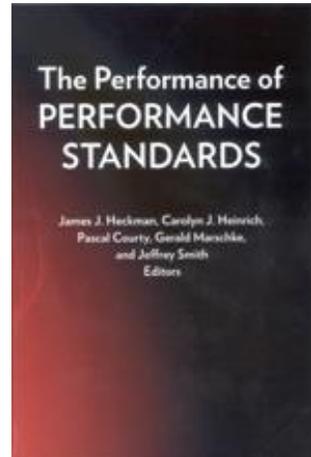

Upjohn Institute Press

The JTPA Incentive System: Implementing Performance Measurement and Funding

Pascal Courty
University of Victoria

Gerald Marschke
University at Albany, SUNY



Chapter 4 (pp. 65-94) in:

The Performance of Performance Standards

James J. Heckman, Carolyn J. Heinrich, Pascal Courty, Gerald Marschke,
and Jeffrey Smith, eds.

Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2011

DOI: 10.17848/9780880993982.ch4

Copyright ©2011. W.E. Upjohn Institute for Employment Research. All rights reserved.

The Performance of Performance Standards

James J. Heckman
Carolyn J. Heinrich
Pascal Courty
Gerald Marschke
Jeffrey Smith
Editors

2011

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

The performance of performance standards / James J. Heckman . . . [et al.], editors.
p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-88099-292-3 (pbk. : alk. paper)

ISBN-10: 0-88099-292-1 (pbk. : alk. paper)

ISBN-13: 978-0-88099-294-7 (hardcover : alk. paper)

ISBN-10: 0-88099-294-8 (hardcover : alk. paper)

1. Government productivity. 2. Performance standards. 3. Civil service—Personnel management. I. Heckman, James J. (James Joseph)

JF1525.P67P476 2011

352.6'7—dc22

2011007877

© 2011

W.E. Upjohn Institute for Employment Research

300 S. Westnedge Avenue

Kalamazoo, Michigan 49007-4686

The facts presented in this study and the observations and viewpoints expressed are the sole responsibility of the author. They do not necessarily represent positions of the W.E. Upjohn Institute for Employment Research.

Cover design by Alcorn Publication Design.

Index prepared by Diane Worden.

Printed in the United States of America.

Printed on recycled paper.

4

The JTPA Incentive System

Implementing Performance Measurement and Funding

Pascal Courty
Gerald Marschke

This chapter outlines the specifics of JTPA's performance incentives, which provide necessary background information for subsequent chapters in this monograph. It also speaks generally to the challenges that must be met in formulating performance measures and incentives anywhere in government.¹ In particular, we argue that the decisions about what should be measured, and how, when, and by whom it should be measured, make a critical difference for the success of incentive-backed performance measurement.

The JTPA organization was 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. In JTPA, states indeed used their discretion to produce a wide variety of performance measurement and incentive structures. The federal government retained control over some important aspects of the incentive system. The discretion left to the states, however, was important for determining the character of performance measurement and the incentives.

By providing an analytical description of performance measurement in JTPA, this chapter complements the institutional literature on the JTPA bureaucracy (Barnow 1992; Svorny 1996). Although this literature provides a good 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, this chapter lays the foundation for under-

standing how the JTPA incentive system determined bureau-crat behavior and program outcomes. Previous studies of the impact of JTPA's performance incentive scheme on bureaucratic behavior (Anderson, Burkhauser, and Raymond 1993; Courty and Marschke 2004; Heckman, Smith, and Taber 1996; and Marschke 2002) have used only the federal guidelines, which provide an incomplete and possibly misleading representation of the true incentive systems.² Along with others (see, e.g., Wholey 1999, p. 305), we believe that such case studies of operational performance measurement systems are important inputs into developing theories to understand and recommendations to improve performance measurement in the public sector.

We limit our description to the state incentive and performance measurement policies for the years 1987–1989 for a sample of 16 states (identified in Table 4.1).³ In describing JTPA's performance-based incentive system, we address three questions. First, what was the nature of the training center's incentive? Or, how much was at stake? Second, which dimension of performance did JTPA reward? That is, what mattered? According to the act, Congress intended the performance incentives to measure the training centers' success in developing participants' labor-market specific human capital (U.S. Congress 1982, Section 106[a]). Because direct measures of human-capital value-added are unavailable, the program's federal overseers have resorted to proxies of value-added. At the heart of the JTPA incentive system is a set of performance measures based on the labor market outcomes of enrollees at or shortly after training. We describe these performance measures.

Third, 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 factors that were specific to each training center and outside the centers' control. The performance standards were intended to establish reasonable counterfactual 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.

The chapter is organized as follows. The first section investigates the bureaucrats' motivations for seeking awards. The second section explains the JTPA performance measures and also how performance outcomes across the population of training centers were adjusted to level the playing field. The third section describes the incentive award, and the last section concludes and summarizes our main findings.

