Performance Incentives and Organizational Behavior: Evidence from a Federal Bureaucracy

Gerald Marschke[1,2]

October 2002

Abstract: This paper examines the effects of performance incentives in a federal job training program for the economically disadvantaged. A natural experiment that exogenously varies the incentives that government workers face allows me to identify incentive responses, which I find are consistent with a simple model of organizational behavior. Additionally, I show that the program's incentive designers have difficulty identifying performance measures that are aligned with the program's objectives. I discuss my results in the context of the greater incentive literature (especially the multi-tasking literature), as well as the literature on incentives in job training programs. (*JEL* J33, J24, D73, I38)

Keywords: Performance incentives, performance measurement, multi-tasking, job training programs, government bureaucracy.

1 Introduction

The absence of accountability in government agences is often blamed for the disappointing record of government provision of many public services. Performance measurement backed by financial or other kinds of incentives have a surface appeal because they offer the hope of aligning the preferences of governments' stakeholders and its agencies. Advocates of such accountability devices in government argue that by shifting the focus away from input-monitoring, supervision, and rules toward objectives, these devices encourage government agencies to use their local information and initiative to achieve more efficient outcomes.[3] This is the argument behind recent legislation requiring the overseers of federal agencies to identify and implement performance measures that track performance.[4] The intuitive appeal of this argument, however, contrasts with the empirical evidence of its support.

One place to look for evidence is the incentive system in place in the federal government's largest stand-alone job training program for the economically disadvantaged. In the early 1980s, federal legislation replaced many of the bureaucratic rules that regulated organizational *conduct* with objectively measured, financially-backed performance incentives designed to influence *outcomes*. The central aim of this paper is to understand the influence of job training performance incentives on organizational behavior and outcomes, and to test whether incentives affected the delivery of job training services. Did the financially-backed incentive system raise performance as its advocates claim or did it instead distort productivity? A natural experiment that exogenously varies cross-sectionally and across time the incentives training agencies face allows me to identify the effect of incentives on productivity.

While a federal job training agency is its focus, this study sheds light on broader issues of the incentive literature. The analytic basis of the study of performance measurement is the principal-agent model, which has traditionally focused on the risk-productivity trade-off in compensation. There is a trade-off because while pay-for-performance elicits higher productivity it also imposes higher risk on the agent, which the principal pays for through higher wages.[5] Recent extensions of the model address incentive problems that are characteristic of government generally and government job training programs in particular. Baker (1992, 2002) and Holmstrom and Milgrom (1991) consider the danger to the organization of imposing explicit incentives when good proxies of worker effort are unavailable.[6] These models show that when only poor proxies of agent effort are available, the optimal contract is low-powered. The results reported here underscore the challenge government designers of

performance measures face when organizational value-added is difficult to measure.

While the theoretical literature is well-developed, empirical evaluations of specific incentive contracts in organizations remain rare. An empirical study of the incentives in this program contributes to economists' understanding of incentives in organizations in at least three important ways. First, the basic assumption underlying the principal-agent model is that agents respond to contracts that reward performance. Interestingly, with few exceptions, the question "Do incentives in organizations matter?" has not been addressed empirically using *direct measures of agent-level incentives and outcomes.*[7] My ability to relate precisely formulated, agent-level incentives to precisely measured agent-level performance outcomes, activities, and productivity sets this paper apart from most other empirical studies of incentives in organizations.

Second, it is a widely recognized limitation of the incentive literature that it does not explain well the relative scarcity of high-powered incentive systems in the private sector (e.g., Baker, Jensen, and Murphy, 1988). Because incentives play an important role in eliciting optimal effort from workers in the theoretical literature, the scarcity of incentives is especially peculiar. The possible explanations for the absence of incentives include the costs of writing and enforcing contracts, risk-aversion of workers, and difficulty measuring performance. By understanding how organizational incentives influence behavior in practice we may better understand why they are scarce and under what circumstances various contractual forms will arise. This paper is an empirical exploration of the problem of measuring performance. Here, one observes exogenous variation in the rewarded performance measures, which are not direct measures of the social value added of government agencies. This paper shows how this variation changed agency activities, and how these changes in turn affected value-added.

Finally, in the data this paper exploits one observes the evolution of an incentive system. The process by which incentives evolve is little studied in the incentive literature, but may be crucial to how compensation systems arise in organizations and how management in organizations can be improved (Bruns, 1992). Congress provided funds for the ongoing scrutiny and monitoring of the effects of the incentives on the federal job training effort. The program's incentive designers, in an effort to improve the measurement of the program objective, have from time to time overhauled the set of performance measures used to reward training agency managers. An important contribution of this paper to show how these adaptations have changed training outcomes and the effectiveness of training incentives. In

other words, one can observe whether the reforms yielded performance measures that are less "misaligned" and "distortionary" (in the sense of Baker, 2002).

The remainder of the paper is organized as follows. Section 2 explains the organizational environment in which the job training program agencies operated and the incentive system they faced. A key point of this section is that the agency manager enjoyed extensive independence from state and federal authority. Section 2 also briefly reviews the empirical literature examining the effects of incentives in federal job training programs. Section 3 lays out the model of incentive response and a strategy for testing it. The data used in this study and the training agency's training choices are described in Section 4. Section 5 details how the data are used to implement the scheme outlined in Section 3. Section 6 describes the results of the estimation of a model of incentive response and discusses the limitations on the analysis imposed by the data. Section 7 uses the results reported in Section 6 to simulate the effects of incentive changes on organizational behavior, performance outcomes, and a measure of organizational efficiency. Finally, Section 8 concludes the paper.

2 Organizational environment: discretion and incentives

The current federal effort to provide job training to the poor can be traced to the Kennedy administration. The job training bureaucracy evolved from the initial enabling legislation in the 1960s through the present. With the enactment of Job Training Partnership Act (JTPA) in 1982, federal job training efforts became much more decentralized. In addition to permitting unusual discretion at the local level, JTPA implemented a financially-backed incentive system. Since July 2000, the Workforce Investment Act (1998) has governed these efforts. The Workforce Investment Act retains the decentralized nature, most of the jurisdictional borders (see below) and administrative entities, and, generally, the performance incentives of JTPA. Because this paper focuses on the period under JTPA, however, this section sketches the relevant features of JTPA.[8]

JTPA divided the U.S. into approximately 640 non-overlapping, collectively exhaustive jurisdictions. 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.

Within each sub-state region, a single administrative agency managed the region's budgetary appropriation.

JTPA was not an entitlement. Thus, who among the eligible were enrolled was left to the agency. Also, through the period that I study these agencies (the late 1980s), federal rules did not limit the length, kind, or expense of training activities.[9] Thus agencies enjoyed wide discretion over who they enrolled, how many they enrolled, and the kinds of training they offered their enrollees. That agencies exercised their discretion is clear from the cross-sectional variation in training strategies. Average expenditure per enrollee and the distribution of services varied widely. For example, in 1986, across the sixteen states[10] that I study here, the average expenditure per enrollee varied from a low of approximately $300 to a high of approximately $3000. In one agency in California, 89 percent of enrollment spells contained a classroom based training activity. By contrast, in one agency in Missouri, 94 percent of enrollment spells contained on-the-job training.[11] This study examines the influence of incentives on this variation.

2.1 Performance incentives

In the early 1980s, Congress restructured the federal job training effort to make more decentralized. By locating discretion within training agencies, Congress hoped to exploit case workers' specific knowledge of local economies and enrollees. This delegation, however, created an agency problem. To encourage self-interested agencies to allocate training resources in ways consistent with the program's goals, Congress provided for a system of financially-backed performance incentives. This subsection describes the incentive system.

Under the new rules, states were required to set aside six percent of their annual appropriation to fund an incentive system that rewarded successful agencies. For some agencies, the size of the incentive award was substantial. In 1986, for example, the most successful agency won the equivalent of almost 60 percent of its budget, while the median award was equivalent to 7 percent of the agency's budget.[12] Awards training agencies won augmented their budget in the following year.[13]

The act directed the U.S. Department of Labor to establish performance measures that promoted JTPA's mission. JTPA's stated mission was to increase the earning ability of the economically disadvanaged through job training (Section 106(a)).[14] Measuring worker output then required knowing what each trainee would have earned in the absence of training,

which of course is not observable. In the absence of a measure of human capital impact, the Department of Labor used performance measures based upon easily measurable labor market outcomes of trainees at or shortly after training. The Department of Labor attempted to get from labor market measures to a measure of value-added by comparing the average performance outcome to an outcome predicted for a hypothetical representative agency in an identical economic environment and with the same enrollee population. In practice, the Department of Labor computed predictions of outcomes—performance standards—as a fixed percentile of the distribution of performance based on historical data for all agencies nationwide. These standards were then adjusted for local economic factors which were outside the agency's control and for characteristics of the participant population.[15] In adjusting the standards by enrollee characteristics, the Department of Labor meant to limit incentives for the agency to choose enrollees based on enrollees' projected scores on performance measures, that is, to *cream skim*.[16]

The incentive schemes were in effect for one year, which was coincident with the program's funding cycle, beginning on July 1 and ending on June 30 of the following calendar year. This unit of time is referred to as a *program year*.[17] In any program year, all agencies within a state faced identical performance incentives. Table 1 defines all outcome-based performance measures that were used in at least one state's incentive scheme between 1987 and 1989. As Table 1 shows, performance measures had two characteristics in common: all performance measures were (1) based on aspects of the enrollee's labor market status either at the date the enrollee officially exited—or *terminated*—from the program, at 90 days after termination, or over the 90 day period following termination; and (2) *averages* of outcomes over the population of the program year's *terminees*. Thus agencies were rewarded annually according to their overall labor market success *rates* among enrollees *terminated*. For example, the employment rate at termination (ERT) for the fiscal year 1987 was defined as the fraction of enrollees who terminated between July 1, 1987 and June 30, 1988 who were employed on their termination date. The system's other performance measures were the *average wage at termination* (AWT), the *cost per employment* (CE), the *employment rate* at 90 days after termination—*i.e.*, *at follow-up*—(ERF), the *average weeks worked at follow-up* ($AWWF$), and the *average weekly earnings at follow-up* ($AWEF$). Adult enrollees who were welfare recipients counted as adults in the measurement of all of the performance outcomes described above. In addition, the employment rates of *welfare* recipients was sometimes

judged additionally under separate measures, the *welfare employment rate at termination* ($WERT$) and the *welfare employment rate at follow-up* ($WERF$). The Department of Labor imposed lower standards for the measures $WERT$ and $WERF$ than for the measures ERT and ERF.

Table 2 reports which of the performance measures were in effect by state and year of the study, identifying the sixteen agencies under study (see Section 4) by the name of the city, county or region they serve. Table 2 shows that performance measures varied considerably between states, and within a state across time. The years of their use across states were not uniform because in some years the Department of Labor allowed states to choose from a menu of measures. Table 2 documents two important trends in the use of performance measures: the elimination of 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 based upon outcomes at follow-up. By measuring labor market outcomes after some time had elapsed since training concluded and by dropping the cost measure, the Department of Labor hoped to encourage agencies to offer more substantive training that would produce longer-lasting impacts on enrollees' skills. See Barnow (1992, 1993) for discussions of the Department of Labor's reasons for these changes. The Department of Labor required states to make these changes, but gave states some leeway in how quickly they were implemented.[18] Thus, different states made these transitions in different years. This exogenous variation in the performance measures is the basis of the methodology in this study.[19]

Subject to the requirement that awards be paid only for performance that exceeded standards,[20] states determined how federal performance measures were to be translated into incentive awards. Consequently the functional form of the award varied from state to state. Nevertheless, the defining feature of the award for all states is that by simply meeting the numerical standards corresponding to a set of performance measures agencies won a substantial monetary award.[21]

2.2 JTPA incentive literature

A small literature has grown up around the subject of JTPA incentives. One question addressed in this literature is whether JTPA bureaucrats responded to performance incentives.[22] The effect of incentives on enrollment, i.e. the cream skimming issue, has been the focus of this inquiry.[23]

The evidence that training centers chose enrollees based on their expected effects on performance outcomes is, at best, mixed. Anderson, Burkhauser, and Raymond (1993) examined the enrollment patterns of training centers in Tennessee for evidence that JTPA bureaucrats prefered persons who were likely to score high on a single JTPA performance measure—the employment rate at termination. They find that some of the personal characteristics that predict high post-training employment rates are positively correlated with the likelihood of enrollment. Nevertheless, it is impossible to tell from their study whether this relationship represents bureaucratic screening behavior, the applicant's decision whether to enroll, or a combination of the two. They do not consider performance measures other than the employment rate.

