained. In addition to being decentralized, JTPA was characterized by a performance incentive system in which training centers received part of their budget contingent on attaining standards in several areas of measured performance. Our description of the incentive scheme reveals some of the challenges that must be met in formulating performance incentives in government. In particular, the decisions about what should be measured, how, when, and by whom it should be measured make a critical difference for the success of incentive-backed performance measurement.

JTPA was a government organization conceived in the spirit of New Federalism. Proponents of New Federalism have argued that more decentralized decision-making leads to “laboratories of the states” that foster innovation and creativity and hence, in the end, superior policies. We have found that states indeed used their discretion to produce a wide variety of performance measurement and incentive structures under JTPA. In JTPA, the federal government retained control over some important aspects of the incentive

system. We show that the discretion left to the states, however, was important also for determining the character of performance measurement and the incentives, possibly in ways unanticipated by the program's federal administrators. We show below that these differences generated substantial variation across states in the size of an award due for a given level of performance.

By providing an analytical description of performance measurement in JTPA, this paper complements the institutional literature on the JTPA bureaucracy.<sup>6</sup> Although this literature provides an understanding of how JTPA's performance incentives worked at the federal level, it has little to say about its implementation at the state level. By offering a more complete description of JTPA performance incentives in practice, this paper lays the foundation for understanding how performance incentives determined agency behavior and program outcomes. Previous studies of the impact of JTPA's performance incentive scheme on agency behavior<sup>7</sup> have considered only the effects of federal guidelines and these provide an incomplete and possibly misleading representation of the true incentive systems.<sup>8</sup> We agree with Wholey<sup>9</sup> that case studies, such as this one, of operational performance measurement systems can contribute importantly to developing theories to understand and recommendations to improve performance measurement in the public sector.

The paper is organized as follows. The next section briefly introduces the two key aspects of JTPA's performance-based incentive system: the measures of bureaucratic performance upon which the JTPA bureaucrats' awards are contingent, and how JTPA uses bureaucrats' performance to compute their award. We organize the paper around these two aspects, answering the fundamental questions, *what counts?* and *how does it count?*, because these two questions are central to any performance measurement system, and how the questions are answered explain to a large extent the success of the system. The third section describes the organization of JTPA and how federal and state authorities share the tasks of constructing and implementing performance-based incentives. The fourth section explains the set of performance measures in place in JTPA and how performance outcomes across the population of training centers are adjusted to "level the playing field." The incentive award is described in the fifth section. The sixth section describes what is known about the incentives under the JTPA's new predecessor program, the Workforce Investment Act. The final section concludes and summarizes our main findings.

## **JTPA'S INCENTIVE SYSTEM: WHAT MATTERS AND HOW IT WORKS**

Our description of the JTPA incentives is organized around two fundamental questions. First, *Which performance outcomes are rewarded?* According to the Act, Congress intended the performance incentives to measure the training centers' success in developing participants' labor market-specific human capital.<sup>10</sup> Because direct measures of human-capital value added were unavailable, JTPA's federal overseers resorted to proxies of value added. A focus of this paper is a discussion of these measures, their consequences for programmatic outcomes, and their variation over time and across states.

Second, *What is the relationship between performance and awards?* In JTPA, states determined awards in three steps. First, they 'standardized' the performance outcomes to make them comparable across training centers. States rewarded training centers not for the absolute level of performance but for their performance in excess of a numerical threshold, or performance standard. The performance standard depended on control variables (e.g. measures capturing the state of the local labor market) that were specific to each training center. The performance standards were intended to establish 'reasonable' counter-factual levels of performance that one would expect given the environment in which the training center operated. The second step was to establish the training center's eligibility for an award. Training centers were usually eligible only if they exceeded the standards associated with all or a defined subset of the performance measures. Finally, the states formulated award functions that translated training centers' excess performances into budgetary awards. The sensitivity of the award to excess performance determined the strength of the incentive and varied across states and over time. Although the functions were idiosyncratic, from a behavioral point of view they varied along only four important dimensions, which we discuss in detail in fifth section.

### **THE JOB TRAINING PARTNERSHIP ACT**

This section sketches the relevant features of JTPA.<sup>11</sup> The Act called for the delivery of employment and training services to "those who can benefit from, and are most in need of, such opportunities," while requiring that training centers' success be judged on their ability to raise the earnings and employment ability of participants and reduce their welfare reciprocity.<sup>12</sup> The focus of this study is on Title IIA of the

Act, which provided services to adults and youth whose income fell below the poverty line.<sup>13</sup> Title IIA was the largest of the JTPA titles, comprising the majority of the enrollments and expenditures under JTPA.

JTPA divided the U.S. into approximately 640 non-overlapping, collectively exhaustive jurisdictions, each jurisdiction contained wholly within a single state. Congress allocated the annual JTPA appropriation to the states by a formula based on the states' shares of the national JTPA-eligible and unemployed population and by the states' unemployment rates. Each state distributed its allocation among its job training regions by the same formula used in the Congressional allocation to the states. Each sub-state region was administered by a single administrative agency, which we call in this paper, a “job training center” or “agency.”

JTPA was not an entitlement. While the annual budget for JTPA averaged approximately \$5 billion during the period under study, the budget was insufficient to fund the training of all JTPA-eligible persons. Thus, training centers decided who among the eligible to enroll. Moreover, federal rules did not substantially limit the length, kind, or expense of training activities.<sup>14</sup> Training centers, therefore, enjoyed wide discretion over who they enrolled, how many they enrolled, and the kinds of training they offered their enrollees.

The Act charged the U.S. Department of Labor (DOL) with interpreting the law and overseeing the program's implementation. Among the DOL's responsibilities was the design and implementation of an incentive-backed performance measurement system. JTPA set aside six percent of the annual appropriation for distribution as awards to successful training centers. A state was to allocate awards to each training center based on its performance relative to numerical standards. The federal government established the basic structure of these incentives. The DOL defined mandatory performance measures for states to use. The DOL also defined the numerical standards by which states were to judge performance. Finally, the DOL established performance accounting rules---rules for obtaining, compiling, and reporting performance data.

While the federal government established the basic structure of the incentive-backed performance system, the states maintained important discretion over its implementation. States were permitted to

formulate and implement their own performance measures to add to the DOL-defined performance measures. States defined the terms under which training centers received budgetary awards, subject to the federal requirement that only performance above a standard was rewarded. The DOL gave states discretion over whether and how these federally established numerical standards were adjusted for differences in demographics and economic conditions across training centers. While the Act provided for funding managerial training activities---“technical assistance”---for poorly performing training centers, the states determined the rules governing how technical assistance resources were allocated.<sup>15</sup>

Because of the difficulty and expense of obtaining state incentive policies, we limit our description of the state incentive and performance measurement policies to the years 1987-1989 for a sample of sixteen states (identified in Table 2).<sup>16</sup> We collected data on these sixteen states because these states contained the sixteen training agencies that participated in the late 1980s DOL-commissioned National JTPA Study (NJS). The NJS was an important experimental study involving approximately 20,000 enrollees that was designed to measure the impact of job training in JTPA on participants’ earnings and employment prospects.<sup>17</sup> The incentive policy data that we collected can be used with the National JTPA Study data to study the impact of JTPA’s incentive-backed performance measurement on bureaucratic behavior.<sup>18</sup>

Training agencies that were successful during a fiscal year received budgetary awards that they could use to augment their regular budgets in the following fiscal year. States decided how to allocate their portion of the incentive program set-aside between the awards for successful training centers and administration of the incentive programs.<sup>19</sup> This discretion produced variation in the size of the awards available for training centers; the states that we examine over the years 1987 through 1989 set aside between 66 (e.g. Montana) and 100 percent (e.g. Missouri) of their incentive award funds for award disbursements. The average training center received an award equivalent to seven percent of its budget.<sup>20</sup> The highest disbursements were equivalent to 60 percent of the training center’s budget. Rewards this large were not possible in all states, however.