AWARDS AND BUREAUCRATIC PREFERENCES

The JTPA incentive system takes the form of increases in training centers' training budgets. We start by describing the sizes of these increases. By focusing solely on the absolute award amounts, however, one risks overlooking potentially important influences of performance awards on bureaucratic behavior and program outcomes. These non-financial reasons are reviewed next.

Award Size

The act required that states allocate about 7 percent of their total JTPA training budget to their incentive programs.⁴ States then decided how to allocate this sum among the three categories of expenditures: 1) awards for successful training centers, 2) administration of the incentive programs, and 3) "technical assistance" for unsuccessful training centers. JTPA mandated that the state set aside funds for technical assistance, but left the amount set aside up to the state. Technical assistance provided resources to improve managerial performance to training centers that failed to meet performance standards.

While the act intended that the training centers use the awards primarily for training, awards could also be used for staff bonuses and payroll increases. A training center was required to spend at least 70 percent of the award on training activities, leaving 30 percent, at most, for staff compensation. By comparison, training centers were permitted

Table 4.1 State Funding of the JTPA Incentive Program States of the National JTPA Study, 1987–1989

State	Program year 1987		Program year 1988		Program year 1989	
	Award fund as percentage of 6 % funds	Percentage of award funds allocated to federal performance standards	Award fund as percentage of 6 % funds	Percentage of award funds allocated to federal performance standards	Award fund as percentage of 6 % funds	Percentage of award funds allocated to federal performance standards
CA	85	100	85	88	85	88
CO	95	100	85	100	85	100
FL	—	—	—	—	—	—
GA	75	100	75	90	75	90
IA	75	99	75	99	75	99
IL	75	100	75	100	75	100
IN	93	40	92	60	78	60
MN	85	50	85	50	85	50
MO ^a	100	100	100	100	100	100
MS	100	70	98	70	98	70
MT ^a	66	100	66	100	66	100
NE	85	60	85	60	85	60
NJ	80	33	80	33	80	33
OH	80	100	80	60	70	60
RI	75	100	75	100	75	100
TX	76	90	77	90	81	80

NOTE: JTPA allowed each state to use up to 6 percent of its JTPA appropriation for direct cash payments to job training centers for performance on federal and state performance measures. The first column for each year lists the fraction of the 6 percent that was set aside for the award fund. This fraction represents the maximum amount that would be rewarded to training centers for performance relative to federal and state standards. The second column for each year represents the fraction of the award fund that is set aside for federal performance standards alone. — = data not available.

^aThe 1987 values for Missouri and Montana in the first column are taken from the NCEP-SRI survey, which interviewed training center personnel in 1986 concerning the program year 1986. They are not calculated from training center policy documents.

to spend no more than 15 percent out of budgetary funds on payroll. The higher cap for payroll expenses alone may have motivated training centers to pursue awards.

Table 4.1 shows the share of the incentive fund set aside for the award for a sample of 16 states in the years 1987–1989. We collected data on these 16 states because they contained the 16 training agencies that participated in the late 1980s USDOL-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. The analyses contained in Chapters 5, 6, and 9 are at least in part based on the NJS data. The second, fourth, and sixth columns show the award size as a percentage of the 7 percent incentive fund for 1987, 1988, and 1989, respectively. Although states varied in the portion set aside for awards, as Table 4.1 shows, most states made available a majority of the incentive fund to training centers as potential awards. Montana set aside the smallest share—66 percent of the incentive fund—while Missouri and Mississippi (in 1987) set aside the entire fund to be paid out as an incentive award. While the portion set aside varied significantly across states, it was stable over the 1987–1989 period. For example, between 1987 and 1989, Illinois devoted 75 percent of its 7 percent fund to the award fund. It reserved the other 25 percent for technical assistance. Only in Indiana did a significant change take place. Between 1987 and 1989, Indiana lowered the portion set aside for the award from 93 percent of the incentive allocation to 78 percent. The other states devoted a constant (or nearly constant) share of the incentive funds to awards.

In some states, not all of the funds that the state made available for awards were paid out. Not all funds were awarded because many training centers did not perform well enough to meet the states' award eligibility requirements. Nevertheless, as Table 4.2 shows, the amounts paid out were substantial. Table 4.2 shows the actual award amounts disbursed as a fraction of the training center's budget for a sample of 448 training centers in program year 1987. The average disbursement was equivalent to 7 percent of the training center's budget.⁵ The highest disbursements were equivalent to 60 percent of the training center's budget. Rewards this large, however, were not possible in all states.

Table 4.2 Size of Incentive Award

	Allocation (\$)	Incentive award (\$)	Fraction of training center's budget
Mean	2,326,191.42	119,663.79	0.07
(Std. dev.)	(3,043,936.68)	(145,751.12)	(0.07)
First quartile	1,003,308.50	33,000.00	0.03
Median	1,627,151.50	93,550.00	0.06
Third quartile	2,398,462.00	160,534.00	0.09
Maximum	29,408,455.00	1,407,853.00	0.57
Number of obs.	448	385	385

NOTE: Data are from the National Commission for Employment Policy and SRI, Inc. (see their description in Dickinson et al. [1988]).