Cragg (1997) contrasted the characteristics of a sample of about 200 JTPA enrollees distributed across the fifty states to a sample of JTPA eligibles. He found some evidence that in states in which the award was more challenging to obtain, training centers pursued high-aptitude eligibles more aggressively.Because the eligible person's decision to apply to a program was not likely to be influenced by whether the state he or she resides in makes the award easy or difficult to obtain, Cragg argues this is evidence of cream skimming and not a depiction of the applicant's decision.

Heckman and Smith (2002) investigated the cream skimming issue using a subset of the data used in this study. In the training agencies they study, Heckman and Smith could identify who among the accepted eventually enrolled in the program. They found that applicants who were expected to produce low performance outcomes were less likely to enroll conditional on acceptance. Their evidence also suggests a selection process that some would consider inequitable: blacks and the least educated and experienced were less likely than others in the applicant pool to enroll. Nevertheless, as was the case with Anderson et al. and as the authors acknowledge, some or all of the evidence may be due to self-selection on the part of applicants.

Heckman, Smith, and Taber (1996) studied the determinants of the enrollment decision at a single Texas training center. For this training center, they had data not only on which persons were accepted into the program but also on who applied. They argued that whether the applicant reaches the acceptance stage mostly reflected the preferences of training center bureaucrats, not the preferences of applicants. The authors found that applicants with the lowest prospects for employment after training were more likely to be accepted in the

program. They also found that the applicants with the greatest expected earnings gain were neither more nor less likely to be accepted.[24]

Thus, the evidence as to whether training center bureaucrats expressed a preference for eligible persons who were most likely to produce high measures of performance is mixed. The best evidence that such cream skimming occurs comes from Cragg and Heckman and Smith. Only the Cragg study is a test of incentive *response*—that is, only Cragg contrasts training center behavior under two different incentive regimes—and his evidence is neither strong nor consistent. Moreover, these studies test whether training agencies prefer to the enroll the kind of person who would likely be employed whether or not they received training. Yet training agencies who were motivated by awards cared how their enrollment choices boost expected *net* employment, that is, expected employment in excess of the standard. The Department of Labor's adjustment model forced the training center to consider how a person's attributes not only raised the performance outcome but also the standard the outcome was measured against. Thus, the absence of strong evidence of cream skimming may be because the performance standard adjustment procedure was doing its job—that is, reducing the incentive to cream skim—and not because JTPA bureaucrats did not respond to incentives.

This paper's empirical strategy contrasts with previous JTPA studies in three important ways. First, this study examines the influence of JTPA's performance on agencies' training choices. Because performance standards are not adjusted by training choice, the influence of incentives on training choice is much simpler to model empirically. A two stage procedure uses estimates of performance outcomes that reflect all performance measures to explain training choice. The test conducted in this paper is a simple one: are training agencies that are rewarded for high employment outcomes, say, encouraged to choose the kinds of training that promote higher employment rates more often than training agencies who are not so rewarded? Second, this study examines the influence of all—not just the employment rate-related—performance measures facing training agencies. Third, this study disentangles bureaucratic response to incentives from enrollee behavior. The empirical strategy accounts directly for enrollee preferences and allows them to act on the activity choice through a predicted earnings impact measure. This study also exploits the cross-sectional and longitudinal variation in incentives to difference out the unmeasured influence of enrollee preferences. The empirical methodology is explained fully in the next section.

3 The incentive scheme and training decision: an empirical strategy

3.1 Training decisions and the award

In this subsection, I sketch a model of training agency behavior. Training agencies that are strict award maximizers choose enrollees and training to maximize the probability of meeting the performance standards they face subject to meeting their budget constraint. Agencies choose from the applicant pool persons based upon observed characteristics. The characteristics—education, age, and skills—of the enrollees they serve as well as the state of the local labor market, are important to agencies because they interact with training to influence outcomes. Between 1987 and 1989 the states in this study followed a Department of Labor procedure that adjusted the performance standards each agency faced by the characteristics of the enrollee population and the local economy (see Section 2.1). Assuming that the adjustment procedure perfectly adjusted the standard for differences in the characteristics of enrollees, agencies won no advantage by cream skimming, but because the standards were not adjusted by the kinds of training offered they may benefit by carefully matching training to enrollees. The agency chose from a discrete set of activities for each enrollee a single training activity. Let $\boldsymbol{X}_i$ be the vector of characteristics describing enrollee i and the area's economy at the time of i's training. Let τ_i be the agency's activity choice for enrollee i. Assuming that the performance standard adjustment procedure perfectly adjusts the standard for differences in $\boldsymbol{X}_i$, only the training choice matters for the award, and the agency matches enrollees to training, subject to its budget constraint, to maximize

$$\sum_k \psi^k \, Pr_k[\{\tau_i\}_{i=1}^N], \tag{1}$$

where ψ^k is the award set aside for measure k.

The awards for the performance measures, $\{\psi^k\}$, varied across states and over time. ψ^k was zero in some state-years and positive in others. My analysis exploits this variation to identify the effects of incentives on the assignment of individual enrollees to training activities. I take a reduced form approach in testing for incentive effects. I argue that if incentives matter, one should observe that a performance outcome will more often guide the training choice when it is compensated then when it is not compensated. I test agency

incentive response by using the agency's training choices to infer whether adding a measure to the award increased the agency's supply of the corresponding outcome through a reallocation of enrollees across training activities.

3.2 Empirical model

I assume that the marginal contribution of a single enrollee i in an activity τ to the agency's ex ante utility is

$$U_{i\tau} = g(\tilde{\boldsymbol{S}}_{i\tau}, \boldsymbol{\psi}, \tilde{\Delta}_{i\tau}, \boldsymbol{R}_i, u_{i\tau})$$

$\tilde{\boldsymbol{S}}_{i\tau}$ is a $1 \times K$ vector of expected performance outcomes. That is, the training agency makes a prediction about i's performance for each measure k in each training activity τ, and these are contained in the vector $\tilde{\boldsymbol{S}}_{i\tau}$. The training agency bases its prediction on the observable characteristics of the clients and on the state of the labor market. $\boldsymbol{\psi}$ mediates the effect of $S_{i\tau}$ on the award and therefore enters the agency's utility. In addition, $g(\cdot)$ allows the performance on a measure to have some value to the training agency unrelated to the award. $\tilde{\boldsymbol{S}}_{i\tau}$ includes the expected cost of enrollee i's activity τ training. Any agency regardless of its objectives and the cost-pay off in the award takes cost into consideration because of its budget constraint.

$\tilde{\Delta}_{i\tau}$ and $\boldsymbol{R}_{i\tau}$ capture non-award influences on training choices. Some writers have maintained that bureaucrats in the welfare bureaucracy, because of their training and the culture of the social work profession, behave as their clients' advocates (e.g. Lipsky, 1980, p. 72). The advocacy may take several forms. First, because it is a likely concern of enrollees, case workers in training agencies may be influenced by earnings impacts. $\tilde{\Delta}_{i\tau}$ is an estimate of training activity τ's impact on enrollee is earnings. Second, much popular writing and public discussion about welfare workers and case workers in job training in particular imagines social workers as benefactors to the least advantaged in their case load. The "social worker mentality" impels them to save the most intensive and expensive training and schooling slots for the enrollees who are welfare recipients or high school dropouts—that is, for enrollees who have the most extreme labor market "barriers" to finding gainful employment. In fact federal and state directives frequently exhorted training agencies to provide special attention to enrollees with these barriers. Heckman et al. test for the social worker motive in the enrollment decision and find that case workers in one JTPA agency preferred to enroll

applicants who were relatively disadvantaged, where disadvantage is indicated by low predicted earnings in the months following enrollment. $\boldsymbol{R}_i$ includes measures of disadvantage. $u_{i\tau}$ are the unmeasured determinants of i's contribution to the training agency's utility from activity τ.

To make the problem empirically tractable, let us assume that the marginal contribution of a single enrollee in an activity τ to the agency's utility can be written as

$$U_{i\tau} = \theta_{\tau 0} + \boldsymbol{R}_i\boldsymbol{\theta}_\tau + \tilde{\Delta}_{i\tau}\theta_\Delta + \tilde{\boldsymbol{S}}_{i\tau}\boldsymbol{D}_i\boldsymbol{\pi}_A + \tilde{\boldsymbol{S}}_{i\tau}(\boldsymbol{I} - \boldsymbol{D}_i)\boldsymbol{\pi}_N + u_{i\tau}, \tag{2}$$

where $\tilde{\boldsymbol{S}}_{i\tau}$ and $\boldsymbol{R}_{i\tau}$ are defined as above, and $\theta_{\tau 0}$ is a constant. $\boldsymbol{D}_{i\tau}$ is a $K \times K$ diagonal matrix; the kth element on the diagonal is 1 if i is assigned to training under a scheme that rewards the kth performance outcome ($\psi^k > 0$), and zero otherwise ($\psi^k = 0$). $\boldsymbol{I}$ is a $K \times K$ identity matrix. $\boldsymbol{R}_i$ and $\boldsymbol{D}_i$ are observed, but $\tilde{\boldsymbol{S}}_{i\tau}$ and $\tilde{\Delta}_{i\tau}$ must be estimated. $\theta_{\tau 0}$, $\boldsymbol{\theta}_\tau$, θ_Δ, $\boldsymbol{\pi}_A$, $\boldsymbol{\pi}_N$, are unknown parameters.

The agency selects choice $\hat{\tau}_i$ if it maximizes the agency's utility for enrollee i, i.e.

$$\hat{\tau}_i = \underset{\tau}{argmax}\{U_{i\tau}\}_{\tau=1}^{T}. \tag{3}$$

$u_{i\tau}$ is an independently (across individuals and choices) and identically distributed stochastic component of the agency's utility. Further, the distribution of $u_{i\tau}$ is assumed normal with mean zero, and variance σ_{u_τ}. Under these assumptions, equations (2) and (3) produce a conventional probit model.

The focus of the model's estimation is the utility weights $\boldsymbol{\pi}_A$ and $\boldsymbol{\pi}_N$ corresponding to the predicted performance outcomes. By measuring the responses of the utility weights in (2) to differences in $\{\psi^k\}$ I am able to test whether agencies respond to incentives. I test whether ceteris paribus an increase in a performance outcome's marginal return in the award leads to an increase in its utility weight. That is, I test whether $\pi_A^k - \pi_N^k > 0$ for each measure k. (For the cost measure, I test whether this difference is negative.) Although the above approximating model is not an explicit structural model, it can shed light on whether the estimation of a structural model is likely to be successful.

4 The sample and training options

The data source for this study is the National JTPA Study (NJS), an experimental study of the effectiveness of JTPA commissioned by the U.S. Department of Labor and conducted

between 1987 and 1989. Sixteen of the organization's roughly 640 job training agencies participated in the NJS.[25] The study was conducted using a classical experiment methodology according to which JTPA applicants were randomized into treatment and control groups. The control groups did not receive JTPA training services for at least 18 months after random assignment. 20,601 JTPA-eligible adults and youth participated in the study: 13,972 were randomized into the treatment group and 6,629 into the control group.

The empirical analysis in this study is based on 13,338 usable adults from the set of participants in the NJS. The data generated by the NJS are often likely equal in scope and detail to the quantifiable information available to training center bureaucrats. The data contain participant-reported information on his/her education level, labor market history, family composition, welfare program participation and demographic characteristics, as well as labor market, training, and schooling activities for approximately 18 months after random assignment.[26] In addition, the data contain enrollment and termination dates for all experimental participants who received training services. These program dates can be used with the participant employment spell, earnings, and wage data to produce accurate measures of performance outcomes.[27]

Before the experimental participants of the NJS were randomized into treatment and control groups, agency bureaucrats assigned each participant to one of three activities: classroom occupational training (CTO), on-the-job training (OJT), or job search assistance (JSA).[28] Thus, $T = 3$ in equation (3). The three activities varied significantly in the extent of contact between enrollee and agency, the enrollment length, and training intensity. Enrollees in classroom occupational training learned occupation-specific skills in a classroom environment. Classroom occupational training lead to the longest spells on average. Enrollees in on-the-job training worked a maximum of 6 months, during which their employers received as much as 50 percent of the wage they pay as a subsidy. The extent and character of training undertaken during the subsidy period was at the discretion of the employer. Job search assistance provided enrollees an "assessment of [their] job skills and interests, along with training in job-finding techniques and help in locating job openings" (Orr *et al.,* p. 4). The mean enrollment lengths for classroom occupational training, on-the-job training, and job search assignments were 7.2, 3.5, and 4.5 months, respectively.