## **PERFORMANCE MEASURES**

Section 106(a) of the Act directed the DOL to formulate measures of performance that captured “the increased employment and earnings of participants and the reduction in welfare dependency.” Reliable measures of earnings and employment *impacts* of job training, however, were prohibitively costly to obtain. Instead, the DOL issued measures of labor market *outcomes* of enrollees at training end.

The Act required states to use the DOL’s performance measures in constructing their incentive system. The DOL permitted states to impose additional performance measures, and many states did. Whereas the federally designed performance measures were concerned with labor-market outcomes, state measures (these are described below) tended to focus on inputs. While many states developed their own measures, they typically devoted a disproportionately small share of the award to them, leaving the bulk of the award for the federally designed measures.

This section discusses the federal and state-designed measures and how states have divided their awards between them. In addition, this section discusses the possible consequences of the construction of these measures for JTPA training practices. Over time, state and federal authorities have replaced or redefined those performance measures whose effects have possibly been counter-productive. We also discuss these modifications.

### *Federal Performance Measures*

Table 1 defines the federal performance measures in place during the period 1987-1989. For the adult portion of the program, the system’s performance measures were *employment rate at termination*, *average wage at termination*, *cost per employment*, *employment rate at 90 days after termination—i.e., at follow-up*, *average weeks worked at follow-up*, and *average weekly earnings at follow-up*. For the youth portion of the program, the system’s performance measures were the *employment rate at termination*, the *cost per employment*, the *positive termination rate*, and the *employability enhancement rate*. The youth employment rate at termination and youth cost per employment were defined as for adults. Youth positive termination rate and youth employability enhancement rate evaluated the acquisition of certain kinds of general or labor

market skills, such as the completion of a major level of education, or completion of a G.E.D. certification (see Table 1, especially note 3).

[Put Table 1 about here]

All federal performance measures were (1) year-end summaries of yearly cumulated performance, (2) based on aspects of the enrollee's labor market status on the date the enrollee officially terminated the program or at 3 months after termination, and (3) averages of outcomes over the population of the year's *terminees* (not participants). Thus, training centers did not face a piece rate in the sense that training centers received compensation per unit of output: e.g., per enrollee employed, or per dollar increase in an enrollee's earnings ability. Instead, training centers received awards for achieving high *average* labor market outcomes. For example, the employment rate at termination for the fiscal year 1987 was defined as the fraction of persons terminated between July 1, 1987 and June 30, 1988 who were employed on their termination date.<sup>21</sup> Awards were thus independent of the number of persons who obtained high outcomes.

Table 2 reports the use of performance measures by state and year of the study for the adult portion of JTPA. (We omit the youth measures from Table 2 due to space considerations.) Table 2 documents two important trends in the use of performance measures in the 16 states of the study: the elimination of the cost measures in many states, and the gradual de-emphasis of measures based on labor market outcomes at the time of termination in favor of measures of outcomes at follow-up.<sup>22</sup> The DOL originally included cost measures to encourage training centers to use their resources efficiently. In the wake of a series of studies suggesting that cost measures discouraged the enrollment of the least skilled---and hence the most costly to serve, the DOL began phasing out cost measures. By 1992, the DOL prohibited the states from using any cost measures of performance. Additional studies seemed to show that training centers were emphasizing "quick fixes" with job placement-oriented services that had no long-term impact on enrollees' skills. Thus, at the same time the DOL was phasing out the cost measures, it began formulating a number of follow-up measures, that is, measures based on outcomes assessed 3 months after termination. The employment rate at termination and average wage at termination gave way to the employment rate, the average weeks worked, and average weekly earnings at follow-up (3 months after termination). (See Tables 1 and 2.) The DOL

introduced follow-up measures to “[promote] effective service to participants and [assist] them to achieve long-term economic independence.”<sup>23</sup> After 1992, the DOL eliminated all termination-based measures in favor of follow-up measures.<sup>24</sup>

[Put Table 2 about here]

Table 2 also shows that performance measures in use varied considerably across states, especially in the years 1988 and 1989. In 1987, the DOL required states to base incentives on the seven termination-based measures. In 1988-1989, as the DOL began its phase-out of termination-based measures, it extended the list of measures to the 12 described in Table 1 and required states to choose any 8 or more of these 12 measures.

Performance measures based on labor market outcomes may have influenced training center behavior in several important and unintended ways. First, because they measure aspects of an enrollee’s employment state and not the impact of job training, they may have lead training centers to select enrollees most able to achieve high levels of employment at high wages, instead of the enrollees most likely to benefit from the program. This behavior has been called *cream-skimming* in the JTPA literature.<sup>25</sup>

Second, just as the DOL suspected, the performance measures may have over-encouraged training centers to offer employment-oriented services. As opposed to the more intensive kinds that result in greater skill development, these services would more frequently lead to an employment on the measurement date, whether or not the employment match was suitable and likely to last beyond the end of training. The reason why quick fixes may have been preferred is that the employment and wage measures focused on labor market success at a point in time rather than over a period of time.

Third, basing performance measures on average outcomes instead of aggregate outcomes may have encouraged training centers to spend less than their entire budget. The optimal training strategy from a pure performance point of view was to enroll only the most promising applicant. More generally, enrolling a smaller than efficient population would typically be an optimal strategy in areas where able applicants were scarce. In these areas, rather than enroll less able enrollees who lower per capita scores, training centers would prefer to leave some of their budget unspent, or to enroll small populations and greatly increase per capita spending, possibly to inefficient levels (Barnow makes this point).<sup>26</sup>

Eliminating the cost measures in the last years of JTPA would have added an additional incentive to reduce the number of enrollees served. A cost measure was defined as the total expenditure divided by the number of persons employed at training end. Holding spending constant, a cost measure would have been an incentive to produce as many employed trainees as possible. This is because, holding the employment rate constant, enrolling more applicants decreased the cost measure.

In short, training centers' responses to the federal performance outcomes may have been dysfunctional. See Blau for an early discussion of unintended responses to performance measures.<sup>27</sup> With the exception of the cream skimming issue, however, these responses have not been studied in JTPA.

### *Federal Performance Standards*

The impact of the performance measures on bureaucratic behavior depended critically on the training center's minimum performance threshold. A training center did not receive an award for the wage measure, for example, if its year-end average wage outcome (the average wage at termination) failed to meet the average wage standard. Thus, the wage measure would produce no effect on behavior if the wage standard were set too high so that no amount of effort would push the wage outcome over the standard. Likewise, standards that were too low and easy to meet would produce no behavioral effect. This subsection describes the heights of the standards and how they were tailored to the different environments faced by the training centers.

For each performance measure, the numerical standard started with the national "departure point". The DOL set the departure point for all but the cost and wage measures at the 25th percentile of the distribution of performance in the system in the preceding two years. That is, 75 percent of the training centers would have exceeded the departure point on average. For the cost measure, successful training centers had to produce an outcome *below* the standard. The DOL set the departure point for the cost measure at the 90<sup>th</sup> percentile.<sup>28</sup>

The difficulty of meeting these departure points varied across training centers. The difficulty varied because labor markets, training costs, and the characteristics of the eligible populations varied. Imposing

uniform standards would have favored low-difficulty training centers by increasing their resources relative to high-difficulty training centers. Only in the case that low-difficulty training centers tended also to be more efficient would such incentives enhance the efficiency of the allocation of training resources. Believing that this probably was not the case, the DOL established an adjustment model that took into account features of the training center's environment presumed to be correlated with the difficulty of attaining departure points. For example, by taking into account local unemployment measures and other measures of the labor market, the adjustment methodology lowered the employment rate standard for training centers in depressed job markets, compared to training centers in robust job markets (e.g., an above average unemployment rate would earn a training center a lower standard).

An important class of characteristics for which the DOL adjusted standards was the composition of the enrollment pool. While the enrollment pool reflected in part the composition of the local eligible population (an influence beyond the training center's control), it was at least partly a choice variable. Adjusting the performance standard in this way may have encouraged the training center to enroll not only persons who would boost performance outcomes, but also persons who would lower standards. To our knowledge, this has not been tested in the cream skimming literature.