We expect that if incentives matter, the intensity of the behavioral responses to the incentives should depend on the size of the award. Tables 4.1 and 4.2 show that the award amounts available varied significantly by state. Everything else equal, we might expect that JTPA incentives produced greater responses among training centers in Missouri than Montana where the award disbursement was the greatest (i.e., 100 percent of the award fund). In addition, the kind of behavioral response we observe would likely have depended upon how states divided award money between state and federal measures of performance. The division matters because state and federal performance measures stressed different aspects of training center output.

Why Should Training Centers Care about Awards?

As a rough approximation, the awards might have increased salaries by as much as 15 percent.⁶ This figure corresponds to the purely financial part of the incentive. This by itself represents a substantial increase even compared with private sector bonuses. There is, however, little evidence that training center administrators paid out such a large share of the award as salary bonuses.⁷ Why, then, should bureaucrats care about the awards?

There are at least three reasons why a training center may wish to increase its budget. First, Niskanen (1971) and others argue that everything the bureaucratic manager desires (salary, staff, power, professional reputation, and perquisites) derives from the bureau's budget.

Because of their training and the orthodoxy of the social work profession, welfare bureaucrats—like the ones in JTPA—may behave as selfless advocates for their clients (Lipsky 1980). Nevertheless, even as client advocates, JTPA bureaucrats would desire extra award funds to expand their clients' training resources.

Second, performance levels and the awards they represent might have been used by local elected officials for political gain. The act gave local elected officials—often these are city mayors—some authority over the operation of training centers. The administrative headquarters of training centers were frequently situated in these elected officials' offices. These elected officials often touted performance awards as measures of their administration's success in the local fight against poverty.

Third, bureaucrats might have sought performance awards for the professional recognition they convey. Bureaucrats faced no other objectives—and the public and bureaucratic superiors have no other evaluation criteria—as precisely defined, quantifiable, and available as these performance measures. Tirole (1994, p. 7) argues that government bureaucrats might be “concerned by the effect of their current performance not so much on their monetary reward, but rather on their reputation or image in view of future promotions, job prospects in the private and public sectors.”⁸ Moreover, by performing well compared to the standards, bureaucrats could protect themselves against attack from outside critics and political enemies.

As a final point, note that because awards were based on group performance, individual bureaucrats had an incentive to free-ride on the effort of colleagues. Free-riding may have muted the influence of incentives at the level of caseworkers. Ultimately, the issue of the influence of budget-based awards and the significance of the free-riding problem must be resolved empirically.

PERFORMANCE MEASURES

The act directed the USDOL to formulate measures of performance that captured the gains produced in the employment and earnings of participants and the reductions in their reliance on welfare programs (Chapter 2). Reliable measures of earnings and employment impacts of

job training, however, were prohibitively costly to obtain, and instead, the USDOL issued measures that assessed the effectiveness of job training through the labor market outcomes of enrollees at training end.

JTPA required states to use the USDOL's performance measures in constructing their incentive system. Nevertheless, the USDOL 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 (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.

The following section discusses the federal and state-designed measures and how states 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 replaced or redefined those performance measures whose effects were possibly counterproductive. We also discuss these modifications.

Federal Performance Measures

Table 4.3 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, the average wage at termination, the cost per employment, the employment rate at 90 days after termination—i.e., at follow-up, the average weeks worked at follow-up, and the 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 GED certification (see Table 4.3, especially the last part of the note).

Table 4.3 National JTPA Performance Measures in Effect in Years 1987–1989

Performance measure	Definition
Adult performance measures	
Employment rate at termination	Fraction of trainees employed at termination
Welfare employment rate at termination	Fraction of trainees receiving welfare at date of application who were employed at termination
Average wage at termination	Average wage at termination for trainees who were employed at termination
Cost per employment	Training center's year's expenditures on adults divided by the number of adults employed at termination
Employment rate at follow-up	Fraction of trainees who were employed at 13 weeks after termination
Welfare employment rate at follow-up	Fraction of trainees receiving welfare at date of application who were employed at 13 weeks after termination
Average weekly earnings at follow-up	Average weekly wage of trainees who were employed 13 weeks after termination
Average weeks worked by follow-up	Average number of weeks worked by trainees in 13 weeks following termination
Youth performance measures	
Youth employment rate at termination	Fraction of youth trainees employed at termination
Youth employability enhancement rate	Fraction of youth trainees who obtained employment competencies (see note below)
Youth positive termination rate	Fraction of youth trainees who were "positively terminated" (see note below)
Youth cost per employment	Training center's year's expenditures on youths divided by the number of youths positively terminated

NOTE: The date of termination is the date the enrollee officially exits training. A *terminee* is an enrollee after he has officially exited training. All measures are calculated over the year's *terminee* population. Therefore, the average follow-up weekly earnings for 1987 was calculated using earnings at follow-up for the *terminees* 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. A positive termination is entering unsubsidized employment, attaining youth employment "competencies" (through coursework, 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.