Table 3 documents the variation in performance outcomes by training activity assignment. The range was approximately 30 cents an hour for the wage measure and $10 per

week for the earnings measure. The range in employment rates was more substantial: 16 percentage points, or about 27 percent of a base of 60 points. On-the-job training appears to have produced the highest employment rates, earnings at follow-up, and weeks worked outcomes, but yielded a low wage at placement. Job search assistance surpassed classroom occupational training in all measures except for the employment at termination measure, where the difference between the two training assignments was minor. This table suggests that different activities were advantageous for different performance outcomes, implying that if agencies behaved as budget-maximizers, we should see different mixes of activities under different incentive schemes.

Of course, enrollee differences across training activities accounted for some of the differences in performance outcomes across training activities. Table 4 shows the distribution of adults across activities by subgroup. (See Table 1 in the appendix for definitions of characteristics.) Table 4 shows that training activity participation varied by the demographic characteristics, earnings, education, and labor market history of participants in predictable ways. For example, the never-employed were much less frequently found in OJT compared to persons that had a recent history of employment. Classroom occupational training was less frequently assigned to the least-educated groups, probably because classroom training teaches advanced skills for specific occupations. Human capital models predict that older persons, with fewer years remaining in the labor force, invest in less training. Thus it is not surprising that the likelihood that someone was assigned to job search assistance—the shortest and most employment-directed of the activities—increased slightly with their age. It is also not surprising that the likelihood of assignment to job search assistance increased with earnings—a measure of the opportunity cost of training—over most of its range. The implication here is that whether for example classroom training was more suited to generating an employment than job search assistance depended upon the characteristics of the enrollee. Because of the systematic relationship between enrollee characteristics that influence outcomes and training activity types, it is natural to model the agency's expectations of outcomes $\{\tilde{S}_{i\tau}^{k}\}$ as functions of enrollee characteristics. The next section describes how this study models the agency's outcome expectations.

5 First stage estimation

The estimation of (2) proceeds in two stages. In the first stage, I estimate for each non-cost performance outcome the agency's expectation of an enrollee's performance in each of the various training activities. These estimates become regressors in the second stage multinomial probit.[29] Although we do not observe $\tilde{S}^k_{i,\tau}$ we do observe the actual outcomes $S^k_{i,\tau}$ for each enrollee in the training option actually assigned. Assume that the actual outcomes are related to the agency's prediction in the following way.

$$S^k_{i,\tau} = \tilde{S}^k_{i,\tau} \; + \; \nu^k_{i,\tau}, \tag{4}$$

where $\mathrm{E}\nu^k_{i,\tau} = 0$ for each k. The deviations $\nu^k_{i,\tau}$ arise because bureaucrats lack full information about all of the factors that influence outcomes.

The employment rates at termination and at follow-up measure the state of employment at a point in time. These variables are binary: an enrollee is either employed or not employed on the measurement date. The other measures—of wages, earnings, and number of weeks worked—are continuous. Let the superscript E denote employment measures and the superscript C denote continuous measures. The agency's expectation for the employment measures range between 0 (certain unemployment) and 1 (certain employment). Assume $\tilde{S}^E_{i,\tau}$, the agency's expectation of i's outcome for an employment measure, is described by

$$\tilde{S}^E_{i,\tau} = F(\boldsymbol{X}_i \boldsymbol{\gamma}^E_\tau) = \frac{\exp(\boldsymbol{X}_i \boldsymbol{\gamma}^E_\tau)}{1 + \exp(\boldsymbol{X}_i \boldsymbol{\gamma}^E_\tau)}, \tag{5}$$

where $\boldsymbol{X}_i$ is a vector of observed attributes of i and of the labor market at the time of the training assignment, and $\boldsymbol{\gamma}^E_\tau$ is a vector of unknown parameters. Assuming that $F(\cdot)$ is the logistic distribution implies the third term in (5). Assume that the bureaucrat's prediction of the outcome for the continuous measure is given by

$$\tilde{S}^C_{i,\tau} = \boldsymbol{X}_i \boldsymbol{\gamma}^C_\tau \tag{6}$$

where $\boldsymbol{\gamma}^C_\tau$ is a vector of unknown parameters.

This framework implies that the agency forms its expectations according to (5) and (6) but that the realized performance outcomes occur by (4). The realized outcomes differ from

the expectations due to the incompleteness of the agency's information set. My assumptions imply that the realized outcomes arise from (4) after substituting in (5) and (6), that is,

$$\begin{aligned} S^E_{i,\tau} &= \frac{\exp(\boldsymbol{X}_i\boldsymbol{\gamma}^E_\tau)}{1+\exp(\boldsymbol{X}_i\boldsymbol{\gamma}^E_\tau)} + \nu^E_{i,\tau}, \text{ and} \\ S^C_{i,\tau} &= \boldsymbol{X_i}\boldsymbol{\gamma}^C_\tau + \nu^C_{i,\tau}. \end{aligned}$$

The use of $\exp(\boldsymbol{X}_i\boldsymbol{\gamma}^E_\tau)/1+\exp(\boldsymbol{X}_i\boldsymbol{\gamma}^E_\tau)$ and $\boldsymbol{X}_i\boldsymbol{\gamma}^C_\tau$ for $\tilde{S}^E_{i,\tau}$ and $\tilde{S}^C_{i,\tau}$ is justified if consistent estimates of $\tilde{\boldsymbol{\gamma}}^E_\tau$ and $\tilde{\boldsymbol{\gamma}}^C_\tau$ are available. An estimator of $\boldsymbol{\gamma}^E$ from (5) is obtained under the assumption that F is the logistic distribution. $\boldsymbol{\gamma}^E$ is then estimated using maximum likelihood. An obvious candidate for $\hat{\boldsymbol{\gamma}}^C$ is the least squares estimator.[30]

To summarize the first stage proceeds as follows. First, using the outcomes and enrollee characteristics, $\hat{\boldsymbol{\gamma}}^E$ is obtained from estimating (5) under the assumption that F is the logistic distribution. Similarly, $\hat{\boldsymbol{\gamma}}^C$ is obtained by estimating (6) by least squares. For each individual i in the data, estimates for predicted values of $S^E_{i,\tau}$ and $S^C_{i,\tau}$ for all activities are formed from the $\hat{\boldsymbol{\gamma}}^E$, $\hat{\boldsymbol{\gamma}}^C$, and the enrollee's characteristics. Predicted earnings impacts in each activity are estimated using the experimental participants' earnings.[31] Estimates of the per capita cost of each activity are borrowed from an analysis of JTPA training contracts with external vendors from a suburban Illinois training agency (Heinrich, Marschke, and Zhang, 1998). These estimates of training costs vary with the characteristics of enrollees.

The second stage in the empirical strategy is to estimate equation (2) using maximum likelihood, including as regressors the estimated performance outcomes, costs, and earnings impact predictions for each training activity τ.[32] The next section presents estimates of $\boldsymbol{\pi}_A$ and $\boldsymbol{\pi}_N$, and the results of a test for an incentive response, a test of whether $\pi^k_A > \pi^k_N$.

6 Probit results

13,338 training choices under eight distinctive incentive regimes provide the basis for the empirical analysis. My most important finding is that the set of performance measures in place influences bureaucratic decision-making. In addition, I find that cost but not predicted earnings impacts influence bureaucratic decision-making. Finally I find mixed support for the hypothesis that JTPA bureaucrats allocated the least well-off among their enrollee population to the most intensive training activities.

Table 5 displays the results from the estimation of four alternate specifications of (2).[33] Model 1 displays the simplest specification, wherein the choice-specific predicted earnings and costs are the sole determinants of training choice. Model 2 allows case workers to take into consideration predicted earnings and costs but also measures of "disadvantage" or "hardship". Model 3 adds to this specification measures of predicted performance, i.e. award concerns. Model 4 allows for more subtle effects of performance incentives on performance outcome utility weights then does Model 3. The table includes the corresponding z statistic under each coefficient estimate.[34] The base of the table contains diagnostic statistics, including the likelihood ratio index, a measure of fit.[35]

6.1 Non-award determinants of choice

First consider column 1. The signs of the coefficients are consistent with a net earnings maximand. The coefficients suggest that agencies prefer low-cost activities and activities which promote the greatest boost to the enrollee's earnings ability ceteris paribus. The inference on costs holds across specifications. The coefficient estimates on earnings impact are not statistically significant in more general specifications, however. Table 6 contains the estimated elasticities derived from the coefficient estimates on the earnings impact and cost variables (these estimates are computed from Model 4's results). According to the estimated elasticities in Table 6, a doubling of the training's predicted cost will reduce the probability of classroom training, on-the-job training, and job search assistance by 50, 47, and 42 percent, respectively. Model 2 allows for the operation of a social worker motive in which case workers seek to match the most disadvantaged enrollee with the most intensive and expensive services (i.e., classroom training). Model 2 includes three measures of disadvantage: earnings in the year prior to enrollment, high school completion, and welfare recipiency.[36]

A likelihood ratio test shows these characteristics are jointly significant at conventional levels of significance. The z values are large for all coefficient estimates in all models except for the variables "Previous year's earnings less than $3,000" and "High School Dropout" in the job search assistance choice, for Model 2. In the more general specifications, their coefficient estimates are significant. Table 7 shows the average effect of these characteristics on the relative probability of being assigned to classroom training. The average effects are computed from estimates produced by Model 4. The signs of the effects of these characteristics on the three choices are consistent across the specifications. A number greater (less) than one

indicates that the relative probability of being assigned to classroom training falls (rises) with the exhibition of the characteristic. Being on welfare raises the likelihood of being assigned to classroom training relative to on-the-job training and to job search assistance. The evidence shows that being on welfare lowers the ratio of the probability of on-the-job training assignment to the probability of classroom training assignment by about one half. Having low pre-training earnings had little effect on these probability ratios. Being a high school dropout, however, lowers the probability of being assigned to classroom training relative to on-the-job training and to job search assistance; being a high school dropout raises the ratios by 46 percent and 15 percent, respectively. The finding that high school dropouts were assigned less often to classroom training may not be surprising. Dropouts may have rejected classroom training assignments because they disliked or lacked an aptitude for classroom training. While the finding that welfare recipients were assigned more often to classroom training is consistent with the social worker mentality hypothesis, the findings taken together do not consistently support the social worker mentality hypothesis.

6.2 Performance incentives and training choice

Models 3 and 4 allow for the influence of predicted performance outcomes on the activity choice. The coefficients on performance outcomes are allowed to be different for persons assigned to training under incentive schemes that rewarded the outcomes and schemes that did not reward them. Panel B presents only the estimated differences between the two sets of coefficients. I interpret these estimates as showing the effect on the choice of training of adding a performance measure to the set of rewarded measures. Panel B contains both the estimated differences and their z statistics.

Panel B is laid out as follows. The far left-hand column contains the set of performance measures whose status in state-years varied. In each line the second column corresponds to a performance outcome whose utility weight should respond to the activation of the measure in column 1. The third column contains the predicted sign of the utility weight response. For example, in the first line, the columns headed by "(3)" and "(4)" present estimates of the effect of activating the employment rate at follow-up measure on the value of increasing the follow-up employment outcome for a single enrollee.

Model 3 tests for an effect of activating a performance measure on the corresponding utility weight. Model 3 tests whether the utility weight is higher when the award places any

weight on the measure then when it places no weight on the measure. Activating either of the wage-based measures—AWT or $AWEF$—had little effect on the corresponding utilities. The utilities are statistically indistinguishable at conventional criteria of significance. Likewise, activating the cost measure had no significant effect on the cost outcome's utility weight. Nevertheless, the evidence of Model 3 provides some support for the simple model of incentive response. The activation of the employment rate at follow-up (ERF) and average weeks worked ($AFWW$) measures increased the value of their respective outcomes to the training center (the coefficient changes are positive with p values close to zero).[37] Model 4 produces the same inferences for this set of variables.

It is perhaps not surprising that the activation of employment-based measures show a greater effect on training decisions than does the activation of wage-related measures. As Table 3 shows, the variation in wage-based performance outcomes is far less than the variation in employment-based measures, suggesting that through its training choices, training agencies are able to exert more influence over employment than over wages.