The DOL adjustment model did not take into account training services as a relevant control variable, although these data were available to the DOL. Thus, the nature of the adjustment procedure meant that the incentive system held training centers accountable for the kinds of training provided but not the kinds of enrollees enrolled, yet both choices have consequences for the effectiveness and efficiency of training.<sup>29</sup> Although the DOL allowed states some flexibility in developing standards, most states used the DOL's adjustment methodology.<sup>30</sup>

### *State Performance Measures and Standards*

While state measures played a smaller role than federal measures in the determination of the awards, the number of NJS states using their own measures increased from 10 to 13 between 1987 and 1989. The increase in importance of state-formulated measures was apparently a nation-wide trend.

Although state-defined measures were common among the NJS states, they comprised a relatively small fraction of the award. Excluding New Jersey, the average split between federal—or performance-based—measures and state measures was 82 percent/18 percent. The split ranged from 50/50 to a full allocation of the money to the federal award. The most important category of state measures comprised input or enrollment measures. In the mid-1980s, states became increasingly concerned that federal performance incentives were driving training center managers to enroll from the eligible population only those enrollees who were likely to get jobs at the end of training—i.e., who were “job ready”. Many states implemented a set of enrollment-based performance measures designed to encourage training centers to enroll the more difficult cases. Table 3 shows the use of these measures by state and year. In 1988, for example, 9 out of 16 states set up standards that compensated training centers for the number of or the rate at which persons in target groups were enrolled. These target groups varied by state but were typically the least successful in the labor market among the eligible, and included high-school dropouts, minority youths, Women/Infant Nutrition program participants, and older workers. For example, in addition to compensating performance based on federal performance measures, Minnesota rewarded training centers for the fraction of enrollees who were receiving public aid. In these cases, compensation was contingent upon meeting a numerical standard, frequently based on the fraction of eligible persons in the local population who belonged to the target group. Studies of the effect of incentives on the enrollment decision can be further refined by allowing for these enrollment quotas to influence the enrollment decision.

[Put Table 3 about here]

The other categories of state measures were more idiosyncratic. Some states compensated training centers for the fraction of their budgetary allotment spent. For example, in 1987 and 1988 Mississippi paid a portion of its award money to training centers that spent at least 85 percent or more of their budgets. These states may have felt this kind of standard necessary because training centers might have left portions of their budgets unspent when, as we noted above, the kinds of applicants who would produce high performance outcomes were scarce.

Three of the 16 states sought to encourage JTPA training centers to coordinate their activities with other state agencies that helped the poor. In promoting these goals, states typically evaluated performance subjectively, without well-defined performance standards. Finally, although state performance measures usually were not based on participant labor market outcomes, some states used their own measures to encourage training centers to seek longer-term employment matches before the DOL began offering the follow-up measures in 1988. In 1987, New Jersey used a separate measure for employment retention, similar to the federal employment rate at follow-up measure.

### **THE JTPA AWARD**

States' eligibility rules determined which training centers received awards. The award functions translated the performance of eligible training centers to award amounts. The states were free to design the eligibility rules and the performance awards as they saw fit, which led to great variation in both across states. Because of space constraints, we do not report the exact computation rules for the awards. Instead, we define some broad dimensions that are important from a behavioral point of view, to categorize the different types of award functions, illustrating where appropriate with specific details from the state incentive systems.

#### *Qualifying for Awards*

Table 4 shows the qualifying criteria for the 16 NJS states in years 1987-89. Some states (e.g., Indiana) required training centers to meet all standards as a prerequisite for earning any award money. Other states required training centers to exceed a subset of standards to qualify. For example, in 1987 Minnesota required training centers to exceed five of seven performance standards to qualify for awards. Other states had no qualification criteria. These states (e.g., Iowa) simply rewarded training centers for each performance standard exceeded.

[Put Table 4 about here]

Some qualification criteria were quite complicated. For example, in 1987, Illinois divided the seven federal measures in place at the time into three groups. To qualify for an incentive grant, a training center

had to meet both standards in the first group, one of two standards in the second group, and one of three standards in the third group.<sup>31</sup> In addition, the training center had to meet a slightly higher version of the standards for at least one measure.

The number and kind of standards a training center had to meet to win an award should have been an important determinant of the influence of incentives upon behavior. States that required the training center to meet all standards discouraged training centers from specializing in the production of certain performance outcomes at the expense of others. Moreover, the greater the number of performance standards that training centers were required to meet, the lower their likelihood of obtaining an award. By lowering the chances to qualify for an award, such qualification criteria may have discouraged training centers from attempting to win awards.

#### *Award Function*

The award functions varied along four important dimensions. First, different state objectives imposed different weighting schemes for the various performance measures. For example, to encourage more intensive training services, some states limited the size of the payoff for meeting the federally-defined cost standard. Second, states varied in the potential size of the award for any single training center. In many states, the award was scaled by the size of the training center. In other states, the incentive system resembled a winner-take-all tournament that could enlarge a successful training center's budget by 50 percent or more. Third, states differed in the extent to which a training center's performance ranking determined its award. Under certain conditions, judgments based on performance rankings have an advantage over judgments based on absolute performance because they allow states to more accurately observe and reward effort (see, e.g. Gibbons and Murphy).<sup>32</sup> Fourth, the marginal award for performance in excess of the standards varied. Some states offered an award that increased with the difference between performance and the standard while others paid only for meeting the standard. One would expect that the former type of policy would produce more high-performers than the latter type. These dimensions of the award function are explored in detail below.

*Performance measure weighting.* We have shown that by exercising their discretion over which measures they included in their awards (see Tables 2 and 3) and in their construction of the eligibility criteria (see Table 4) states could emphasize some performance measures and de-emphasize others. In addition, states used explicit weighting schemes in the award function for the same purpose. Although many states weighted each measure they used equally, some states weighted performance measures differently to emphasize some measures over others.

New Jersey (in 1987 and 1988) illustrates the use of performance weighting. A large fraction of the award in New Jersey depended on a summary performance variable that was computed independently for each training center. This summary variable was equal to the sum of the weighted excess performance (difference between outcomes and standards) for six federal performance measures excluding the cost measures. While training centers in New Jersey were required to meet the adult and youth cost standards to qualify for an award, the sum put no weight on these measures; that is, cost measures did not directly affect the size of the award.<sup>33</sup> In other states, cost measures figured into both the qualifying criteria and the calculation of the awards. In addition, New Jersey placed higher weights on follow-up performance measures relative to termination-based measures to encourage training centers to offer the kinds of training that produced long-lasting impacts on enrollees' labor market success.

*Competition among training centers.* In many states (e.g., Texas), a training center's award depended only on its own performance. Such states divided the total fund into training center-specific award accounts, limiting each training center's award to no more than the funds in its award account.<sup>34</sup> In Illinois' scheme, the size of a successful training center's award depended on the number of training centers that qualified: the lower the number of training centers that qualified, the greater the allocation to the qualifying training centers. Illinois illustrates one kind of incentive scheme in which the performance of one training center affected the compensation of others.

Other examples of interdependence more closely resemble the relative performance evaluation schemes found in the incentive literature in economics. New Jersey is an example of a state that pitted training centers against one another in a form of head-to-head competition. Recall that New Jersey computed a single

summary performance measure for each training center. Then, for a portion of the award fund, New Jersey ranked the training centers and paid only the top five performing training centers.<sup>35</sup> In New Jersey's tournament system, a training center received an award based not on an absolute level of performance but on its position in a ranking of its fellow training centers.