All federal performance measures for adult participants had three characteristics in common. All 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 three 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.⁹ Awards were thus independent of the number of persons who obtained high outcomes.

The use of performance measures varied across states. In 1987, the USDOL required states to base incentives on seven termination-based measures: the adult employment rate at termination, adult welfare employment rate at termination, adult cost per employment, average wage at termination, youth employment rate at termination, youth positive termination rate, and youth cost per employment. In 1988–1989, the USDOL extended the list of measures to the 12 described in Table 4.3 and required states to choose any 8 or more of these 12 measures. The years 1988 and 1989 marked the beginning of the phaseout of termination-based performance measures and the cost measure.¹⁰ After 1992, the USDOL prohibited the states from using any cost measures of performance. Moreover, after 1992, all performance measures based on measures of output became follow-up measures.¹¹

Performance measures based on labor market outcomes may have influenced training center behavior in several ways. First, because they measure aspects of an enrollee's employment state and not the impact of job training, they may have led 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 (see Chapters 6 and 9).

Second, the performance measures may have encouraged training centers to offer "quick fixes," that is, employment-oriented job search or on-the-job training services. These services, rather than more

intensive kinds that result in greater skill development, would more frequently lead to 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 because the employment and wage measures focused on labor market success at a point in time rather than over a period of time. Cream skimming is an example of dysfunctional behavior that results when the performance measures are not well-aligned with the goals of the organization. If quick fixes generate smaller welfare gains per dollar spent than more intensive services they also constitute dysfunctional behavior (see Bloom et al. [1997] for evidence on the relative effectiveness of different training types in JTPA). See Blau (1955) for an early discussion of unintended responses to performance measures.

Third, because these measures were based on averages instead of aggregate outcomes, training centers had no incentive to spend their entire budget. Actually, 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.

Fourth, the role of the cost measure was ambiguous. The cost measure was defined as the total expenditure divided by the number of persons employed at training end. Holding spending constant, the cost measure becomes an incentive to produce as many employed trainees as possible. Bureaucrats could produce greater numbers of employed trainees either by increasing their employment rates at termination or by enrolling more applicants (holding the employment rate constant). Viewed in this way the cost measure would have countered the incentive to serve small populations. The elimination of the cost measures in the last years of JTPA would have added an additional incentive to reduce the number of enrollees served. Training centers would enroll less than the efficient number of enrollees because small enrollee populations increased the per capita spending, and greater per capita spending increased the per capita performance outcomes (Barnow [1992] makes this point).

Federal Performance Standards

The impact of the performance measures on bureaucratic behavior depended critically on the threshold the bureaucrat had to meet. For example, a training center did not receive an award for the wage measure 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 was set too high so that no amount of effort would push the wage outcome over the standard. Another reason why the performance standards were crucial was because the performance awards did not depend on absolute performance but on excess performance, that is, on the difference between the performance outcome and the performance standard.

This subsection describes the heights of the standards and how they were tailored to the different environments faced by the training centers. The USDOL adjusted the performance standards to the local conditions faced by training centers in an attempt to level the playing field.

For each performance measure, the numerical standard started with the national “departure point.” The USDOL 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 performance standard on average. For the cost measure, good performances were low performances. Successful training centers had to produce an outcome below the standard. The USDOL set the departure point for the cost measure at the 90th percentile. For the wage measure, the department set the departure point at the 50th percentile.¹²

Training centers faced different costs of meeting these departure points. Costs varied because labor markets, training costs, and the characteristics of the eligible populations varied. Imposing uniform standards would have favored low-cost training centers by increasing their resources relative to high-cost training centers. Only in the case that low-cost 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 USDOL established an adjustment model that took into account features of the training center’s environment that may be correlated with costs. 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.

Although the USDOL allowed states some flexibility in developing standards, most states used the department's adjustment methodology.¹³ All states participating in the NJS used the USDOL's adjustment model during the NJS years. We describe this method here.

USDOL adjustment model

Consider an arbitrary performance measure and let S_l be the outcome produced by training center l . The USDOL adjustment scheme posited that the following function generates performance outcome S_l .

$$(4.1) \quad S_l = \alpha + \beta_1(x_{1l} - \bar{x}_1) + \beta_2(x_{2l} - \bar{x}_2) + \dots + \beta_M(x_{Ml} - \bar{x}_M) + \varepsilon_l,$$

where $x_{1l}, x_{2l}, \dots, x_{Ml}$ are training center l 's realizations for the M factors chosen, $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M$, are the average realizations of these factors over all JTPA training centers, and ε_l is a site-specific error term. Biannually the USDOL estimated the coefficients β with the most recent two years of training center-level data using ordinary least squares. β_m expresses the impact of an increase in the factor x_{ml} , on the outcome S_l , holding other factors constant. The USDOL chose a different set of factors for each performance measure. It chose those economic factors and demographic variables based upon their availability and whether the factors were statistically correlated with the performance outcomes. In addition, political considerations may have played a role.¹⁴

