

# Measuring Government Performance: Lessons from a Federal Job-Training Program

By PASCAL COURTY AND GERALD MARSCHKE\*

The history of the job-training programs funded under the Job Training Partnership Act (JTPA) offers two cautions to the many who believe that government bureaucracies can be reformed with market-like performance incentives. First, bureaucratic discretion and accurate measurement of performance outcomes sometimes conflict. This finding has turned up from time to time in the small empirical incentive literature.<sup>1</sup> We find that JTPA bureaucrats whose performance is aggregated for evaluation may manipulate performance outcomes to maximize intertemporal awards. Moreover, we identify a moral-hazard problem specific to the use of labor-market-based performance measures in training programs. This moral-hazard problem occurs because local-level bureaucrats privately observe participants' performance outcomes and manipulate the measurement procedures to maximize their private rewards rather than social return.

Second, the lack of well-defined goals and the lack of residual claimants in government bureaucracies make it unlikely that incentives

will improve efficiency in public bureaucracies in the same way as in private ones. We focus on the process by which the incentive-designer learns about and responds to the agent's unintended reaction to performance incentives. The issues of unintended responses and the process by which incentives evolve are little studied in the incentives literature but are crucial to understanding how compensation systems arise in organizations and how management in government organizations can be improved (William Bruns, 1992).

## I. JTPA and Its Performance-Incentive System

Our conclusions about the effects of performance incentives in government bureaucracies are drawn from our study of the large-scale incentive system created under the Job Training Partnership Act (JTPA) of 1982.<sup>2</sup> The 1982 Act drastically remodeled the bureaucracy that runs the federal government's most important job-training program for the poor. As a result, the JTPA bureaucracy is highly decentralized: job training is carried out by more than 620 semiautonomous sub-state training centers. Instead of a rigid, comprehensive set of rules to regulate bureaucratic *conduct*, the federal government uses a loose set of financially backed performance incentives to influence *outcomes*. According to the Act, Congress intended the performance incentives to measure the training centers' success in developing participants' human capital [Section 106(a)].<sup>3</sup> The translation of this goal

\* Department of Economics, Universitat Pompeu Fabra, Ramon Fargas 25-27, 08005 Barcelona, Spain, and Department of Economics, SUNY-Buffalo, 415 Fronczak Hall, Box 601520, Buffalo, NY 14260-1520, respectively. We acknowledge support from NSF-SBR-91-31143 to J. Heckman. Courty acknowledges support from the Sloan Foundation and the University of Chicago. Marschke acknowledges support from a grant from the Upjohn Institute and a grant from the American Bar Foundation. We are grateful for the support and comments of James Heckman. We thank Robert LaLonde, Canice Prendergast, Sherwin Rosen, and seminar participants at the University of Chicago, the University of Arizona, the Upjohn Institute, and the Milken Institute for helpful suggestions. All errors are exclusively the responsibility of the authors.

<sup>1</sup> See, for example, Paul Healy (1985) on manager compensation, Paul Oyer (1995) on sales-force compensation, and Beth Asch (1990) on military-recruiter compensation.

<sup>2</sup> For a description of the JTPA incentive system between 1987 and 1989 see Courty and Marschke (1995).

<sup>3</sup> It should be noted that human capital is not the exclusive goal of JTPA. See James Heckman and Jeffrey Smith (1995) for a discussion of other goals such as equity and special service to the eligible persons in the very bottom of the income distribution.

into measures of performance has raised problems because human capital is not directly measurable. The remainder of this section considers two critical aspects of the JTPA incentive system: (i) the choice of the measures of performance and (ii) the measurement procedures for the performance outcomes.

The Act gave the U.S. Department of Labor (DOL) the responsibility to develop a workable set of performance measures based on the Act's mandate. The necessity that the incentive system keep measurement costs low while giving training managers rapid feedback on their performance led designers of the JTPA incentive system to choose a set of short-term labor-market measures based on an enrollee's employment status, wage, and earnings.

The incentive-designers faced several problems in formulating procedures for measuring enrollee labor-market outcomes. First, job-training service plans do not have a clear beginning and end. If training were of a uniform length, say, six months, then measuring performance would be simple. The DOL would require the training center to report enrollee outcomes starting in the seventh month following the first day of enrollment. But individual training services run for different lengths of time, and an enrollee may receive any number of these services depending on how events unfold. Second, applicants enroll and receive services at any point in the year. An arbitrary date such as June 30 would catch some enrollees in the middle of a program and others at the beginning or end. Thus, the incentive-designers did not impose a uniform date of measurement because they wished to encourage training centers to tailor training services to the enrollee's characteristics and the circumstances of the local economy.