Evaluating a training center on its relative performance may have stimulated competition among training centers by accentuating social comparisons. Another reason for relative performance evaluation is that it holds training centers harmless for influences that are beyond the training center's control and affect all training centers uniformly. Thus, relative performance evaluation should be most effective in states where training centers operate in similar environments and where performance standards do a poor job at controlling for factors that are outside the training centers' control.<sup>36</sup>

*Marginal incentive.* The marginal incentive measures the change in incentive award for a small change in performance. Marginal incentives are constant when the award function is linear as in a piece rate compensation system. When the marginal incentives vary with the level of performance, incentives are said to be nonlinear. In JTPA, the main source of nonlinearity was the performance standard: in many states, training centers were paid contingent upon achieving standards. As discussed in previous section, for most performance measures the standard was set quite low, at the 25th percentile of system-wide performance. Moreover, many states (e.g., Georgia) paid out the entire award merely for meeting the standards. Such states provided no pecuniary incentive to exceed the standards. Other states, however, compensated training centers for performance in excess of the standard, at least over some range of performance.

Figures 1.I-V show the different award functions represented among the states of the NJS. For the sake of presentation, we present only five stylized award functions that capture most of the heterogeneity in our sample. Figures 1.I-IV depict the size of the award paid out for different levels of *excess* performance, that is, the performance outcome ( $S$ ) minus the performance standard ( $\bar{S}$ ), under different incentive schemes for qualifying training centers. In panels I-IV, negative excess performance ( $S - \bar{S} < 0$ ) generated no award. Figure 1.V depicts the award for one performance measure in New Jersey in the years 1988 and 1989.<sup>37</sup> Table 5 identifies each state with each award type depicted in Figure 1. Note that three states—California,

New Jersey, and Illinois—used several different kinds of awards simultaneously. These three states split the award and allocated the various parts under different rules. For example, California distributed the award to eligible training centers in two steps. First, it gave an automatic and uniform lump sum for each standard met. Then, it allocated the remainder across the training centers in proportion to their excess performance.

[Put Figure 1 about here]

[Put Table 5 about here]

The award represented by Figure 1.I is flat; that is, the training centers received lump sums for achieving the standard. There was no marginal incentive for performance beyond the standard. Figures 1.II and 1.III show the training center receiving a lump sum at the standard, and additional amounts for performance beyond the standard. Nevertheless, in both figures, the award plateaus and the marginal incentive eventually drops to zero.

The Illinois incentive system in 1987 resembled Figure 1.II. That is, a training center in Illinois received an award based on the number of standards exceeded and on the extent to which each standard was exceeded. Illinois set aside for each training center a portion of the state's incentive fund based upon the training center's relative size. This amount was divided equally into seven shares with one share associated with each standard. Each of these seven shares was further divided into five margins. For example, if a training center in Illinois exceeded a standard by ten percent, it qualified to receive up to 60 percent of the incentive grant funds. By exceeding the standard by 15 percent, it qualified to receive another ten percent (for a total of 70 percent) of the incentive award funds. It qualified to receive its entire award if it exceeded its standard by 40 percent. For performance in excess of 40 percent of the standard, the training center received no additional compensation. The marginal incentives for performances above 40 percent of the standard were zero.

*Number and size of the training centers.* Incentive systems varied by whether a training center's incentive award reflected the number of training centers in its state and the relative sizes of their budgets. The number of training centers within a state varied greatly across states. In addition, some training centers

managed larger training budgets than others. Given a level of performance, should the award depend on the number of training centers? Moreover, should the award depend on their relative sizes?<sup>38</sup>

Different states addressed the issues of relative size and number of training centers differently. Under the Texas system, for example, the awards did not depend on the number of training centers because there was an individual award account set aside for each training center. In addition, the award account was scaled by the budget sizes so that for a given level of performance larger training centers earned proportionally larger awards.

Figure 1.IV illustrates how the number of training centers in a state mattered in the calculation of a training center's award. This figure describes an incentive scheme under which a training center's marginal incentive at the standard was positive and continuous. For the five states in our sample that had such award schemes, a training center's award was proportional to its share of the sum of excess performance over all training centers. Such training centers' awards increased with excess performance, but slowly, especially in states with many training centers. Awards increased more slowly in excess performance in states with many training centers because a single training center's contribution to the total excess performance generated by all training centers in the state was small. In states with few training centers, a training center's award was more sensitive to changes in its excess performance. Figure 1.IV depicts both cases.

*Discontinuities and bunching.* Nonlinearities (e.g. curvature or discontinuities) in the award function may have influenced the timing of effort over the program year. To start, consider the award function whose profile resembles Figure 1.I. This simple incentive system has the disadvantage that the training center faces no incentive to increase performance once the target is reached or after a poor start leaves the target hopelessly out of reach. Linear systems, on the other hand, exert a uniform "incentive pressure" that makes training centers want to increase performance regardless of how things have gone in the past.

More generally, under nonlinear incentive systems, the training center's effort and output should typically vary depending on past performances and the location of the nonlinearities. A model of training centers based on incentives when outcomes are costly to produce would predict bunching in outcomes at the discontinuities in the award functions. Because states paid a large fraction of the potential award for simply

meeting the standard, one would expect bunching in performance outcomes at the standard. In addition, for Figures 1.II and 1.III, the model would predict bunching at the performance thresholds denoted  $u$  on the horizontal axis. For states using an award function resembling 1.II, bunching may occur at the discontinuities in the interior of 1.II, as well.

These discontinuities in the award may have produced another kind of bunching. As with many incentive systems based upon attaining standards, training centers may have been able to manipulate the award intertemporally by selectively choosing when to report good and bad performances. In this way, training centers would have been able to show increased performance outcomes without actually increasing the effectiveness of their training. See Courty and Marschke for a discussion and some evidence of deadline effects in training centers.<sup>39</sup>

#### *The Consequences of the Interstate Variation in Incentives: An Example*

We consider the following example to illustrate the consequences of the interstate differences in incentive policies on training center incentives. Consider a training center in 1987 whose annual budget was \$2 million (this represents a training center that serves a substantial, but not atypically large, population of JTPA-eligible persons). In 1987, most states included seven performance measures: ERT, WERT, AWT, CE, YERT, YPTR, and YCE (see Table 1 for definitions). Suppose the training center's performance with respect to each of these measures in 1987 was the following. For each measure ERT, WERT, YERT, YPTR, and YCE, the training center exceeded its standard by three percent. For measures AWT and CE the training center did not achieve its performance standards.

Now we consider the awards due for such a training center with such performance in Montana, Colorado, and Missouri. In Montana, because the training center did not exceed all standards, it would not qualify for an award (see Table 4). In Colorado, the training center would be eligible for an award. In 1987, Colorado set aside for each training center a portion of the state award pot that was proportional to the training center's budget. Training centers that exceeded all standards by more than five percent earned their entire set-aside. Training centers that exceeded standards by less than five percent, however, earned far less. Our

hypothetical training center would have earned in Colorado, based on its size and performance and the state's award pot, only about \$2,443. (If instead of exceeding its standards by three percent, the training center exceeded them by six percent, it would have earned over \$81,000.) In Missouri, training centers qualified for an award if the weighted sum of their excess performances exceeded zero (see Table 4, column 1). The performance hypothesized would have qualified this training center for an award. In Missouri, a training center's award was a function of both the training center's size and its performance relative to the size and performance of other eligible training centers in the state. Assuming all training centers in the state of Missouri performed at the same level as the hypothetical training center, given the state's award pot that year, this training center would have earned \$39,920, about two percent of its budget, or 16 times what it would have earned in Colorado.

#### **WORKFORCE INVESTMENT ACT<sup>40</sup>**

In 2000, a new program created under the Workforce Investment Act (WIA) of 1998 supplanted JTPA. WIA retains much of the decentralized nature, the jurisdictional borders, administrative entities, and the performance incentives of JTPA. WIA represents a departure from JTPA in several potentially important ways. For example, WIA contrasts welfare and job training services geographically, a principle it calls "one stop shopping." Also, rather than assigning job training enrollees to job training providers, WIA training centers issue vouchers to eligible participants who then shop among a list of approved service providers.