Model 4 allows for more subtle effects of the incentives on utility weights. In the state-years studied here, some agencies made their training selections under incentive regimes that rewarded agencies separately for welfare recipients' employment outcomes. Under these regimes a welfare recipient's end-of-training employment outcome counted twice, once toward the overall employment outcome and again toward the separate employment outcome for welfare recipients. Thus by adding the welfare employment at termination ($WERT$) and the welfare employment rate at follow-up ($WERF$) measures, states added incentives to ensure employment for welfare recipients, beyond the incentive imparted by the overall employment rate measures (ERT and ERF). Lines 6 and 7 in Panel B describe the impact of activating the two welfare employment measures on the utility weights of their employment outcomes. The activation of the welfare employment rate at termination and welfare employment rate at follow-up measures as predicted increase the estimated utility weights for the outcomes $WERT$, and $WERF$, respectively, but these increases are insignificant by conventional criteria.

In addition, Model 4 allows for activating the cost performance measure to affect employment outcomes. Because the activation of the cost measure increases the value of training activities that more often terminate in employment, it should increase increase the value of the coefficient on employment outcomes for all adults. This accounts for the predictions of

positive effects on the utility weights for both employment rate at termination outcomes in the last two lines of Panel B. Activating the cost measure increases the utility weights of the two employment outcomes, ERT and $WERT$, as predicted, and both p values are very close to zero.[38]

6.3 Discussion

The results reported in Table 5 support a simple model of incentive response. Nevertheless, it should be emphasized that in the estimation of the cost (CE) effects, the contrast is provided by only two agencies; Omaha, Nebraska and Heartland, Florida; they are the only sites in this study that operated under an incentive regime without cost measures. In the cases where only two agencies are generating estimates of utility weights, the differences observed in utility weights may not reflect differences in incentive schemes so much as differences in non-award-related preferences for activities or differences in other factors not observed. In the cases where the number of agencies providing the contrast is small, this unobserved variation may not cancel out.[39]

While the cost results were produced by the Omaha and Heartland training agencies, the other strong findings in the analysis described in Table 5 arise from contrasting the behavior of two fairly representative groups of agencies. The effect of remunerating the employment rate at follow-up on its utility weight was estimated by comparing choices made by 13 agencies where the measure counted in the award to the choices made by 9 agencies where the measure did not count in the award. The average weeks worked at follow-up measure impact was estimated by comparing 5 agencies where the measure was compensated to 12 agencies where the measure was not compensated. Because they arise from the training choices of many more agencies, these results are more likely to be incentive-related.[40]

Another concern is that states had (limited) discretion over which of the federal performance standards they implemented. This opens up the possibility that the performance measures were not exogenous to the training agency's training assignment decisions. In his investigation of JTPA incentives and cream-skimming, Cragg (1997), using 1986 data on the JTPA incentive policies of all fifty states, also tests and rules out the hypothesis that incentives are endogenous. In a separate analysis, I test whether the variation in performance measurement is endogenous and reach the same conclusion.[41]

7 Simulating the effects of incentive changes on agency outcomes

Economists often argue that explicit incentives increase agent productivity because they tie agent compensation directly to measures of agent output. Yet this argument has rarely been tested using direct measures of performance and of incentives. This section explores the economic implications of the incentive responses documented in Panel B of Table 5 by simulating the effects of performance incentives on productivity.

The Department of Labor changed the set of performance measures during the 1980's to encourage agencies to offer lengthier and more intensive training activities. An important dimension of the social product of JTPA is the improvement training brings to participants' earnings ability net of training costs. Using the estimates from Table 5 one can test whether these changes likely improved total earnings gains net of training costs.

I use the estimates in column 4 of Table 5 and from the first stage analysis to simulate the effect of incentive changes on performance outcomes, earnings impacts, out-of-pocket training costs, and activity choice. I first simulate a basic incentive regime that is typical of the early years of JTPA. The basic regime consists of the cost per employment (CE), employment at termination (ERT), and wage at termination (AWT) performance measures. These three measures were present in all states in the early years of the program. In addition to the movement toward follow-up based measures and away from cost measures, the late 1980s saw the addition of performance measures of earnings and employment that measured performance over the three months following termination, in contrast to snap-shot measures of performance that measured employment and wage status on the last day of training. Federal and state JTPA authorities hoped that these changes would encourage agencies to offer more intensive training services that would produce longer-lasting effects.

To investigate the effects of these changes I contrast the base incentive regime with three alternative regimes. The alternative regimes differ from the base regime by the addition or elimination of one performance measure. The regimes are defined as follows.

1 Base regime (cost per employment at termination (CE), employment at termination (ERT), and average wage at termination (AWT)),

2 Base regime minus the cost measure (CE),

3 Base regime plus employment at follow-up (ERF), and

4 Base regime plus follow-up weeks worked ($AWWF$)

The comparisons I have chosen reflect the contrasts from Table 5 whose estimation are statistically significant.

Table 8 reports the results of the simulation. The rows of the table display the outcomes simulated. The outcomes shown are averages for the experimental sample of adult enrollees who enrolled in the program and had valid data.[42] The first column displays the outcomes under the base regime. The second, third, and fourth columns display the outcomes under regimes 2, 3, and 4.

In comparing column 2 to column 1 we isolate the effect of dropping the cost measure. Deactivating the cost measure reduces total earnings impacts by 16 percent, or by $160 per capita. Interestingly, deactivating the cost measure decreased total programmatic costs (by about .01 percent), but increased the cost outcome. This is possible if dropping the cost measure leads to a sufficiently large decrease in employments at termination.[43] The reduction in earnings impact is partly a consequence of the rise in the use of classroom training and a fall in the use of on-the-job training services. (This result is consistent with findings of Orr et al. that classroom training is less effective than the other services at increasing earnings after training.) The use of on-the-job training services falls because deactivating the cost measure lowers the value to the training agency of an employment immediately upon enrollment. The reduction in earnings impacts is also partly a consequence of the change in the composition of persons within each training type; this is not displayed in Table 8. The outcome changes reported in Table 8 generally result from both the overall change in the number of assignments to each activity and the change in the person-types assigned to each activity.

Column 3 shows the effect of adding a measure of the duration of employment over the three month period following training. Column 4 shows the effect of adding a measure of the employment status on the 90th day following training. In both cases, earnings impacts rise with the introduction of the new measure—by approximately $241 and $189, respectively—and costs rise slightly. Note that with the addition of follow-up measures, agencies use more on-the-job training and less classroom training overall. The total increase in net impacts—aggregate earnings impacts minus total out-of-pocket costs—due to the introduction of the average follow-up weeks worked and employment rate at follow-up measures are, in per capita terms, $232 and $183, respectively. Observe that the changes in per capita earnings impact

are modest, at best only $13 per month.

The intention of this section is to provide some rough estimates of the effects of incentive changes on agency productivity. These simulations provide credible estimates of the impact of changes in incentives on earnings impacts, costs, and training choice, provided the agency does not alter its screening rule and provided changes in the agency's training strategies do not substantially alter the applicant pool.[44] The simulation shows that by reducing the use of classroom training, introducing the follow-up measures increased per capita earnings impacts. Dropping the cost measure increased the use of more intensive and lengthier training activities and reduced the use of job search assistance, as intended. Yet, because more classroom training activities produced smaller increases in earnings over the 18 months following enrollment than other activities, deactivating the cost measure reduced the aggregate earnings impact of training.

8 Conclusion

The job training agencies in JTPA faced varying sets of performance measures. This paper shows that they responded to changes in performance measures by reallocating their resources across inputs as would be predicted by a simple model of incentive response, and similar to how multiproduct firms reoptimize their mix of products in response to changes in relative output prices. Such evidence of incentive response in real-world principal-agent relationships using direct measures of agent-level incentives and performance is rare.

The incentive-response findings raise the question of whether these incentive changes affected program efficiency. Without observing the operation of the program in the absence of an incentive system, we cannot shed much light on the question. A major concern of the theoretical incentive literature, however, is the market for unobserved inputs, or the market for effort. Holmstrom and Milgrom and Baker (1992, 1999) analyze this market in a multitask or multigoal context. JTPA compensates agencies for a number of goals but earnings gains is not one of them. This paper tests whether effort devoted to the performance objectives and to social welfare objectives are substitutes. This paper reports—in the simulation this is apparent—that when awards are provided for some measures, the return in earnings ability on a dollar spent on training declines, indicating substitutability in efforts across goals; other times, however, the return increases, indicating complementarity in efforts. Moreover, the simulation results point out the difficulty of performance measurement. This exercise shows

that some of the Department of Labor's reforms in the 1980s may have reduced the net earnings impacts of job training participants. That is, the Department of Labor appears to have had some difficulty in devising performance measures that correctly track organizational value-added, if indeed net earnings was its objective.[45]

The failure of some of the reforms to improve productivity suggests a shortage of easily identifiable performance measures that are aligned with the objectives of the organization and non-distorting.[46] While one can take the general implications of this case study of a government organization too far, this study lends some support to the argument that the rarity of the use of performance incentives in the workplace (as noted by Baker, Jensen, and Murphy, 1988, as well as others) and the dominance of fixed-wage contracts may be a consequence of the unavailability of performance measures sufficiently correlated with productivity. Until now the literature on this subject has focused on the risk-aversion of enrollees as the primary cause of this phenomenon.

My findings that JTPA bureaucrats choose services partly to meet performance standards are not inconsistent with the findings of Heckman, Smith, and Taber, who find no evidence that training agencies choose enrollees on the basis of their predicted employment outcomes. It is possible that the enrollee-selection decision is the domain of case workers, and the service selection decision is the domain of training center managers. This interpretation is bolstered by the case study of a training center done by Heinrich (1995), which describes the process that *managers* follow in deciding which outside vendors of training services to utilize. Moreover, as noted above, the two findings are not inconsistent because the incentive system is designed to influence the selection of training activities, more than it is designed to influence the selection of enrollees (see Section 2.1). Nevertheless, like Heckman et al., I do not find that predicted earnings impacts play a role in the assignment of training activities.

Notes

[1]Department of Economics and Department of Public Administration and Policy, University at Albany, State University of New York.

[2]This research was supported by NSF-SBR-91-31143 to J. Heckman, and grants from the Upjohn Institute and the American Bar Foundation. I thank Steve Cameron, Pascal Courty, Eric Gould, Hidehiko Ichimura, Robert Lalonde, Sherwin Rosen, Jeff Smith, Chris Taber, and seminar participants at The University of Chicago, University at Buffalo, University at Albany, Case Western Reserve, and the Upjohn Institute, as well as participants at conferences organized by the North American Econometric Society and the Society of Labor Economists. I am especially grateful to James Heckman and Canice Prendergast for many useful suggestions. The usual disclaimer applies.

[3]The organizational case studied here was a prototype for the Clinton/Gore National Performance Review initiatives, which makes this argument (Gore, Chapter 2, 1993).

[4]The Government Performance and Results Act of 1993 requires federal agencies (most for the first time) to formulate measures of performance to improve public accountability and permit closer scrutiny by Congressional oversight committees.

[5]At an empirical level, these models have been applied mostly in the analysis of executive compensation contracts (see, e.g., Jensen and Murphy, 1990, Haubrich, 1994, and Garen, 1994).

[6]See Dixit (1999) on why the worker output in the public sector is often more difficult to measure than in the private sector.

[7]Exceptions are Lazear (2000) and Paarsch and Shearer, (1996). Much of the literature on incentive response uses estimated or indirect measures of incentives, in larger part because organizations are reluctant to divulge their compensation practices. See, for example, Kahn and Sherer (1990) and Abowd (1990). Prendergast, 1999, and Gibbons, 1996 offer recent and thorough surveys of empirical studies of principal-agent models.

[8]See, e.g. Johnston (1987) for a detailed description of JTPA.

[9]The two significant exceptions: agencies were not allowed to subsidize a client's on-the-job training for more than 6 months at a time, and the agency was not permitted to place enrollees in public sector employment.

[10]See section 4 for a description of the data used here.

[11]These figures are based on data collected by SRI, International and Berkeley Planning Associates, which are described in Dickinson, et al. (1988). I thank Carol Romero of the National Commission for Employment Policy for making these data available to me.

[12]*ibid.*

[13]Even though the Act required training agencies to use award money to provide additional training, it placed fewer restrictions on the use of award monies than on budgetary monies. For example, training agencies were permitted to spend a larger fraction of their award on staff compensation, although few training agencies appeared to use the awards this way. Courty and Marschke (2002) argue that career concerns, the desire for professional recognition, and political motivations also made training center bureaucrats covet the award.