Table 4.4 presents an example of a JTPA worksheet for adjusting the adult employment rate at termination in 1987. The first six adjustment factors in the table are enrollment population characteristics (the percentage of the participant population that is female, black, Hispanic, Asian, handicapped, and welfare recipients). The last two adjustment factors are measures of the local economy (unemployment rate and population density). Column B presents factor values for a hypothetical training center. Columns C and E present the actual national factor averages and the weights from the USDOL adjustment model for 1987. These weights are the estimated effects of each characteristic on the performance outcome adult employment rate at termination (estimated

Table 4.4 U.S. Department of Labor's Performance Standard Adjustment Model Performance Standard: Adult Employment Rate at Termination

A. Local factors	B. Training center factor values	C. National averages	D. Difference (B - C)	E. Weights	F. Effect of local factors
% Female	49.9	52.8	-2.9	-0.020	0.058
% Black	41.2	23.8	17.4	-0.081	-1.41
% Hispanic	30.1	7.9	22.2	-0.009	-0.20
% Asian	2.1	2.4	-0.3	-0.022	0.01
% Handicapped	9.5	9.1	0.4	-0.093	-0.04
% Welfare recipient	35.0	29.8	5.2	-0.276	-1.44
Unemployment rate	8.8	8.0	0.8	-0.623	-0.50
Population density	0.21	0.6	-0.39	0.771	-0.3
G. Total effect of local factors on performance expectations					-5.77
H. National departure point					62.4
I. Model-adjusted performance level (G + H)					56.6

NOTE: Local factors listed in column A are determined by the USDOL. Percentages are of year's participant population. Values for columns C, E, and H are given by the USDOL. Values for column B are for a hypothetical training center.

β 's from Equation 4.1). The training center's realization of each of the factors is compared to the national average and the difference is multiplied by a weight. For example, suppose the hypothetical training center served 1,000 persons during 1987, of which 499 were female. Thus, its percentage female factor was 49.9 percent. To obtain the adjustment to the standard for the female participation factor, one multiplies the difference between the training center's factor value and the national average ($49.9 - 52.8 = -2.9$) by the adjustment weight (-0.020). The adjustment weight reflects how the enrollment of women historically affected the employment rate outcome.

For the other factors, weighted differences were calculated similarly. The total effect of local factors on performance expectations (column F)—the sum of the weighted differences—was added to the national departure point. The departure point (the 25th percentile value) for the measure was 62.4. The final performance standard (56.6) is the sum of the departure point (62.4) and adjustment factor (-5.8). The state used the final standard to establish whether the training center had met its adult employment rate target.

The USDOL intended that the bar be set to a height appropriate to the training center's circumstances. Thus it included measures of the local unemployment rate and of local population density to capture aspects of the local labor market in which the training center operated. For example, as one can see from Table 4.4, the weight on the local unemployment rate measure was negative: a one-point increase in the local unemployment rate lowered the standard by about two-thirds of a point. Because training centers were small relative to the local labor market, the unemployment rate is an example of an influence on the performance outcome which was likely to be beyond the training center's control.

As Table 4.4 also makes clear, an important class of characteristics for which the USDOL 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.

Table 4.4 also reveals that the USDOL adjustment model did not take into account training services as a relevant control variable, although these data were available to the department. 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; both choices have consequences for the effectiveness and efficiency of training. Neither did the adjustment model directly control for the training services available in the training centers' local area. This is more surprising; by not controlling for the availability and costs of training facing training centers, the incentive system implicitly favored those training centers located in markets where there was a competitive and efficient training industry.¹⁵

State Performance Measures and Standards

States were permitted to develop their own performance measures. 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 nationwide trend.

In Table 4.1, the second column under each year shows the percentage of the total award set aside allocated to federal measures, as opposed to state measures. 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 82/18. The split ranged from a low of 50/50 to a full allocation of the money to the federal award.

We broke down the state measures into four categories. 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 bureaucrats 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. In 1988, for example, 9 out of 16 states set up standards that com-

compensated training centers for the number of or the rate at which persons in target groups were enrolled.¹⁶ 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.

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. Training centers might have left portions of their budgets unspent because, as we noted above, the kinds of applicants who would produce high-performance outcomes may have been scarce. Rather than enrolling less able enrollees who lower per capita scores, training centers might have enrolled fewer enrollees than the maximum their budgets would have allowed.

Three of the 16 states sought to encourage JTPA training centers to coordinate their activities with other state agencies that helped the poor.¹⁷ 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 USDOL 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

This section describes how states used performance outcomes and standards to reward training centers. In particular, we discuss the states' eligibility rules that determine which training centers received awards and the award functions themselves. The award functions translate the performance of eligible training centers to award amounts.