WIA's performance incentive system is new and therefore unstudied. We know from the legislation and from the DOL's policy letters to the states the general form of the system, however. At the federal level, there are many similarities and some differences between performance incentives under JTPA and performances incentives under WIA. As in JTPA, in WIA most of the performance measures are based on the labor market outcomes of enrollees. All labor market outcomes are measured after training ceases (as opposed to on the date of termination), as in the latter day JTPA. Interestingly, WIA includes among the JTPA-style performance measures a new before-after measure of enrollees' earnings. Conceptually, the difference between an enrollee's earnings before enrollment and after termination is more similar to an

earnings or employment gain---and thus more similar to the objective of job training under JTPA and WIA--  
-than is a post-training labor outcome. While it is more similar to a job training impact, a before-after  
earnings measure too suffers from potential problems.<sup>41</sup> Whether before-after measures lead to less cream-  
skimming is an open question. The performance measures under WIA also include a measure of “customer  
satisfaction” produced from post-training surveys of enrollees and their employers. As in JTPA, states are  
permitted to formulate and implement their own performance measures to add to the federally mandated  
ones.

As in JTPA, in WIA training centers are rewarded for exceeding performance standards. WIA  
performance standards are not formulaically adjusted by the characteristics of the training center, however.  
Instead, states “negotiate” these standards with the Department of Labor. We do not yet know whether this  
negotiation results in standards that are sensitive to training center economic conditions or enrollment  
populations.

Interestingly, under WIA’s performance measurement system, a state’s incentive allocation is no longer  
fixed, but depends on the aggregated performance of its training centers. Only states whose performance  
exceeded the standards associated with all federal and state-designed performance measures are eligible for  
awards. Thus, states cannot qualify by “specializing” along some dimensions of performance---that is, by  
making up for sub-standard performance in one area with exceptional performance in another. A state’s  
incentive allocation also depends on its improvement over previous years’ performance. This is also an  
innovation over JTPA but may lead to the so-called "ratchet effect". That is, basing a state’s award on  
performance improvements may encourage a state to suppress performance, because exceptional  
performance raises the bar and lowers the state’s compensation in subsequent years, everything else  
being equal.

The Workforce Investment Act and the DOL policy guidelines establish how a state’s aggregate  
performance determines its incentive fund allocation---just as under JTPA. The Act and the DOL  
guidelines, at least through the present, appear to leave to the states the same kind of discretion to reward  
individual training centers as the states had under JTPA. Thus, while we know few details about the

structure of the performance-incentive system at the state level, we surmise there exists the same state-by-state variation in the degree of award competition among training centers, in the number of performance standards that must be met for award eligibility, and in the shape of the “award function” as was found under JTPA.

## CONCLUSION

This paper presents a detailed description of the incentives in place in the government program created by the Job Training Partnership Act. Job training programs illustrate well the difficulty of devising performance measures that are aligned with programmatic goals. In the early years of JTPA, performance measures were based on employment outcomes measured at training end, thus possibly encouraging training centers to pursue high employment rates instead of increased earnings capacities. Because they were based on average outcomes, performance measures may have reduced the number of disadvantaged people served, raised the expenditure per enrollee, and produced budget surpluses.

Training centers received awards when the outcomes of performance measures exceeded numerical standards, which the DOL made sensitive to training centers’ environments in an attempt to ‘level the playing field.’ The DOL’s adjustment scheme offers a real-world example of strategies for adjusting performance measures that have been proposed in the literature.<sup>43</sup> Performance standards were also adjusted for the characteristics of persons enrolled, to discourage cream-skimming, but not adjusted for the kinds of training offered. Excluding training from the adjustment method may have promoted employment-oriented services—such as job club and on-the-job training—over intensive job training. This study has revealed numerous ways in which performance measures might have been misaligned with the agency’s goals. With some exception, these distortionary effects have not been investigated. We believe that further research on the behavioral responses to JTPA’s and WIA’s incentives would lead to much useful information for developing and refining performance measures for many kinds of public sector organizations.

JTPA and the DOL also allowed states discretion over the formulation of many of the details of the incentive system designed to hold local agencies accountable. With this discretion, states established

eligibility criteria that modified the JTPA incentive system by reducing the incentive to specialize in the production of one or two performance measures, lowering the training center's likelihood of obtaining an award, holding effort constant, and emphasizing some performance measures over others. The strength of the award varied greatly across states, as well. In some states, training centers received additional money for higher performance, but in other states, they did not. Because it retains JTPA's decentralized nature, WIA may likely produce the same kind of across-state variation in incentive system's policies.

The wide diversity of measurement systems across states has several possible interpretations. First, it could be that there is no "best" measurement system and the observed diversity only reflects the fact that different states have selected different, although equally efficient, systems. Second, the differences in measurement systems may reflect differences in state preferences. There is some evidence that states have designed and implemented their own performance measures to promote different goals than the goals expressed in the federal measures. Inter-state differences in the implementation of the performance incentives grew over time. The findings reported here suggest that the objectives the states' incentive systems encouraged at the end of JTPA's run diverged from the original federal ones. Finally, we have also found that the measurement systems have systematically changed over time both at both the federal and state levels, with trends that seem to indicate a deeper understanding of their impacts. This observation is consistent with the original intent of Congress that decentralization would lead to experimentation and innovation. In that respect, it seems that a valuable objective of research would be to study if the design of WIA has leveraged the JTPA experience.

#### NOTES

The authors wish to thank Jim Heckman, Jeff Smith, and Carolyn Heinrich, as well as the editor and two anonymous reviewers for their helpful comments. All remaining errors are the authors' responsibility.

<sup>1</sup> We use the definitions cited in Megan M. Jordan, and Merl M. Hackbart, "Performance Budgeting and Performance Funding in the States: A Status Assessment," *Public Budgeting and Finance* 19, no. 1 (Spring 1999): 68-88. Performance budgeting is the requirement that an agency submit with its budgetary request to the legislature a set of performance measures by which the legislature will assess the agency's success. Performance funding makes the link between performance and reward explicit. According to Jordan and Hackbart (p. 69), performance funding is "the allocation or distribution of a percentage of the appropriated funds contingent upon the assessment of the performance measures identified in the budget." Note that the heart of both schemes is performance measurement, a focus of this study.

<sup>2</sup> David Osborne and Ted Gaebler, *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. (Reading, Massachusetts: Addison-Wesley, 1992).

<sup>3</sup> For evidence of the degree to which state agencies practice performance measurement, see, for example, Jordan and Hackbart, and Robert D. Lee and Robert C. Burns, "Performance Measurement in State Budgeting: Advancement and Backsliding from 1990 to 1995," *Public Budgeting and Finance* 20, no. 1 (Spring 2000): 38-54. For evidence on county-level agencies, see Xiaohu Wang, "Performance Measurement in Budgeting: A Study of County Governments," *Public Budgeting & Finance* 20, no. 3 (Fall 2000): 102-118.

<sup>4</sup> Discussions of general criteria for choosing performance measures and constructing performance measurement systems are found in, for example, Harry Hatry, "Performance Measurement Principles and Techniques: An Overview for Local Government," *Public Productivity Review* 4, no. 4 (December 1980): 313-15; Brian Usilaner and Edwin Soniat, "Productivity Measurement," in *Productivity Improvement Handbook for State and Local Government*, ed. George J. Washnis (New York: Wiley, 1980); Joseph S. Wholey, "Performance-Based Management: Responding to the Challenges," *Public Productivity and Management Review* 22, no. 3 (1999): 288-307; and U.S. General Accounting Office, *Executive Guide: Effectively Implementing the GPRA*, GAO/GGD-96-118 (Washington, D.C., June, 1996).

<sup>5</sup> In 2000, a new program created under the Workforce Investment Act (WIA) of 1998 supplanted JTPA. Its performance funding scheme is sketched below.