[14]Other goals mentioned in the Act include reducing welfare dependency among the poor (Section 106(a)), the equitable provision of services (Section 141(a)), and the placement of women in non-traditional jobs (Section 141(d)(2)). This study construes social value-added in terms of participant earnings impacts only. See Heckman, Heinrich, and Smith (1997, 2002) for discussions of the goals of JTPA and how they were expressed through the JTPA incentives.

[15]Because the training agency's performance standards were based on the past performance of many training agencies, no single agency could appreciably lower its future standards by suppressing performance. I.e., there was no ratchet problem in JTPA.

[16]The adjustment procedure is based on regression estimation of the determinants of performance outcomes, with training center characteristics as dependent variables. See Barnow (1992) or Johnston for a more complete description of this procedure.

[17]The incentive schemes in effect for a given year are set forth in policy letters on file in state governors' offices and available to the public. They document the exact performance measures in place as well as the terms under which agencies received incentive money for each program year.

[18]States added or subtracted a performance measure without changing the size of the potential award. Each state set aside (roughly) six percent of its Congressional JTPA allocation to be distributed among its training agencies as a financial award. Changing the kind and number of performance measures did not increase the size of the potential award a training agency won but simply changed the way the award pot was divided.

[19]The exogeneity assumption is discussed in section 6.3.

[20]Job Training Partnership Act, Sec. 202(a)(1)(3)(B).

[21]In some states agencies received additional awards for *exceeding* standards. In some states the performance-award relationship displayed many sorts of non-linearities, such as piece-wise linear forms and nested qualifying conditions. Because of space constraints, I do not describe these various functional forms but instead refer the interested reader to the descriptions in Courty and Marschke (2002).

[22]For some findings on the effect of incentives in other public sector contexts, see Asch (1990) on performance incentives facing Navy recruiters and Croxson, Propper, and Perkins (2001) on incentives offered to doctors in the U.K.'s NHS to reduce health care costs.

[23]Courty and Marschke (1997 and forthcoming), however, have provided evidence that training centers manipulate performance reporting in response to the incentives' thresholds and deadlines.

[24]Another set of papers on JTPA have attempted to "validate" the job training performance measures by correlating the performance measures in place in JTPA with various economic measures of training agency value-added (e.g., Heckman, Heinrich, and Smith, 2002, and Barnow, 2000.) For a survey and discussion of such studies on job training performance measures, see Heckman, Heinrich, and Smith.

[25]See Doolittle and Traeger (1990) for a description of the implementation of the National JTPA Study, and Orr et al., 1994 for a detailed description of the its results.

[26]For one quarter of the experimental participants, data were collected for an additional 18 months. This paper utilizes only the 18 month data.

[27]Burkhauser et al. and Cragg did not have access to such precise measures of outcomes.

[28]An appendix describing the construction and interpretation of the activity choice variable is available from the author upon request.

[29]To estimate a person's outcome in each of the choices, I use the resulting outcomes from the choices that persons actually make. Other researchers have used actual payoffs to estimate payoffs for choices not taken, then used the resulting estimated payoffs in a second stage to attempt to explain the choices in the first place. For example, Haveman, Wolfe, and Warlick (1988) used expected income flows in alternative work and public transfer program options to explain workers' choices over those options.

[30]Because the enrollee is observed only in the training activity selected by the agency, one has to be concerned with unobserved heterogeneity. If unobserved selectivity is present these estimators may yield inconsistent estimates. In the analysis for each performance outcome I test for and rule out the presence of selection. Because selection is based on more than two choices, one cannot use the procedure proposed by Heckman (1979). Instead, I use the procedure outlined by Lee (1983) to test for selectivity bias. This procedure involves constructing the inverse Mills ratios for the three choice case. The inverse Mills ratios are entered on the righthand-side of the performance outcome equations and their coefficients are estimated along with the other coefficients of the models. (The employment equations are estimated as linear probability models for the purposes of this test.) In no case did I find that the coefficient estimate on the inverse Mills ratio approached statistical significance by conventional significance criteria. These results are available upon request. The absence of evidence of unobserved selectivity is not surprising given the richness of the each enrollee's description contained in the data. I therefore proceed with the estimation and argue that these estimates are consistent.

[31]Appendix 1 describes further details of the estimation of the predicted outcomes and earnings impacts that serve as the regressors in the second stage estimation. The first stage regressions' right-hand-side variables, the characteristics of enrollees and of the local labor

market, are described in Table A.1.

[32]Appendix 2 describes in detail the likelihood function of the second stage estimation.

[33]The program I use to estimate this model was developed under NICHD Grant R01-HD19226 to James Heckman. Stephen Cameron and George Yates wrote the software. I thank James Heckman for making this program available to me. Notes by Cameron and Heckman (1989) describe this program and the estimation of a general version of the model presented here.

[34]The test (z) statistics are corrected for the estimated regressors using the delta method. See Appendix 1 for an explanation.

[35]The likelihood ratio index reported is defined by

$$\rho^2 = 1 - \frac{\text{log-likelihood at maximum}}{\text{log-likelihood at } \boldsymbol{\pi} = \mathbf{0}},$$

which lies in the $[0, 1]$ interval and higher values correspond to closer fit. The likelihood ratio statistic is also reported.

[36]In selecting these measures, I have tried to follow the pronouncements of the Department of Labor, Congress, and the states concerning special service to the "hard to serve." Hard to serve was a loosely-defined term that described persons among the JTPA-eligible who had extraordinary difficulty finding employment, either because they had few skills, or some "stigma" that made them poor employment risks (see Barnow and Constantine, 1988). The persons most commonly described as stigmatized in this way included welfare recipients and high school dropouts. Case workers likely also used earnings in the year prior to application (or some similar measure of recent labor market history) as an indicator of the labor market skills the applicant brought to JTPA.

[37]The estimates of the utility weights that produced these differences are presented in Table A.2 in the appendix. As Table A.2 shows, the utility weight estimates for several outcomes under incentive regimes that did not reward them sometimes are positive and significant. Nevertheless, the estimated utility weights for several outcomes under incentive regimes that rewarded them sometimes have negative signs. That estimated utility weights appear with surprising signs suggests that predicted outcomes in training activities are often correlated with other determinants of choice that I cannot measure. By looking at how choices change under different incentive regimes, however, we can detect the influence of performance incentives. By comparing behavior across regimes we difference out the fixed, non-award determinants of training choice that cannot be controlled for.

[38]The reader may wonder about the magnitude of the pecuniary incentive these agencies are responding to. Although the marginal valuation of exceeding a standard varied by state and training agency, we can illustrate its magnitude by considering the training agency in Northwest, Minnesota. As Table 2 shows, the Northwest training agency faced the average

weeks worked at follow-up ($AWWF$) measure for the first time in 1988. In exceeding this standard in 1988, the Northwest training agency earned an additional $3,685 in training resources for 1989. Note that its budget for 1989 for adults was $919,590. Taking Minnesota's incentive policy as representative, meeting or failing to meet a performance standard concerned something less than one percent of the training agency's budget.

[39]At least two sources of unobserved agency heterogeneity may be confounded with the changes in incentives. First, the study lacks agency-specific measures of training costs. Thus, two agencies with identical incentive schemes may have been driven by differences in the relative input prices alone to choose different inputs. Second, important non-incentive related objectives which I am unable to measure and thus account for in my empirical model undoubtedly drove some of the behavior I am trying to explain. Possible preferences that I cannot control for include preferences for service providers with political ties to the local elected officials. (Westat, Inc (1985) and MDC, Inc and Grinker, Walker and Associates (1984,1985) discuss the objectives of agencies and the influence of area politics on training decisions.) The Act gave the locally elected executive politician (usually a city mayor or county president in the city or county in which the agency is located) important authority in the conduct of training operations. Such preferences, for example, may have led the agency to favor a classroom training program operated by a large, influential community-based organization, or to reward subsidized on-the-job training slots to local businesses that have contributed to political campaigns. Thus, in the analysis depicted in Table 5, instead of picking up incentive effects, in some cases I may be picking up differences in political climate.

[40]Another reason for confidence that the results reported in Table 5 are responses to changes in incentives and do not suffer from a cross-sectional heterogeneity bias is by conducting separate event studies for each agency that faced a change in incentives. For each agency a separate model—similar to the model that generates the results reported in Table 5—has been estimated where the data are the persons randomly assigned in the two years straddling the change. Four of the largest agencies in the study—Fort Wayne, Indiana; Jersey City, New Jersey; and Springfield, Missouri—generated statistically significant shifts in training strategies coinciding with the activation of the welfare employment rate, the cost measure, the employment rate at follow-up, and the weeks worked at follow-up. The results of the event studies, which are not susceptible to unobserved agency heterogeneity, support the pooled analysis. These results are available from the author upon request.

[41]These results are available to the interested reader upon request.

[42]Each simulation is based on probabilities constructed for each observation from 500 draws from a multivariate normal, the observations characteristics, and the parameter estimates from Model 4 in Table 5.

[43]The *net* impact estimates I present are low in comparison to the estimates from the National JTPA Study. The *impact* estimates, however, are similar. Using roughly the same data Orr et al. estimate an average per-capita earnings impact of $877 (computed from their table 4.6, p. 87). The differences between my estimates of net impacts arise from the

cost calculations. Orr et al. derive daily cost estimates for each activity from expenditure data obtained from the sixteen experimental sites covering all training dispensed in 1987. Not having access to this data, this study instead borrows from an analysis of a set of training contracts from a suburban Illinois agency with external training vendors (Heinrich, Marschke, and Zhang). These data allow estimation of costs by demographic characteristics, such as race, gender, and age. Orr *et al.* estimated that the activities classroom training, on-the-job training, and job search assistance cost $1341, $1319, and $585, respectively. The estimates for these activities used here, under the actual JTPA training assignments, are $1956, $2064, and $1582, respectively. Clearly, using a different set of cost estimates would change the magnitudes of this study's outcomes. Nonetheless, because the relative magnitude of training costs between the studies are similar, using their cost estimates would not likely change the inferences seen in Table 5.

[44]The predicted total expenditure under the actual training assignments is $11,508,823. Thus predicted outlays under any of the simulated incentive regimes are possible without substantial additional funds. (The total predicted earnings impact under actual training assignments is $6,610,389.)

[45]The finding that the follow-up measures produced more effective job training, however, suggests that incentive-designers learn about the distortions caused by performance measures and attempt to replace them with less-distorted measures.

[46]Indeed, Heckman, Heinrich, and Smith and Barnow (2000) show little statistical correlation between JTPA performance measures and measures of social product.

References

[1] Abowd, J. M. (1990). Does Performance-Based Managerial Compensation Affect Corporate Performance? *Industrial and Labor Relations Review*, 43:52S–73S.

[2] Anderson, K., Burkhauser, R. V., and Raymond, J. E. (1993). The Effect of Creaming on Placement Rates under the Job Training Partnership Act. *Industrial and Labor Relations Review*, 46(4):613–24.

[3] Asch, B. J. (1990). Do Incentives Matter? The Case of Navy Recruiters. *Industrial and Labor Relations Review*, 43:89S–106S.

[4] Baker, G. P. (1992). Incentive contracts and performance measurement. *Journal of Political Economy*, 100(3):598–614.

[5] Baker, G. P. (2002). Distortion and risk in optimal incentive contracts. *Journal of Human Resources*, 38(4).

[6] Baker, G. P., Jensen, M. C., and Murphy, K. J. (1988). Compensation and Incentives: Practice vs. Theory. *Journal of Finance*, 43:593–616.

[7] Barnow, B. (1992). The Effects of Performance Standards on State and Local Programs. In Manski, C. and Garfinkel, I., editors, *Evaluating Welfare and Training Programs*. Harvard University Press.

[8] Barnow, B. (1993). Thirty Years of Changing Federal, State, and Local Relationships in Employment and Training Programs. *Publius*, 23(3):75–94.

[9] Barnow, B. (2000). Exploring the Relationship between Performance Management and Program Impact. *Journal of Policy Analysis and Management*, 19(1):118–41.

[10] Barnow, B. S. and Constantine, J. (1988). *Using Performance Management to Encourage Services to Hard-to-Serve Individuals in JTPA*. Washington, D.C.: National Commission for Employment Policy.

[11] Bloom, H. S. (1984). Accounting for No-Shows in Experimental Evaluation Designs. *Evaluation Review*, 8:225–25.