The states were entirely free to design the eligibility rules and the performance awards as they saw fit, which led to great variation in both across states. In fact, no two award functions (or eligibility rules, for that matter) were identical. 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

Here we discuss the qualifying criteria for the 16 NJS states in years 1987–1989. Some states (such as 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 (such as Iowa) simply rewarded training centers for each performance standard exceeded.

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.¹⁸ 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 may 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 special-

izing 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 several important dimensions. The three most important dimensions, discussed below, address three broad questions: Which performance measures mattered? How competitive was the incentive system? Did excess performance cease to matter after some point?

Performance measure weighting. By exercising their discretion over which measures they included in their awards (see Table 4.2) and in their construction of the eligibility criteria, 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 used equally, some states weighted performance measures differently to emphasize some measures over others.

Consider, for example, New Jersey in 1987 and 1988. While a training center there had to meet its cost standards to qualify for an award, its cost outcomes did not figure into the award's calculation. Moreover, the award calculation up-weighted the follow-up-based measures compared to the termination-based measures. In de-emphasizing the cost- and termination-based measures, New Jersey intended to encourage training centers to provide more intensive training and enroll more difficult-to-train enrollees (see Note 12).

Competition among training centers. In many states (such as Texas), a training center's award depended only on its own performance. In Illinois's 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 successful ones.

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. 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 between 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 that affect all training centers uniformly. Thus, relative performance evaluation should be most effective in states where training centers are homogeneous (e.g., operate in similar environments) and if the performance standards do a poor job of controlling for factors that are outside the training centers' control.¹⁹

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. However, they may depend on the level of performance. 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 only contingent upon achieving standards. Many states (such as 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. For example, in Illinois in 1987, a training center's award increased with its performance, until its performance exceeded the 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.

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 increase their performance outcomes without actually increasing the effectiveness of their training. Deadline effects have been described in the behavior of military recruiters whose bonuses depend on recruitment quotas (Asch 1990), salespersons whose commissions depend on achieving sales targets (Oyer 1995), and CEOs whose bonuses depend on performance targets linked to measures of corporate earnings (Healy 1985). See Courty and Marschke (1996, 1997) for evidence of deadline effects in JTPA, and Courty and Marschke (2004) for an attempted estimation of the efficiency costs of these effects.

CONCLUSION

This chapter details a description of the incentives in place under JTPA. The incentive awards amount to budgetary increases that training center bureaucrats valued for professional, personal, and political reasons. The potential size of the bonus award varied by state and program year. In program year 1986, for example, the total bonus was about 7 percent of the training budget. These awards, however, could have been a substantial added source of training funds. Depending on the state and the year, this amounted to as much as 60 percent of a training center's yearly allocation of funds.

The heart of the JTPA incentive system was the set of performance measures and their standards. Job training programs illustrate 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. Moreover, 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. In later years, some states adopted expenditure-based performance measures, possibly to counteract the incentives created by federal measures to leave some of their budgets unspent.

Training centers received awards when the outcomes of performance measures exceeded numerical standards. The USDOL attempted to adjust performance standards to reflect the environment in which the training center operated so that training centers in healthy and in depressed economies had to exert complementary levels of effort to achieve their standards. The USDOL's adjustment scheme offers a real-world example of strategies for adjusting performance measures that have been proposed in the literature (Stiefel, Rubenstein, and Schwartz 1999). Performance standards were also adjusted for the characteristics of persons enrolled, so as 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 clubs and on-the-job training—and these services may not have produced either stable employment relationships or significant improvements in the long-term employability of enrollees. This study has revealed numerous ways in which performance measures might have been misaligned with the agency's goals. With the exception of some cream skimming studies and a study of deadline effects, these distortionary effects have not been investigated. We believe that further research on the behavioral responses to JTPA's incentives would lead to much useful information for developing and refining performance measures for many kinds of public sector organizations.

The federal government left the formulation of many of the details of the award to the states. Consequently, the form of the award varied. 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, suggesting that the magnitude of responses to performance incentives depended on the state. States also differed in the degree of competition among training centers and in the interdependence of awards. Head-to-head competition among training centers in some states may have heightened the impact of the incentive system on behavior. States also differed in the extent to which they compensated exceptional performance. In some states, training centers received no more award money for meeting standards than exceeding them. In other states, training centers

received additional money for higher performance. States also designed and implemented their own performance measures that promoted different goals than the goals implicit in the federal measures. Thus, the objectives transmitted to the training center depended on the state in which it operated.

Notes

This chapter introduces features of the JTPA incentive system that will be used extensively in the remainder of this monograph. Some of these features have been discussed and analyzed in detail elsewhere (Courty and Marschke 1997, 2002, 2003, and 2004), and parts of this chapter borrow from these sources.