<sup>6</sup> See Shirley Svorny, "Congressional Allocation of Federal Funds: The Job Training Partnership Act of 1982," *Public Choice* 87, No. 3/4 (June 1996): 229-42; Burt Barnow, "The Effect of Performance Standards on State and Local Programs," in *Evaluating Welfare and Training Programs*, eds. Charles Manski and Irwin Garfinkel, Cambridge, Massachusetts: Harvard University Press, 1992): 277-309; and Barnow, Burt, "Thirty Years of Changing Federal, State, and Local Relationships in Employment and Training Programs," *Publius: The Journal of Federalism* 23, no. 3 (Summer 1993): 75-94.

<sup>7</sup> Kathryn Anderson, Richard Burkhauser, and Jennifer Raymond "The Effect of Creaming on Placement Rates Under the Job Training Partnership Act," *Industrial and Labor Relation Review* 46, no. 4 (July 1993): 613-624; James J. Heckman, Jeffrey A. Smith, and Christopher Taber, "What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program," NBER Working Paper No. 5535 (Cambridge, MA: National Bureau of Economic Research, 1996); and Pascal Courty and Gerald Marschke, "An Empirical Investigation of Gaming Responses to Performance Incentives," *Journal of Labor Economics*, in press.

<sup>8</sup> Two exceptions are Michael Cragg, "Performance Incentives in the Public Sector: Evidence from the Job Training Partnership Act," *Journal of Law, Economics and Organizations* 13, no. 1 (1997): 147-168, and Gerald Marschke, "Performance Incentives and Organizational Behavior: Evidence from a Federal Bureaucracy," Manuscript, University at Albany, State University of New York, 2002.

<sup>9</sup> Wholey, 305.

<sup>10</sup> The Job Training Partnership Act, Public Law 97-300, October 13, 1982, Section 106(a).

<sup>11</sup> See, e.g., Janet W. Johnston, *The Job Training Partnership Act: A Report by the National Commission for Employment Policy*, (Washington, D.C.: U.S. Government Printing Office, 1987) for a detailed description of JTPA.

<sup>12</sup> The Job Training Partnership Act, Sections 141(c) and 106(a).

<sup>13</sup> For a more complete description of the JTPA eligibility rule, see Katherine P Dickinson, et al., "Evaluation of the Effects of JTPA Performance Standards on Clients, Services, and Costs," Research Report No. 88-16 (Washington, D.C.: National Commission for Employment Policy, September 1988).

<sup>14</sup> The one significant exception: agencies were not allowed to subsidize a client's on-the-job training for more than six months at a time. In addition to on-the-job training with private sector employers, the main types of training offered under JTPA include vocational classroom training, basic or remedial education, work experience, job counseling, and job search assistance.

<sup>15</sup> In addition, the Act allowed the DOL to revoke the job training center's right to operate if it failed to meet performance standards two years in a row. The DOL apparently did not exercise this right often.

<sup>16</sup> Compiling the state incentive policies was difficult and expensive because each state's incentive policy (the performance measures, performance standard adjustment methodology, the award function, etc.) was typically scattered across several documents and thus dispersed within the state's governor's office. Moreover, because the policies in question were more than three years old---governors' offices often archived, misplaced, or discarded documents that old---we were often forced to seek the information from multiple sources (in all such cases we were eventually able to obtain copies of the state's policies among the state's training centers.) Note that we reconstructed all incentive policies almost exclusively from official state incentive policy guidelines, and relied very little on program officials' recollections. This was necessary because of the high level of detail we were seeking. We will provide these policy documents to the interested reader upon request.

<sup>17</sup> See Fred Doolittle and Linda Traeger, *Implementing the National JTPA Study*, (New York: Manpower Demonstration Research Corporation, 1990), for a description of the implementation of the study, and Larry L. Orr, *The National JTPA Study: Impacts, Benefits, and Costs of Title II-A*, (Cambridge, MA: Abt Associates, 1994), for the JTPA experimental impact findings along with estimates of the program's benefits and costs.

<sup>18</sup> These incentive policy data have been exploited for this purpose by Marschke and James J. Heckman, Carolyn Heinrich, and Jeffrey Smith, "The Performance of Performance Standards," *The Journal of Human Resources* 37, no. 4 (Fall 2002): 778-811. Marschke exploits the variation in incentives across states to measure the influence of incentives on training center behavior (specifically training choice) and efficiency. The NJS data contain information on training centers' enrollment and training decisions and on performance. These data can also be used to construct measures of training center value added. Heckman, Heinrich, and Smith examine the validity of JTPA performance measures by measuring their degree of correlation with various measures of program value added. (See also Carolyn Heinrich, "Do Government Bureaucrats Make Effective Use of Performance Management Information?" *Journal of Public Administration Research and Theory* 9, no. 3 (1999): 363-393, and Burt Barnow, "Exploring the Relationship between Performance Management and Program Impact," *Journal of Policy Analysis and Management* 19, no. 1 (2000): 118-141, for evidence on the relationship between performance measures and program goals.) Obviously, many more questions about performance measurement in JTPA can be addressed with these data. A purpose of this article is to inform interested researchers of the existence of these data and encourage use of these data.

<sup>19</sup> States also used the six percent incentive fund to provide training resources and supervision to training centers' management that failed to meet their performance standards.

<sup>20</sup> These figures are based on the data set of SRI, International and Berkeley Planning Associates for the program year 1987. See Dickenson et al. for a description of these data. The JTPA funds were allocated in three sub-funds: 78 percent were set aside for training services, six percent were set aside for the incentive system, and the remaining 16 percent were set aside for other special services. The award fund as a fraction of total training budget was 7.1 percent ( $6/(78+6)$ ) if one assumes that all award funds were eventually distributed as training budget. The actual figure may have been a little lower because some of the incentive set-aside was spent on incentive fund administration.

<sup>21</sup> Note that persons who entered the program already employed, and then terminate employed, holding the same job they began training with, say, were numbered among the successfully trained—i.e., were *employed at termination*—for award accounting purposes.

<sup>22</sup> From JTPA's start through 1987, the only adult federal performance measures year in year out were the three termination-based employment measures (ERT, WERT, and AWT) and the cost measure (CE). Thus, the first major changes to the federal performance measures occurred in 1988, the middle year of our three-year study period.

<sup>23</sup> State of New Jersey Performance Standards Manual, PY1988-89. Division of Employment and Training, New Jersey Department of Labor, April 1990.

<sup>24</sup> JTPA's emphasis on gathering information on the effects of its performance measures, and then, when the evidence calls for it, revising the performance measures is consistent with the claim of Arie Halachmi, "Performance Measurement: Don't Abuse it through GPRA," (paper presented at the Ninth Public Sector Productivity Conference: Performance Measurement, Performance Improvement, Online conference, November 9-20, 1998), that good performance measures take time to develop. The performance measure reforms are also consistent with the recommendation of Robert Kravchuk, Robert and Ronald Schack, "Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993," *Public Administration Review* 56, no. 4 (July/August 1996): 348-358, that the designers of performance measurement systems should periodically review and revise them. The Act provided for a commission to oversee JTPA for this purpose. See Marschke for some evidence on how well these reforms succeeded.

<sup>25</sup> The evidence that this behavior existed is mixed. See Anderson et al., Heckman, Smith, and Heinrich, Heckman, Smith, and Taber, and Cragg.

<sup>26</sup> Barnow, "The Effect of Performance Standards on State and Local Programs."

<sup>27</sup> Peter Blau, *The Dynamics of Bureaucracy: A Study of Interpersonal Relations in Two Government Agencies*, (Chicago, IL: University of Chicago Press, 1955).

<sup>28</sup> The Department of Labor's JTPA Technical Assistance Guide, PY1988, reports that the wage departure point was set "above the 25th percentile. It more closely resembles an estimate of average performance." While one might interpret this statement to mean the DOL set the wage standard at the 50th percentile, its exact formulation is uncertain.