[12] Bruns, W. (1992). Introduction. In Bruns, W., editor, *Performance Measurement, Evaluation, and Incentives*. Harvard Business School Press, Boston.

[13] Cameron, S. V. and Heckman, J. J. (1989). Son of CTM: The DCPA Approach Based on Discrete Factor Structure Models. Manuscript, University of Chicago.

[14] Courty, P. and Marschke, G. (2002). Implementing Performance Measurement and Performance Funding: Lessons from a Federal Job Training Program. Manuscript, University at Albany.

[15] Courty, P. and Marschke, G. (Forthcoming). An Empirical Investigation of Gaming Responses to Performance Incentives. *Journal of Labor Economics*.

[16] Cragg, M. (1997). Performance Incentives in the Public Sector: Evidence from the Job Training Partnership Act. *Journal of Law, Economics and Organization*, 13(1):147–168.

[17] Croxson, B., Propper, C., and Perkins, A. (2001). Do Doctors Respond to Financial Incentives? UK Family Doctors and the GP Fundholder Scheme. *Journal of Public Economics*, 79:375–398.

[18] Dickinson, K., West, R., Kogan, D., Drury, D., Franks, M., Schlichtmann, L., , and Vencill, M. (September, 1988). *Evaluation of the Effects of JTPA Performance Standards on Clients, Services, and Costs*. Washington, D.C.: National Commission for Employment Policy, Research Report No 88-16.

[19] Dixit, A. (2002). Incentives and Organizations in the Public Sector: An Interpretive Review. *Journal of Human Resources*, 38(4).

[20] Doolittle, F. and Traeger, L. (1990). *Implementing the National JTPA Study*. Manpower Demonstration Research Corporation, New York.

[21] Garen, J. (1994). Executive Compensation and Principal-Agent Theory. *Journal of Political Economy*, 102:1175–99.

[22] Gibbons, R. (1998). Incentives and Careers in Organizations. *Journal of Economic Perspectives*.

[23] Gore, A. (1993). From Red Tape to Results: Creating a Government that Works Better and Costs Less. Vice President Albert Gore's National Performance Review.

[24] Haubrich, J. (1994). Risk Aversion, Performance Pay, and the Principal-Agent Problem. *Journal of Political Economy*, 102:258–76.

[25] Haveman, R., Wolfe, B., and Warlick, J. (1988). Labor Market Behavior of Older Men. *Journal of Public Economics*, 36(2):153–175.

[26] Heckman, J. J. (1979). Sample Selection as a Specification Error. *Econometrica*, 47:153–61.

[27] Heckman, J. J., Heinrich, C., and Smith, J. A. (2002). The Performance of Performance Standards. *Journal of Human Resources*, 37(4):778–811.

[28] Heckman, J. J., Heinrich, C., and Smith, J. A. (May, 1997). Assessing the The Performance of Performance Standards in Public Bureaucracies. *American Economic Review*, 87(2).

[29] Heckman, J. J. and Smith, J. A. (1995). The Determinants of Participation in a Social Program: Evidence from the Job Training Partnership Act. Manuscript, The University of Chicago.

[30] Heckman, J. J., Smith, J. A., and Taber, C. (1996). What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program. In Libecap, G., editor, *Advances in the Study of Entrepreneurship, Innovation and Economic Growth, Volume 7, Reinventing Government and the Problem of Bureaucracy*, pages 191–218. Greenwich, CT: JAI Press.

[31] Heinrich, C. (July, 1995). A Case Study of the Participant Selection and Service Assignment Processes and the Intervening Influences of Performance Standards and Other Administrative Practices. Manuscript, The University of Chicago.

[32] Heinrich, C., Marschke, G. R., and Zhang, A. (1998). Measuring the Costs of Training Under JTPA. Manuscript, The University of Chicago.

[33] Holmstrom, B. and Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, 7(0):24–52.

[34] Jensen, M. and Murphy, K. J. (1990). Performance Pay and Top Management Incentives. *Journal of Political Economy*, 98(2):225–64.

[35] Johnston, J. W. (1987). *The Job Training Partnership Act: A Report by the National Commission for Employment Policy*. Washington D.C.: U.S. Government Printing Office.

[36] Kahn, L. and Sherer, P. (1990). Contingent Pay and Managerial Performance. *Industrial and Labor Relations Review*, 43:107S–120S.

[37] Lazear, E. (2000). Performance Pay and Productivity. *American Economic Review*, 90(5):1346–61.

[38] Lee, L.-f. (1983). Generalized econometric models with selectivity. *Econometrica*, 51:507–12.

[39] Lipsky, M. (1980). *Street Level Bureaucracy: Dilemmas of the Individual in Public Services*. Russell Sage Foundation.

[40] MDC, Inc./Grinker, Walker & Associates (January, 1985). An Independent Sector Assessment of the Job Training Partnership Act, Phase II: The Initial Implementation. Manuscript, Report to the National Commissionfor Employment Policy, Washington, D.C.

[41] MDC, Inc./Grinker, Walker & Associates (March, 1984). An Independent Sector Assessment of the Job Training Partnership Act, Phase I: The Initial Implementation. Manuscript, Report to the National Commissionfor Employment Policy, Washington, D.C.

[42] Newey, W. K. and McFadden, D. L. (1994). Estimation and Hypothesis Testing in Large Samples. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, volume 4, chapter 2. Amsterdam; London and New York: Elsevier, North-Holland.

[43] Olson, M. (1965). *The Logic of Collective Action*. Cambridge: Harvard University Press.

[44] Orr, L. L., Bloom, H. S., Bell, S. H., Lin, W., Cave, G., and Doolittle, F. (March 1994). *The National JTPA Study: Impacts, Benefits, and Costs of Title II-A*. Abt Associates Inc.

[45] Paarsch, H. and Shearer, B. (1996). Fixed Wages, Piece Rates, and Incentive Effects. Manuscript, University of Laval.

[46] Prendergast, C. (March, 1999). The Provision of Incentives in Firms. *Journal of Economic Literature*, 37(1).

[47] Westat, Inc. (November, 1985). Implementation of the Job Training Partnership Act: Final Report. Manuscript, Report to the U.S. Department of Labor, Washington, D.C.

TABLE 1

JTPA Performance Measures in Effect in Years 1987-1989

Performance Measure	Name	Definition
ERT	Employment Rate at Termination	Fraction of terminees employed at termination
WERT	Welfare Employment Rate at Termination	Fraction of terminees receiving welfare at date of application who were employed at termination
AWT	Average Wage at Termination	Average wage at termination for terminees who were employed at termination
CE	Cost per Employment	Training center's year's expenditures on adults divided by the number of adults employed at termination
ERF	Employment Rate at Follow-up	Fraction of terminees who were employed at 13 weeks after termination
WERF	Welfare Employment Rate at Follow-up	Fraction of terminees receiving welfare at date of application who were employed at 13 weeks after termination
AWEF	Average Weekly Earnings at Follow-up	Average weekly wage of terminees who were employed 13 weeks after termination
AWWF	Average Weeks Worked by Follow-up	Average number of weeks worked by terminees in 13 weeks following termination

Notes:

1. The date of termination is the date the enrollee officially exits training. A terminee is an enrollee after he has officially exited training.

2. 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. The follow-up period was the 13 weeks following termination. Likewise, persons who terminated in 1986 were not included in the 1987 measure, even if their follow-up period extended into 1987.

TABLE 2
Performance Measures
By Training Agency and Year

Training Agency	Year	ERT	AWT	CE	WERT	ERF	AWWF	WERF	AWEF
Corpus Christi, TX	1987	×	×	×	×	×			
	1988	×	×	×	×	×			
	1989†	×	×	×	×	×		×	
Cedar Rapids, IA	1987	×	×	×	×				
	1988	×	×	×	×	×			
	1989†	×	×		×	×		×	×
Coosa Valley, GA	1987	×	×	×	×	×			×
	1988	×	×	×	×	×		×	×
	1989†	×	×	×	×	×		×	×
Heartland, FL	1987	×	×	×	×				
	1988	×	×		×	×		×	
	1989†	×	×		×	×		×	
Fort Wayne, IN	1987	×	×	×	×				
	1988	×	×	×	×	×			
	1989	×	×	×	×	×			
Jersey City, NJ	1987	×	×	×	×	×			
	1988	×	×	×		×		×	×
	1989	×	×	×		×		×	×
Jackson, MS	1987	×	×	×	×				
	1988	×	×	×	×				
	1989†	×	×	×	×				
Larimer, CO	1987	×	×	×	×				
	1988	×	×	×	×	×	×		
	1989	×		×		×	×	×	×
Decatur, IL	1987	×	×	×	×				
	1988	×	×	×	×	×			
	1989†	×	×	×	×	×			
Northwest, MN	1987†	×	×	×	×				
	1988	×	×	×	×		×		
	1989†	×	×	×	×		×		
Butte, MT	1987†	×	×	×	×				
	1988	×	×	×	×	×	×		×
	1989	×	×	×	×	×	×		×
Omaha, NE	1987†	×	×	×	×				
	1988	×	×		×	×	×		×
	1989	×	×		×	×	×	×	×
Marion Co., OH	1987	×	×	×	×				
	1988	×	×	×	×				
	1989	×	×	×	×	×		×	×
Oakland, CA	1987†	×	×	×	×				
	1988	×	×	×	×	×	×		
	1989†	×	×	×	×	×	×		
Providence, RI	1987	×	×	×	×				
	1988	×	×	×	×				
	1989	×	×	×	×				
Springfield, MO	1987	×	×	×	×				
	1988	×	×	×	×	×			
	1989†	×	×			×	×	×	×

Notes:

† indicates a year at a training center for which there are zero adult random assignments.

1. Source: state incentive policies. See Table 1 for definitions of performance measures.

TABLE 3
Performance Outcomes of JTPA Enrollees
By Training Assignment
(Based on adult men and women in treatment group who enrolled)

	Performance Outcome					
Training Assignment	Employment at Termination (ERT)	Employment at Follow-up (ERF)	Wage at Termination (AWT)	Weeks Worked Between Termination and Follow-up (AWWF)	Weekly Earnings at Follow-up (AWEF)	Obs.
All	0.58	0.58	5.75	8.16	240.70	6695
Classroom	0.53	0.53	5.82	7.55	235.48	2642
On-the-job	0.67	0.66	5.59	9.09	245.06	2356
Job Search	0.51	0.53	5.92	7.66	241.27	1697

See Table 1 for definition of performance measures.

TABLE 4
Training Assignments By Participant Subgroup
(Based on adult men and women in treatment and control groups)

Characteristic	Fraction in Classroom Occup. Training	On-the-job Training	Job Search Assistance	Obs.
All	0.34	0.39	0.27	14908
Age 22-29	0.36	0.39	0.25	6539
Age 30-39	0.34	0.37	0.28	5227
Age 40-49	0.28	0.43	0.28	2015
Age 50-54	0.25	0.48	0.28	483
Age > 54	0.26	0.36	0.38	644
Male	0.23	0.46	0.30	6852
Female	0.42	0.33	0.25	8056
White	0.29	0.48	0.23	8181
Black	0.39	0.29	0.32	4591
Hispanic	0.39	0.29	0.32	1597
Asian, Pacific Islander, or Filipino	0.33	0.39	0.28	197
Other race	0.46	0.23	0.32	342
Highest grade completed < 10	0.28	0.43	0.29	2447
Highest grade completed 10-11	0.36	0.39	0.25	3069
Highest grade completed 12	0.36	0.39	0.25	5660
Highest grade completed 13-15	0.34	0.39	0.27	1945
Highest grade completed > 15	0.19	0.36	0.45	510
Married	0.30	0.46	0.24	3735
Divorced, widowed, or separated	0.35	0.39	0.27	4941
Single	0.34	0.37	0.29	5194
No children under age 6 at home	0.30	0.41	0.29	9571
Own children under age 6 at home	0.42	0.35	0.23	4078
Not receiving General Assistance	0.34	0.39	0.27	12419
Receiving General Assistance	0.29	0.45	0.26	1041
Not receiving Food Stamps	0.29	0.41	0.30	8538
Receiving Food Stamps	0.41	0.37	0.23	5740
Not receiving AFDC	0.29	0.43	0.29	11644
Receiving AFDC	0.52	0.27	0.21	3059
Currently employed	0.35	0.41	0.24	2081
Last employed 0-2 months ago	0.27	0.44	0.28	2900
Last employed 3-5 months ago	0.31	0.43	0.26	1970
Last employed 6-8 months ago	0.29	0.43	0.27	1171
Last employed 9-11 months ago	0.33	0.40	0.27	619
Last employed ≥ 12 months ago	0.35	0.38	0.27	2335
Never employed	0.46	0.24	0.30	1792
Last year earnings < $3K	0.36	0.37	0.27	7897
Last year earnings $3K-6K	0.32	0.45	0.23	2194
Last year earnings $6K-9K	0.29	0.46	0.25	1141
Last year earnings $9K-12K	0.27	0.43	0.30	603
Last year earnings $12K-15K	0.29	0.39	0.32	268
Last year earnings > $15K	0.22	0.38	0.40	255

See Table A.1 for definition of variables.