1. Discussions of general criteria for choosing performance measures and constructing performance measurement systems, for example, are found in Hatry (1980), Hurst (1980), Usilaner and Soniat (1980), Wholey (1999), and GAO (1996).
2. Cragg (1997) and Marschke (2002) are exceptions.
3. We chose to limit the number of states and years for which we collected incentive policies because of the difficulty and expense of obtaining these records. We collected data on these 16 states because they contained the training agencies that participated in the late 1980s USDOL-commissioned National JTPA Study, described in Chapter 2.
4. The JTPA funds were allocated in three subfunds: 1) 78 percent were set aside for training services, 2) 6 percent were set aside for the incentive system, and 3) 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 fund was spent on the administration of the incentive funds.
5. These figures are based on the data set of SRI, International and Berkeley Planning Associates. See Dickinson et al. (1988) for a description of these data.
6. Salary payroll represented at most 15 percent of training budget that was itself only 78 percent of total JTPA training funds (see Note 7). The award fund was 6 percent of JTPA funds and at most 30 percent could be distributed as salary. The award salary bonus as a fraction of total salary was at most $(0.06 \times 0.3)/(0.15 \times 0.78) = 0.15$.
7. In a survey of 30 training centers conducted by Dickinson et al. (1988), only three administrators indicated that this was their practice. In a 1994 telephone survey administered to 11 of the 16 training centers of the National JTPA Study, all training center administrators that we spoke to denied that they ever distributed a portion of the award for salary bonuses.
8. Thompson (2000) describes the merit pay system at the Social Security Administration, ushered in with the enactment of the Civil Service Reform Act of 1978. Under this system one half of managers' annual pay increases were determined by

their performance relative to a set of output indicators. While originally set up as a basis for rewarding pay increases, these indicators became the basis for promotion decisions as well. According to one former manager, they became “the basis for your career . . . if you do well on those four measures. You can write your own ticket” (p. 270).

9. Note that persons who entered the program already employed, and then terminated 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.
10. In the first years of JTPA (through 1987), the USDOL required states to use a cost measure. Nevertheless, some policymakers and analysts, alarmed at the short length of training (the average length of training is about five months in JTPA), instigated investigations by the GAO and other interested parties into the link between cost measures and short, low-intensity services. As a consequence of this inquiry, the USDOL encouraged states to phase out the cost measure, “to encourage [training centers] to provide more comprehensive programming and increased services for those individuals who are most in need” (Division of Employment and Training, New Jersey Department of Labor 1990). Moreover, in response to a number of Labor Department investigations, which concluded that training centers were emphasizing “quick fixes” with job-placement-oriented services that had no long-term impact on enrollees’ skills, the department formulated a number of follow-up measures: the measures based on outcomes three months after termination, presented in Table 4.3. The USDOL introduced follow-up measures to “[promote] effective service to participants and [assist] them to achieve long-term economic independence” (Division of Employment and Training, New Jersey Department of Labor 1990).
11. That is, 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 (13 weeks after termination).
12. Actually, the USDOL’s JTPA Technical Assistance Guide, PY1988, reports that the cost standard’s departure point was set “above the 25th percentile. It more closely resembles an estimate of average performance.” We interpret this statement to mean the department set the wage standard at the 50th percentile.
13. 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 USDOL; or 3) propose their own adjustment method to the USDOL. Between the end of the National JTPA Study and the end of JTPA, more states abandoned the USDOL adjustment method for their own (option 3).
14. Barnow (1992) writes that “when estimated coefficients have an unexpected sign, the variables are dropped from the models and regressions are re-estimated” (p. 292). For example, in some regressions, the USDOL dropped an indicator variable for Hispanic enrollees because it apparently showed a positive effect on performance outcomes.

15. For more on the specification of USDOL's regression model, see Barnow (1992) and Trott and Baj (1987). One significant shortcoming they describe is that estimating the Equation with training center-level data, as opposed to enrollee-level data, biased the estimates of the model's coefficients. Using enrollee-level data, Trott and Baj demonstrated that the practice of pooling the data by training center significantly changes some of the estimates of β_m . For a general discussion of the theory of adjusting performance measures, what factors should and should not be used to adjust performance measures, see Stiefel, Rubenstein, and Schwartz (1999).
16. These states were California, Iowa, Indiana, Minnesota, Missouri, Mississippi, Nebraska, New Jersey, and Ohio.
17. These states were Georgia, Iowa, and Ohio.
18. 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.
19. The USDOL's performance standard adjustment methodology—to the extent that it accounted for external factors that affect performance—produced the same effect.