<sup>29</sup> For more on the specification of DOL's regression model, see Barnow, "The Effect of Performance Standards on State and Local Programs," and Charles E. Trott, and John Baj, *Development of JTPA Title II-A Performance Standards Models for the States of Region V*, (DeKalb, IL: Center for Governmental Studies, Northern Illinois University, 1987). For a general discussion of the theory of adjusting performance measures, what factors should and should not be used to adjust performance measures, see Leanna Stiefel, Foss Rubenstein, and Amy Ellen Schwartz, "Using Adjusted Performance Measures for Evaluating Resource Use," *Public Budgeting & Finance* 19, no. 3 (Fall 1999): 67-87.

<sup>30</sup> State governors had the option to: (1) set the performance standard at the national departure point; (2) adjust the national departure point for specific economic, geographic, and demographic factors within the state or local service delivery areas using the regression model established by the DOL; or (3) propose their own adjustment method to the DOL. Between the end of the National JTPA Study and the end of JTPA, more states abandoned the DOL adjustment method for their own (option 3).

<sup>31</sup> The first group consisted of the adult employment rate at termination and the adult cost per employment. The second group consisted of the youth positive termination rate and the youth employment rate at termination. The third group consisted of the average wage at termination, the welfare employment rate at termination, and the youth cost per positive termination.

<sup>32</sup> Robert Gibbons and Kevin J. Murphy, "Relative Performance Evaluation for Chief Executive Officers," *Industrial and Labor Relations Review* 43, Special Issue (February 1990): 30S-51S.

<sup>33</sup> New Jersey de-emphasized the cost measures "to encourage [training centers] to provide more comprehensive programming and increased services for those individuals who are most in need." (State of New Jersey Performance Standards Manual, PY1988-89. Division of Employment and Training, New Jersey Department of Labor, April 1990.)

<sup>34</sup>In years where the training centers performed poorly, this distribution rule left some award funds unallocated. Some states required that the left-over funds be allocated for technical assistance, some rolled over the funds into the next program year, and some re-distributed the left-over funds to the top performers.

<sup>35</sup> Among the NJS states, between 1987 and 1989, only New Jersey managed a tournament system. However, many states used a system where training centers were paid on the basis of their fraction of excess performance relative to the total excess performance within the state. That type of incentive scheme is similar to tournaments.

<sup>36</sup> The DOL's performance standard adjustment methodology—to the extent that it accounted for external factors that affect performance—produced the same effect.

<sup>37</sup> The award depicted in Figure 1.V is for a state-defined job retention measure. The retention measure was defined as the number of enrollees who were employed continuously for the 60 days following termination divided by the total number of enrollees employed at termination.

<sup>38</sup> See Pascal Courty and Gerald Marschke, "Performance Incentives with Award Constraints," *Journal of Human Resources* 37, no. 4 (Fall 2002): 812-845, for a theoretical discussion of these issues.

<sup>39</sup> Pascal Courty and Gerald Marschke, "Moral Hazard Under Incentive Systems," in *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*, ed. Gary Libecap, 7 (1996): 157-190; Pascal Courty and Gerald Marschke; "Measuring Government Performance: Lessons from a Federal Bureaucracy," *American Economic Review* 87, no. 2 (1997): 383-388; and Courty and Marschke, "An Empirical Investigation of Gaming Responses to Performance Incentives."

<sup>40</sup> This section is drawn from U.S. Department of Labor, "Core and Customer Satisfaction Performance Measures for the Workforce Investment System," Training and Employment Guidance Letter No. 7-99. (Washington, D.C.: Employment and Training Administration, 2000) and U.S. Department of Labor, "Negotiating Performance Goals; and Incentives and Sanctions Process under Title I of the Workforce Investment Act," Training and Employment Guidance Letter No. 8-99, (Washington, D.C.: Employment and Training Administration, 2000).

<sup>41</sup> For the average enrollee, earnings dip just before entering job training, suggesting that her earnings would eventually rise even if the job training program had no value. This phenomenon, the so-called "Ashenfelter dip," means that before-after earnings differences are distorted measures of the true impact of job training. (See James J. Heckman and Jeffrey A. Smith, "The Pre-Programme Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies," *Economic Journal* 109, no. 457 (1999): 313-348, for a discussion of this point.)

<sup>42</sup>Steifel, et al.

TABLE 1  
National JTPA Performance Measures in Effect in Years 1987-1989

Performance Measure		Definition
<i>Adult Performance Measures</i>		
Employment Rate at Termination	ERT	Fraction of trainees employed at termination
Welfare Employment Rate at Termination	WERT	Fraction of trainees receiving welfare at date of application who were employed at termination
Average Wage at Termination	AWT	Average wage at termination for trainees who were employed at termination
Cost per Employment	CE	Training center's year's expenditures on adults divided by the number of adults employed at termination
Employment Rate at Follow-up	ERF	Fraction of trainees who were employed at 13 weeks after termination
Welfare Employment Rate at Follow-up	WERF	Fraction of trainees receiving welfare at date of application who were employed at 13 weeks after termination
Average Weekly Earnings at Follow-up	AWEF	Average weekly wage of trainees who were employed 13 weeks after termination
Average Weeks Worked by Follow-up	AWWF	Average number of weeks worked by trainees in 13 weeks following termination
<i>Youth Performance Measures</i>		
Youth Employment Rate at Termination	YERT	Fraction of youth trainees employed at termination
Youth Employability Enhancement Rate	YEEN	Fraction of youth trainees who obtained employment competencies (see note 3 below)
Youth Positive Termination Rate	YPTR	Fraction of youth trainees who were "positively terminated" (see note 3 below)
Youth Cost per Employment	YCE	Training center's year's expenditures on youths divided by the number of youths positively terminated

Notes:

1. The date of termination is the date the enrollee officially exits training. A trainee is an enrollee after he has officially exited training.
2. All measures are calculated over the year's *trainee* population. Therefore, the average follow-up weekly earnings for 1987 was calculated using earnings at follow-up for the trainees who terminated in 1987, even if their follow-up period extended into 1988. Likewise, persons who terminated in 1986 were not included in the 1987 measure, even if their follow-up period extended into 1987.
3. A positive termination is entering un-subsidized employment, attaining youth employment "competencies" (through course-work, training and/or tests in work maturity, basic education, or job-specific skills), entering non-JTPA training, returning to school full-time, or completing a major level of education.



TABLE 2 (Continued)

Training Center	Year	ERT	AWT	CE	WERT	ERF	AWWF	WERF	AWEF
Omaha, NE	1987	X	X	X	X				
	1988	X	X		X	X	X	X	X
	1989	X	X		X	X	X	X	X
Marion Co., OH	1987	X	X	X	X				
	1988	X	X	X	X				
	1989	X	X	X	X	X		X	X
Oakland, CA	1987	X	X	X	X				
	1988	X	X	X	X	X	X		
	1989	X	X	X	X	X	X		
Providence, RI	1987	X	X	X	X				
	1988	X	X	X	X				
	1989	X	X	X	X				
Springfield, MO	1987	X	X	X	X				
	1988	X	X	X	X	X			
	1878	X	X			X	X	X	X

## Notes:

1. Source: state incentive policies.
2. See Table 1 for definitions. All measures reported were DOL-defined measures except: ERF in 1977 in New Jersey, Texas, and Georgia, and AWEF in 1987 in Georgia, which were state-defined measures. These state-defined measures were similar to the DOL-defined measures so are included here.
3. Youth measures are omitted for space considerations. In 1987, all states included YERT, YCE, and YPTR (see Table 1 for definitions). In later years, states dropped YCE and added YEEN.