TABLE 5
Determinants of Training Choice
Probits (|z stat.| in parentheses)
Observations = 13338

A. *Non-award determinants of choice*

	Model			
Variable	(1)	(2)	(3)	(4)
Earnings impact ($1000)	.028	.140	.004	.005
	(2.118)	(.023)	(.001)	(.001)
Cost ($1000)	-.354	-.366	-.328	-.333
	(22.039)	(18.825)	(16.344)	(15.893)
Variables constant across choices				
OJT choice				
Constant	.071	.308	.243	.237
	(3.910)	(2.803)	(2.401)	(2.484)
Previous year's earnings < $3K		-.116	-.095	-.098
		(2.803)	(2.401)	(2.484)
High school dropout		.164	.291	.293
		(4.419	(8.020)	(8.033)
Welfare recipient		-.502	-.374	-.380
		(12.869)	(9.308)	(8.833)
JSA choice				
Constant	-.390	-.432	-.355	-.374
	(20.574)	(11.493)	(9.095)	(9.316)
Previous year's earnings < $3K		.029	.103	.098
		(.711)	(2.504)	(2.371)
High school dropout		.099	.097	.106
		(2.675)	(2.504)	(2.799)
Welfare recipient		-.042	-.349	-.336
		(.934)	(7.273)	(6.978)

(Continued)

TABLE 5 (Continued)

B. *Impact of activating performance measure on outcome utility weight*

Activated measure	Performance outcome	Predicted impact	Estimated impact on outcome utility weight $(\pi_A - \pi_N)$ (1)	(2)	(3)	(4)
ERF	*ERF*	+			1.082	1.365
					(4.126)	(4.206)
AWT	*AWT*	+			-.181	-.161
					(1.407)	(1.309)
AWEF	*AWEF*	+			-.112	-.063
					(1.557)	(.872)
AWWF	*AWWF*	+			.154	.176
					(6.052)	(7.039)
CE	*CE*	-			.018	.048
					(.273)	(.714)
WERT	*WERT*†*	+				.351
						(.885)
WERF	*WERF**	+				.120
						(.201)
CE	*ERT*	+				1.315
						(2.323)
CE	*WERT**	+				3.003
						(3.319)

	(1)	(2)	(3)	(4)
Log likelihood	-14196.847	-14105.278	-13870.513	-13619.277
Likelihood ratio index (ρ^2)	.031	.044	.069	.070
Likelihood ratio statistic	912.89	1296.94	2020.82	2055.92

Notes:

1. ''Estimated impacts on outcome utility weight" are the differences in the estimated utility weights reported in Table A.2. For example, the estimated impact of activating *ERF* in Model 3 (first row above) is 1.082 because the estimate of the utility weight on the predicted *ERF* rises from -.518 to .564 with the activation of the *ERF* measure. The construction of the remaining coefficients are explained in Table A.2.

2. |z statistics| based on asymptotic standard errors in parentheses. Standard errors are corrected for the estimated regressors.

* Based on coefficients estimated off of welfare recipients only.

† Based on coefficients estimated off of persons assigned with the *CE* performance measure in effect. The data allow this impact to be estimated off of persons assigned when the *CE* performance measure is not in effect. Either way the impact is computed, the inference is the same.

TABLE 6
Estimated Elasticities
$\frac{\partial \ln \Pr(\text{Training choice} = \tau)}{\partial \ln Z_\tau}$

Z_τ	Classroom Training	On-the-job Training	Job Search Assistance
Earnings impact	.021	.021	.027
Cost	-.501	-.473	-.444

Computed from model 4's (Table 5) coefficient estimates by averaging effects over sample.

TABLE 7
Estimated Effects of Low Earnings, Low Education, and Welfare Recipiency on the Relative Probability of Assignment to Classroom Training

Low earnings	$\frac{\Pr(\tau=OJT\|ERN\leq\$3K)}{\Pr(\tau=CTO\|ERN\leq\$3K)} / \frac{\Pr(\tau=OJT\|\overline{ERN\leq\$3K})}{\Pr(\tau=CTO\|\overline{ERN\leq\$3K})}$	=	.885
	$\frac{\Pr(\tau=JSA\|ERN\leq\$3K)}{\Pr(\tau=CTO\|ERN\leq\$3K)} / \frac{\Pr(\tau=JSA\|\overline{ERN\leq\$3K})}{\Pr(\tau=CTO\|\overline{ERN\leq\$3K})}$	=	1.007
High School Dropout	$\frac{\Pr(\tau=OJT\|DROP)}{\Pr(\tau=CTO\|DROP)} / \frac{\Pr(\tau=OJT\|\overline{DROP})}{\Pr(\tau=CTO\|\overline{DROP})}$	=	1.455
	$\frac{\Pr(\tau=JSA\|DROP)}{\Pr(\tau=CTO\|DROP)} / \frac{\Pr(\tau=JSA\|\overline{DROP})}{\Pr(\tau=CTO\|\overline{DROP})}$	=	1.148
Welfare Recipient	$\frac{\Pr(\tau=OJT\|WELF)}{\Pr(\tau=CTO\|WELF)} / \frac{\Pr(\tau=OJT\|\overline{WELF})}{\Pr(\tau=CTO\|\overline{WELF})}$	=	.487
	$\frac{\Pr(\tau=JSA\|WELF)}{\Pr(\tau=CTO\|WELF)} / \frac{\Pr(\tau=JSA\|\overline{WELF})}{\Pr(\tau=CTO\|\overline{WELF})}$	=	.558

Notes:
$ERN \leq \$3K$ ($ERN \leq \$3K$) indicates the enrollee earned less than or equal to (greater than) \$3,000 in the year prior to enrollment; $DROP$ ($\overline{DROP}$) indicates the enrollee is (is not) a high school dropout; and $WELF$ ($\overline{WELF}$) indicates the enrollee was (was not) a welfare recipient at the time of enrollment.

TABLE 8

Simulating the Effects of Incentive Regime Changes on
Social Net Benefit and Training Activity Choice
(The base regime contains the employment rate at termination (*ERT*),
the average wage at termination (*AWT*), and the cost per employment (*CE*))

		Base Regime		
Outcome	Base Alone (ERT, AWT, and CE) (1)	− Cost per Employment (CE) (2)	+ Ave. Follow-up Weeks Worked (AWWF) (3)	+ Employment Rate at Follow-up (ERF) (4)
Efficiency:				
Total earnings impact ($1000)	6,051	5,081	7,511	7,194
per enrollee	.999	.839	1.240	1.188
Total cost ($1000)	11,478	11,349	11,515	11,529
per enrollee	1.895	1.873	1.904	1.901
Change in per enrollee net impact relative to Base Regime ($1000)		−.139	+.232	+.183
Performance outcomes:				
CE ($1000/employed enrollee)	3.308	3.344		
AWWF (wks/enrollee)	7.95		8.209	
ERF (fraction)	.573			.586
Assigned activity:				
Classroom	.364	.371	.353	.349
OJT	.368	.340	.397	.386
Job Search	.268	.289	.250	.265

Notes:

1. 6056 enrollees used in simulation. Earnings impacts are cumulated, undiscounted earnings impacts over the 18 months following enrollment.

2. The simulation uses the estimates from Model 4 in Table 5.

3. See Table 1 for definitions of performance outcomes and text for definitions of training.

Appendix 1: Constructing measures of expected outcomes and earnings impacts

Performance outcomes

Left-hand side variables. I ran separate regressions for each performance outcome, using all participants in the experiment who were randomized into the program and had valid enrollment and termination dates. The NJS data file contains participant-reported monthly data on employment in the 18 months following random assignment. Participants in the study supplied separate wage and employment information for each job held in the period after random assignment. The data file also contains enrollment and termination dates from agency records. I constructed all termination-based performance outcomes using the enrollee's reported employment status and wage in the month that the agency reported his/her termination. I constructed all follow-up date-based performance outcomes using the enrollee's reported employment hours and wage information from the calendar month containing the enrollment date through the calendar month containing the follow-up date (the follow-up date occurs ninety days after the termination date).

I ran the wage at termination and earnings at follow-up regressions using enrollees who reported being employed in the month of termination and the month of follow-up, respectively. I constructed the weekly earnings at follow-up from the total earnings reported for all job spells occurring between the months of termination and follow-up, inclusive.[47] For enrollees employed by more than one employer at termination, I computed the wage at termination from the spell with the highest wage.[48]

Right-hand side variables. The agency conditions its expectations of outcomes on the enrollee's attributes and the state of the labor market. Thus, in addition to the enrollee's characteristics, the right hand side of each regression contains the local unemployment rate at the month of random assignment, the fraction of the local population employed in the year of random assignment, and the fraction of the local population working in the service sector in the year of random assignment. Enrollee characteristics used in these regressions are the same as those listed in Table 4 (Table A.1 defines the variables).[49] These variables enter each regression alone and interact with two training assignment dummies (for on-the-job training and job search assistance—classroom training is the omitted training assignment).[50]

To estimate the relationship between an activity choice, τ, and the performance outcomes, I estimate five models separately on the same data. The dependent variables in these models were employment status at termination (ERT) and follow-up (ERF), the wage at termination (AWT), the number of weeks worked in the ninety days following termination ($AWWF$), and labor market earnings in the ninety days following termination ($AWEF$). The kth model is $S_i^k = F(\boldsymbol{X}_i\boldsymbol{\gamma}^k + \boldsymbol{X}_i\boldsymbol{\gamma}_{OJT}^k + \boldsymbol{X}_i\boldsymbol{\gamma}_{JSA}^k)$. The set of five measures include

both continuous measures, like wages and earnings, and binary ones, like employment status at termination. For the continuous regressors, F is linear and $\boldsymbol{\gamma}^k$, $\boldsymbol{\gamma}^k_{OJT}$, and $\boldsymbol{\gamma}^k_{JSA}$, are estimated using least squares. For the binary regressors, $F(\cdot)$ is the logistic distribution, and the coefficients are estimated using maximum likelihood. From $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\gamma}}_{OJT}$, and $\hat{\boldsymbol{\gamma}}_{JSA}$, and individual characteristics, predicted performance outcomes for each measure under each activity are formulated for each individual.

Earnings impact

Both treatments and controls are used to estimate earnings impacts. To produce activity-specific estimates of earnings impact, the following regression model is estimated.

$$Y_i = \delta_0 + T_i \boldsymbol{X}_i \boldsymbol{\delta} + T_i \, d_{iOJT} \, \boldsymbol{X}_i \boldsymbol{\delta}_{OJT} + T_i \, d_{iJSA} \, \boldsymbol{X}_i \boldsymbol{\delta}_{JSA} + \nu_i \tag{7}$$

where Y_i is the the cumulated labor market earnings for the 18 months after random assignment; T_i is an indicator variable equal to one if i is assigned to training; d_{iOJT} is an indicator variable equal to one if i is assigned to on-the-job training; d_{iJSA} is an indicator variable equal to one if i is assigned to job search assistance; $\boldsymbol{X}_i$ is a vector of individual, training and labor market characteristics; and ν_i is a normal i.i.d. error term. (7) is estimated using least squares.[51]

Conditional on i's characteristics, the total impact of training on the earnings (Δ) is equal to (suppressing the i subscript)

$$\begin{aligned} E(\Delta|\ \boldsymbol{X}, d) &= E(Y|T=1,\ \boldsymbol{X}, d) - E(Y|T=0,\ \boldsymbol{X}) \\ &= \boldsymbol{X}\boldsymbol{\delta} + d_{OJT}\ \ \boldsymbol{X}\boldsymbol{\delta}_{OJT} + d_{JSA}\ \ \boldsymbol{X}\boldsymbol{\delta}_{JSA} \end{aligned}$$

Conditional on i's characteristics, estimates of the total impact of training on the earnings for classroom training, on-the-job training, and job search assistance are $\boldsymbol{X}\hat{\boldsymbol{\delta}}$, $\boldsymbol{X}\hat{\boldsymbol{\delta}} + \boldsymbol{X}\hat{\boldsymbol{\delta}}_{OJT}$, and $\boldsymbol{X}\hat{\boldsymbol{\delta}} + \boldsymbol{X}\hat{\boldsymbol{\delta}}_{JSA}$, respectively.