References

- Anderson, Kathryn, Richard Burkhauser, and J. Raymond. 1993. "The Effect of Creaming on Placement Rates under the Job Training Partnership Act." *Industrial and Labor Relations Review* 46(4): 613–624.
- Asch, B.J. 1990. "Do Incentives Matter? The Case of Navy Recruiters." *Industrial and Labor Relations Review* 43(3): S89–106.
- Barnow, Burt. 1992. "The Effect of Performance Standards on State and Local Programs." In *Evaluating Welfare and Training Programs*, Charles F. Manski and Irwin Garfinkel, eds. Cambridge, MA: Harvard University Press, pp. 277–309.
- Blau, P. 1955. *The Dynamics of Bureaucracy: A Study of Interpersonal Relations in Two Government Agencies*. Chicago: University of Chicago Press.
- Bloom, Howard S., Larry L. Orr, Stephen H. Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M. Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study." *Journal of Human Resources* 32(3): 549–576.
- Courty, Pascal, and Gerald Marschke. 1996. "Moral Hazard under Incentive Systems." In *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*, G. Libecap, ed. New York: JAI Press, pp. 157–190.
- . 1997. "Measuring Government Performance: Lessons from a Federal Bureaucracy." *American Economic Review* 87(2): 383–388.

- . 2002. “Performance Incentives with Award Constraints.” *Journal of Human Resources* 37(4): 812–845.
- . 2003. “Performance Funding in Federal Agencies: A Case Study of a Federal Job Training Program.” *Public Budgeting and Finance* 23(3): 22–48.
- . 2004. “An Empirical Investigation of Gaming Responses to Performance Incentives.” *Journal of Labor Economics* 22(1): 23–56.
- Cragg, Michael. 1997. “Performance Incentives in the Public Sector: Evidence from the Job Training Partnership Act.” *Journal of Law, Economics, and Organization* 13(1): 147–168.
- Dickinson, Katherine P., Richard W. West, Deborah J. Kogan, David A. Drury, Marlene S. Franks, Laura Schlichtmann, and Mary Vencill. 1988. *Evaluation of the Effects of JTPA Performance Standards on Clients, Services, and Costs*. Research Report No. 88-16. Washington, DC: National Commission for Employment Policy.
- Division of Employment and Training, New Jersey Department of Labor. 1990. *State of New Jersey Performance Standards Manual, PY 1988–89*. Trenton, NJ: Division of Employment and Training, New Jersey Department of Labor.
- General Accounting Office (GAO). 1996. *Executive Guide: Effectively Implementing the GPR*A. GAO/GGD-96-118. Washington, DC: GAO.
- Hatry, Harry. 1980. “Performance Measurement Principles and Techniques: An Overview for Local Government.” *Public Productivity Review* 4(4): 313–315.
- Healy, P. 1985. “The Effect of Bonus Schemes on Accounting Decisions.” *Journal of Accounting and Economics* 7(1–3): 85–107.
- Heckman, James, Carolyn J. Heinrich, and Jeffrey A. Smith. 1997. “Assessing the Performance of Performance Standards in Public Bureaucracies.” *American Economic Review* 87(2): 389–395.
- Heckman, James J., and Jeffrey A. Smith. 1995. “The Performance of Performance Standards: The Effects of JTPA Performance Standards on Efficiency, Equity, and Participant Outcomes.” Unpublished manuscript. University of Chicago.
- Heckman, James J., Jeffrey A. Smith, and Christopher Taber. 1996. “What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program.” NBER Working Paper No. 5535. Cambridge, MA: National Bureau of Economic Research.
- Hurst, E. 1980. “Attributes of Performance Measures.” *Public Productivity Review* 4(1): 43–46.
- Lipsky, Michael. 1980. *Street Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.

- Marschke, Gerald. 2002. "Performance Incentives and Bureaucratic Behavior: Evidence from a Federal Bureaucracy." Unpublished manuscript. University at Albany, State University of New York.
- Niskanen, William A. 1971. *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton.
- Oyer, P. 1995. "The Effect of Sales Incentives on Business Seasonality." Unpublished manuscript. Princeton University, Princeton, NJ.
- Stiefel, Leanna, Ross Rubenstein, and Amy Ellen Schwartz. 1999. "Using Adjusted Performance Measures for Evaluating Resource Use." *Public Budgeting and Finance* 19(3): 67–87.
- Svorny, Shirley. 1996. "Congressional Allocation of Federal Funds: The Job Training Partnership Act of 1982." *Public Choice* 87(3–4): 229–242.
- Thompson, James R. 2000. "The Dual Potentialities of Performance Measurement." *Public Productivity and Management Review* 23(3): 267–281.
- Tirole, Jean. 1994. "The Internal Organization of Government." *Oxford Economic Papers* 46(1): 1–29.
- Trott, Charles E., and John Baj. 1987. *Development of JTPA Title II-A Performance Standards Models for the States of Region V*. DeKalb, IL: Center for Governmental Studies, Northern Illinois University.
- U.S. Congress. 1982. *Job Training and Partnership Act*. Public Law 97-300, 29 U.S.C. §1501. Washington, DC: U.S. Government Printing Office.
- Usilaner, B., and E. Soniat. 1980. "Productivity Measurement." In *Productivity Improvement Handbook for State and Local Government*, G. Washnis, ed. New York: Wiley.
- Wholey, Joseph S. 1999. "Performance-Based Management: Responding to the Challenges." *Public Productivity and Management Review* 22(3): 288–307.