TABLE 3  
State-Defined Performance Measures for States of the National JTPA Study, 1987-1989

State	Enrollment <sup>1</sup>			Expenditure <sup>2</sup>		System Building <sup>3</sup>		Other <sup>4</sup>	
	1987	1988	1989	1987	1988	1988	1989	1987	1989
CA		√	√						
CO									
FL									
GA						√	√		√
IA	√	√	√			√			
IL									
IN	√	√	√						
MN	√	√	√						
MO	√	√	√						
MS	√	√	√	√	√				
MT									
NE	√	√	√	√	√			√	
NJ	√	√	√						
OH	√	√	√			√	√		
IL									
TX			√						

Notes:

1. Enrollment-based awards were based on the proportions of certain groups served by individual training centers. Most groups were defined in terms of a perceived impediment to employment. They included welfare recipients, teenage parents, and displaced-homemakers.
2. Expenditure-based awards required that the training center spend a certain portion of its budget serving particular groups of individuals (such as youth).
3. System-building awards rewarded training centers for attempting to coordinate job training services with services provided by other state agencies serving the poor.
4. Two states had other standards which do not easily fit into the other three categories. Nebraska rewarded training centers for high ratios of private sector to government sector on-the-job training placements. Georgia gave training centers extra credit for placing in jobs JTPA enrollees who were also recipients of welfare services in other state agencies.

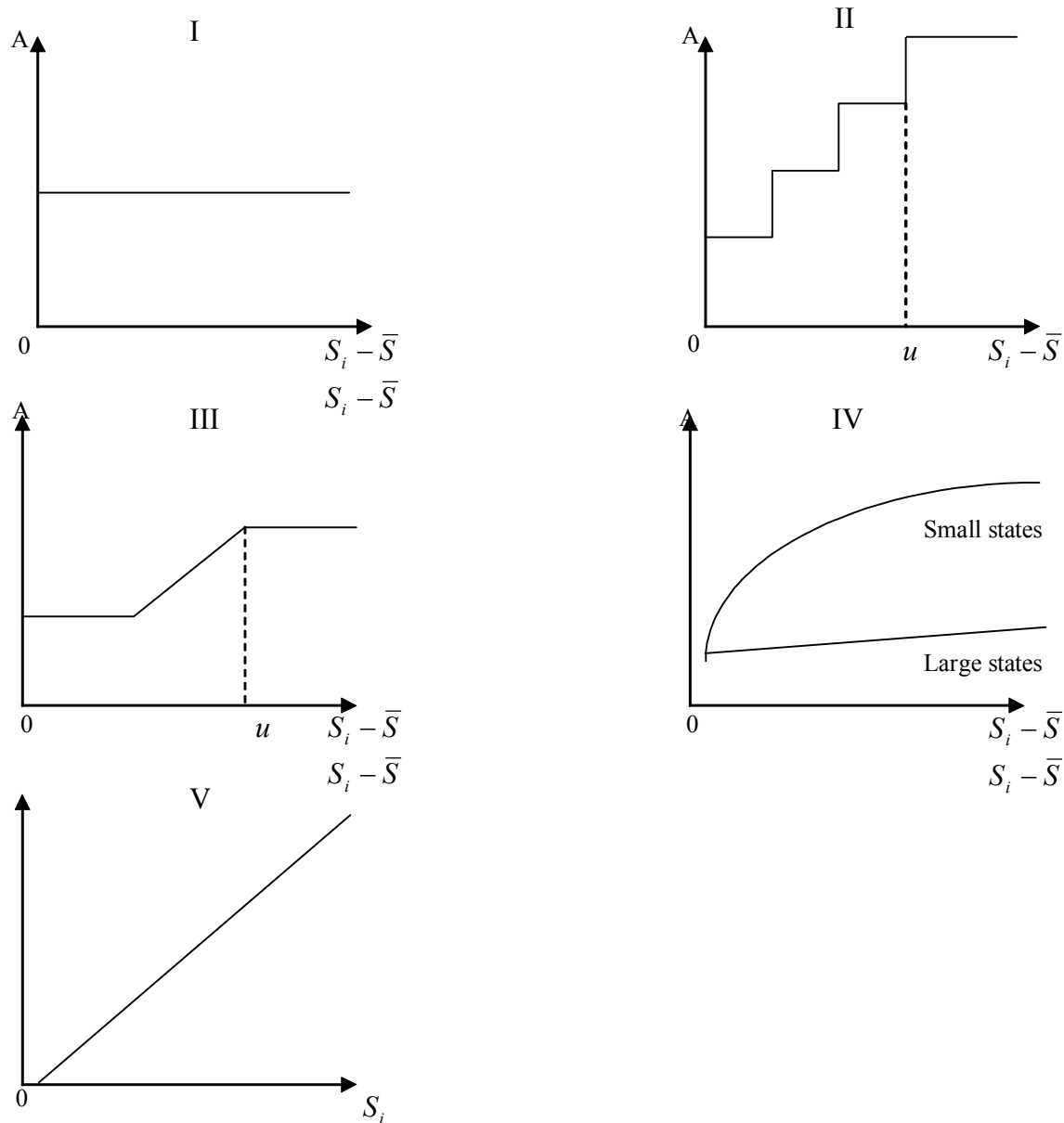
TABLE 4  
Award Qualifying Criteria by State in Program Year

State	1987	1988-1989
CA	Exceed at least 5 standards	Exceed at least 7 standards
CO	Exceed at least 1 standard	Exceed at least 5 standards
FL	$\sum_{k=1}^K \frac{S_k - \hat{S}_k}{\hat{S}_k} > 0$	$\sum_{k=1}^K \frac{S_k - \hat{S}_k}{\hat{S}_k} > 0$
GA	$\sum_{k=1}^K \frac{S_k - \hat{S}_k}{\hat{S}_k} > 0$	State used 8 of 10 measures to determine eligibility, which it divided into 2 groups. <i>Primary group</i> (2 measures): Exceed both standards <i>Secondary group</i> (6 measures): Exceed 4 standards
IA	Exceed at least 1 standard	Exceed at least 1 standard
IL	State divided 7 measures into 3 groups. <i>Primary group</i> (2 measures): Exceed both standards <i>Secondary group</i> (2 measures): Exceed at least 1 standard <i>Tertiary group</i> (3 measures): Exceed at least 1 standard	Exceed at least 5 standards
IN	Exceed all standards	Exceed all standards
MN	Exceed at least 5 standards	Exceed at least 6 standards
MO	$\sum_{k=1}^K w_k \frac{S_k - \hat{S}_k}{\hat{S}_k} > 0$	$\sum_{k=1}^K \frac{S_k - \hat{S}_k}{\hat{S}_k} > 0$
MS	Exceed at least 1 standard	Exceed at least 1 standard
MT	Exceed all standards	Exceed all standards
NE	Exceed at least 1 federal standard, exceed 3 of 4 state standards	Exceed at least 1 federal standard, exceed all state standards
NJ	Exceed all standards	Exceed all standards
OH	Exceed all standards	Exceed 7 standards
RI	$\sum_{k=1}^K \frac{S_k - \hat{S}_k}{\hat{S}_k} > 0$	$\sum_{k=1}^K \frac{S_k - \hat{S}_k}{\hat{S}_k} > 0$
TX	Exceed at least 6 standards	Exceed at least 6 standards

Notes:

1. The standards referred to above are based upon DOL regression-adjusted departure points. In addition, many states adjusted these standards further using DOL-sanctioned options. Thus the standards described above are not uniformly constructed.
2.  $S_k$  is a training center's outcome for performance measure  $k$ .  $\hat{S}_k$  is the corresponding standard.  $w_k$  is the weight applied to performance measure  $k$ .
3. In general, the set of performance measures varies across states. Thus,  $K$ , the number of performance measures activated, is understood to vary across states.

Figure 1  
Award Functions for States Participating in the National JTPA Study, 1987-1989



Notes:  $S$  is the performance outcome,  $\bar{S}$  is the corresponding standard, and  $A$  is award amount. Each state uses one or more of these award types. Table 5 shows award-types by state.

TABLE 5  
Award Functions for States Participation in the National JTPA Study, 1987 - 1989

Function (See Figure 1)	State
I	California, Georgia, Illinois, Indiana, Montana, New Jersey
II	Illinois, Minnesota, Mississippi, Ohio, Texas
III	Colorado, California
IV	Iowa, Missouri, Nebraska, Rhode Island, Florida
V	New Jersey (1988 - 1999)