The results of the first stage estimates are available from the author upon request.

Appendix 2: Likelihood function for second stage

For each enrollee, the agency chooses activity that yields the highest net utility, where the net utility of a choice τ—there are three possible training choices here—is (equation 2 from the text)

$$U_{i\tau} = \theta_{\tau 0} + \boldsymbol{R}_i \boldsymbol{\theta}_\tau + \tilde{\Delta}_{i\tau}\theta_\Delta + \tilde{\boldsymbol{S}}_{i\tau}\boldsymbol{D}_i\boldsymbol{\pi}_A + \tilde{\boldsymbol{S}}_{i\tau}(\boldsymbol{I} - \boldsymbol{D}_i)\boldsymbol{\pi}_N + u_{i\tau},\ \tau = 1, 2, \text{ and } 3, \tag{8}$$

where all variables are defined as in the text. Because we do not observe outcomes in choices not taken, $\tilde{\Delta}_{i\tau}$ and $\tilde{\boldsymbol{S}}_{i\tau}$ are estimated. They are replaced by $\hat{\Delta}_{i\tau}$ and $\hat{\boldsymbol{S}}_{i\tau}$. The kth element of

$\hat{\boldsymbol{S}}_{i\tau}$ $\hat{S}^k_{i,\tau}$, is an estimate of the agency's expectation of outcome k for individual i in training activity τ. It is a function of the estimated vector of parameters, $\hat{\boldsymbol{\gamma}}^k_\tau$ computed from the first stage estimation. Note that $k = 1, ..., 5$ and predicted outcomes are estimated for $AWEF$, AWT, $AWWF$, ERF, and ERT. $\hat{\Delta}_{i\tau}$ is a function of a parameter vector, $\hat{\delta}_\tau$.

The optimal training choice is the choice that maximizes net utility

$$\hat{\tau}_i = \underset{\tau}{argmax}\{U_{i\tau}\}^3_{\tau=1}. \tag{9}$$

Let $\boldsymbol{Z}_{i\tau} = \begin{bmatrix} 1 \\ \boldsymbol{R}_i \\ \tilde{\Delta}_{i\tau} \\ \tilde{\boldsymbol{S}}_{i\tau}\boldsymbol{D}_i \\ \tilde{\boldsymbol{S}}_{i\tau}(\boldsymbol{I} - \boldsymbol{D}_i) \end{bmatrix}$, and $\boldsymbol{\beta}_\tau = \begin{bmatrix} \theta_{\tau 0} \\ \boldsymbol{\theta}_\tau \\ \theta_\Delta \\ \boldsymbol{\pi}_A \\ \boldsymbol{\pi}_N \end{bmatrix}$. The probability that activity 1 will be chosen, conditional on $\boldsymbol{Z}_i$ is

$$\begin{aligned}
P_{i,1}(\boldsymbol{Z}_i) &= Pr\{1 = \underset{\tau=1,2,3}{argmax}\, U_{i\tau} \mid \boldsymbol{Z}_i\} \\
&= Pr(u_{i2} < \boldsymbol{Z}_{i1}\boldsymbol{\beta}_1 - \boldsymbol{Z}_{i2}\boldsymbol{\beta}_2 + u_{i1}) \times \\
&\quad Pr(u_{i3} < \boldsymbol{Z}_{i1}\boldsymbol{\beta}_1 - \boldsymbol{Z}_{i3}\boldsymbol{\beta}_3 + u_{i1}) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\left(\frac{(Z_{i1}\beta_1 - Z_{i2}\beta_2) + \sigma_{u_1}\epsilon_1}{\sigma_{u_2}}\right)} \phi(0,1)d\epsilon_2 \int_{-\infty}^{\left(\frac{(Z_{i1}\beta_1 - Z_{i3}\beta_3) + \sigma_{u_1}\epsilon_1}{\sigma_{u_3}}\right)} \phi(0,1)d\epsilon_3 \phi(0,1)d\epsilon_1 \\
&= \int_{-\infty}^{\infty} \Phi\left(\frac{(Z_{i1}\beta_1 - Z_{i2}\beta_2) + \sigma_{u_1}\epsilon_1}{\sigma_{u_2}}; 0, 1\right) \Phi\left(\frac{(Z_{i1}\beta_1 - Z_{i3}\beta_3) + \sigma_{u_1}\epsilon_1}{\sigma_{u_3}}; 0, 1\right) \phi(0,1)d\epsilon_1
\end{aligned}$$

I have assumed that u_{i1}, u_{i2}, and u_{i3} are independent draws from normal distributions with zero mean and variance σ_{u_τ}, $\tau = 1, 2$, and 3. Thus the above model is a multinomial probit and $\phi(a_1, a_2)$ and $\Phi(a_3; a_1, a_2)$ are the density and distribution functions for the normal and a_1, a_2, and a_3 are the associated mean, variance, and limit of integration. The probabilities of choosing activity 2 and 3 are calculated similarly. The contribution of the ith individual to the single likelihood is

$$\mathcal{L}_i = \sum_{\tau}^{3} P_{i\tau}(\mathbf{Z}_i)d_{i\tau},$$

where $d_{i\tau} = 1$ if τ is chosen, and zero otherwise. Assuming independent sampling across persons, the full likelihood is obtained by producting $\mathcal{L}_i$ across persons. I make the following normalizations to achieve identification: $\sigma^2_{u_1} = \sigma^2_{u_2} = \sigma^2_{u_3} =$ constant (1, in this case), and $\theta_{\tau 0} = 0$ for $\tau = 1$.[52]

TABLE A.1

Definitions of Variables Used in Analysis

Variable	Definition
Race	Whether participant black, hispanic, or other. ‘‘White" is the omitted category in the regressions.
Age	Whether at application, the participants reported they were aged 30 to 39, 40 to 49, 50 to 54, or over 54.
Sex	Whether respondent female.
Own children ≤ age 6 at home	Whether the participant has a child who is less than 6 years of age, and living in the participant’s household.
Marital Status	Whether respondent single, or divorced, widowed or separated.‘‘Married" is the omitted category in the regressions.
Receiving AFDC	Participant’s household at time of application is receiving Aid to Families with Dependent Children (AFDC).
Receiving Food Stamps	Participant’s household at time of application is receiving Food Stamps.
Receiving General Assistance	Participant or a member of participant’s family who lives with the participant at time of application is receiving General Assistance, Home Relief or any other welfare other than AFDC and Food Stamps.
Welfare recipient	Participant who is classified as either an AFDC or General Assistance recipient or both.
Highest grade completed	Whether at application, participant’s highest grade completed was less than the 10th grade, the 10th or 11th grade, 1-3 more years than high school, or 4 or more years than high school. The omitted category in the regressions is 12th grade.
Last year’s earnings	Whether participant’s labor market earnings before taxes and other deductions in 12 months preceding application was less than $3,000, $3,000-$6,000, $9,000-$12,000, $12,000-$15,000, or greater than $15,000. The omitted category in the regressions is $6,000-$9,000 of earnings in the year prior to application.
Employment status	Whether the last date of employment was less than 2 months, 3-5 months, 6-8 months, 9-11 months, or greater than 12 months before application, or never employed. The omitted category in the regressions is ‘‘employed at application".

(Continued)

TABLE A.1 (Continued)

Definitions of Variables Used in Analysis

Variable	Definition
Unemployment rate at RA	Fraction of the labor force looking and unable to find work in month of random assignment. Constructed for county and metropolitan area of training center service area using data from the U.S. Department of Labor publication ''Labor Force, Employment and Unemployment Estimates for States, Labor Market Areas, Counties and Selected Cities" for the years 1987 to 1989. Where service areas include more than one county or metropolitan area, rates are weighted by population (population data taken from U.S. Department of Commerce Regional Economic Information System (REIS) data).
Wage and salary employment ratio at RA	Total wage and salary employment divided by total employed population in year of random assignment. Constructed for county and metropolitan area of training center service area using U.S. Department of Commerce Regional Economic Information System (REIS) data.
Service industry share of employment at RA	Total persons employed in service sector divided by total population in year of random assignment. Constructed for county and metropolitan area of training center service area using U.S. Department of Commerce Regional Economic Information System (REIS) data.

All characteristics used in regression analyses are the experimental participant's characteristics on the date of application.

TABLE A.2

Estimated Utility Weights on Performance Outcomes from Model 4, Table 5

	Variable	Coef. Est.	(|t stat.|)
1.	$d^{AWT} \cdot AWT$	-.117	(8.775)
2.	$(1-d^{AWT}) \cdot AWT$	.044	(.346)
3.	$[(1-WELF)+WELF\cdot(1-d^{WERT})] \cdot d^{CE} \cdot d^{ERT} \cdot ERT$	-.078	(.387)
4.	$[(1-WELF)+WELF\cdot(1-d^{WERT})] \cdot (1-d^{CE}) \cdot d^{ERT} \cdot ERT$	-1.392	(2.945)
5.	$WELF\cdot d^{WERT} \cdot d^{CE} \cdot d^{ERT} \cdot ERT$	.273	(1.123)
6.	$WELF\cdot d^{WERT} \cdot (1-d^{CE}) \cdot d^{ERT} \cdot ERT$	-2.730	(3.248)
7.	$d^{CE} \cdot CE$	-.332	(16.515)
8.	$(1-d^{CE}) \cdot CE$	-.380	(6.256)
9.	$[(1-WELF)+WELF\cdot(1-d^{WERF})] \cdot d^{ERF} \cdot ERF$	.937	(5.311)
10.	$[(1-WELF)+WELF\cdot(1-d^{WERF})] \cdot (1-d^{ERF}) \cdot ERF$	-.439	(2.198)
11.	$WELF\cdot d^{WERF} \cdot d^{ERF} \cdot ERF$	1.057	(1.932)
12.	$d^{AWEF} \cdot AWEF$	.566	(10.427)
13.	$(1-d^{AWEF}) \cdot AWEF$	.629	(17.739)
14.	$d^{AWWF} \cdot AWWF$	.022	(1.152)
15.	$(1-d^{AWWF}) \cdot AWWF$	-.154	(13.388)

Notes:

1. The estimated coefficients reported in Panel A and C of Table 5, and in this table together comprise the entirety of Model 4.

2. Explanation of variables:

$d^k=1$ if performance measure k is rewarded on the applicant's random assignment date, $d^k=0$, otherwise, for $k=AWT, ERT, ERF, AWEF, AWWF, CE, WERT$, and $WERF$.

AWT, ERT, ERF, $AWEF$, $AWWF$, CE describe the applicant's predicted performance in the performance categories conditional on their characteristics, local economic characteristics, and the training category (Table 1 defines these).

$WELF=1$ if the applicant is a welfare recipient, $WELF=0$, otherwise.

3. All applicants enrolled under incentive policies which rewarded ERT (*i.e.*, the cell $d^{ERT}=0$ is empty).

4. All incentive policies that rewarded $WERF$, rewarded ERF also (*i.e.*, the cell $(1-d^{ERF})d^{WERF}=1$ is empty).

5. Table 5's column (4) is created by contrasting the above coefficients as follows:

Activated measure	Performance outcome	Line numbers from above table				Subst. coefficients from above table				Result appearing in Table 5
ERF	*ERF*	(9)	-	(10)	=	.937	+	.439	=	1.375
AWT	*AWT*	(1)	-	(2)	=	-.117	-	.044	=	-.161
AWEF	*AWEF*	(12)	-	(13)	=	.566	-	.629	=	-.063
AWWF	*AWWF*	(14)	-	(15)	=	.022	+	.154	=	.176
CE	*CE*	(7)	-	(8)	=	-.332	+	.380	=	.048
CE	*ERT*	(3)	-	(4)	=	-.078	+	1.392	=	1.315
CE	*WERT*	(5)	-	(6)	=	.273	+	2.730	=	3.003
WERT	*WERT*	(5)	-	(3)	=	.273	+	.078	=	.351
WERF	*WERF*	(11)	-	(9)	=	1.057	-	.937	=	.120