In the absence of a better solution, the incentive designers decided to allow training centers to report the performance of an enrollee *any* time between her last day of training and 90 days after the last day of training. After the date at which the training center chooses to terminate the enrollee and report the enrollee's labor-market outcome, the enrollee cannot receive additional training. Since the one date corresponds to two concepts (termination of the enrollment period and report of the labor-market outcome), we will refer to

that date as either the "termination date" or the "reporting date," depending on the context. Finally, the incentive designers had to aggregate the raw labor-market outcomes into summary measures of performance that would form the basis for rewarding training centers. In JTPA's incentive system, as in many incentive systems, the fiscal calendar and incentive calendar correspond. On June 30, the end of the fiscal or program year, all labor-market outcomes reported in the previous 12 months are averaged to constitute the yearly performance outcome.

## II. What Does the Incentive System Measure?

### A. *Reporting Good Outcomes and Waiting on Bad Ones*

At the end of an enrollee's training, training centers face an option: report the enrollee's labor-market outcomes, or postpone reporting in hopes that the outcome improves. In theory, the optimal reporting strategy depends in a complex way on many variables influencing these measures. In Courty and Marschke (1996), hereafter CM, we analyze this strategy in detail and argue that one can restrict attention to a single performance measure, the employment status at the reporting date, for two reasons: (i) the employment-based performance measure plays a disproportionate role in the determination of the award; (ii) the other performance measures are measures of wages or earnings, which are harder to manipulate by just choosing the reporting date.

Figure 1 presents the distribution of the difference in days between the enrollee's last day of training and the date on which the training center reports the enrollee's labor-market outcomes, by employment status at the end of training.<sup>4</sup> Figure 1 shows that the enrollees employed on the date training ended (repre-

<sup>4</sup> The figures in this paper are based on data from the national JTPA Study, a national experiment conducted at 16 JTPA job-training centers in the late 1980's (see the data's description in Fred Doolittle and Linda Traeger [1990]). The study was designed to measure the impact of JTPA on the earnings of the poor.

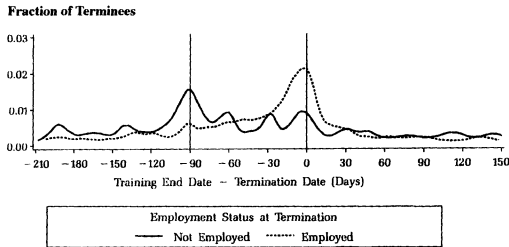


FIGURE 1. INCIDENCE OF DELAYED TERMINATION BY EMPLOYMENT PERFORMANCE OUTCOME

Notes: If an enrollee reported multiple training spells, the spell with the end date closest to the termination date is used for comparison. The plot was smoothed using a cubic spline. There are 2,236 useable observations represented by the graph. Reprinted from Courty and Marschke (1996) with the permission of the publisher.

sented by the dashed line) are reported on or very close to the last day of training, while the enrollees not employed at the end of training (represented by the solid line), are reported on average 26 days after the end of active training. Notice that a large fraction of the enrollees remaining unemployed throughout the 90 days following the end of active training are reported around the 90th day. The difference between the fraction of individuals who are employed on the day they finish their training and the fraction employed on the day the training center officially reports them is a measure of the private gains from waiting to report unemployed individuals. According to this measure, the training centers would produce an employment rate outcome 20-percent lower if it were required to report individuals on the date their actual training ends.<sup>5</sup>

Of course, the decision rule uncovered in Figure 1 and attributed to gaming may be precisely the rule that maximizes the human-capital impact of job training. Case workers may not wish to terminate unemployed enrollees because terminating them precludes case workers from offering any further train-

ing. The evaluation literature measures job training's human-capital impact by changes in earnings. In an estimation not reported here (due to space constraints), we compare the after-training earnings for unemployed enrollees whose terminations are delayed past the end of active training to unemployed enrollees whose terminations are not delayed, holding the characteristics of enrollees constant. We find that, among enrollees who were not employed at the end of training, delaying reporting or termination had no impact on after-training earnings. We conclude that the decision rule uncovered in Figure 1 is consistent with performance outcome-manipulation or gaming, and that it is inconsistent with human-capital maximization.

### B. Reporting Bad Outcomes in Good Times Only

The other reason why training centers have an incentive to manipulate the reporting date is because it determines the program year in which the participants are counted. The Act directs that only those training centers whose performance outcomes exceed numerical standards shall receive awards [JTPA Section 202(a)(1)(3)(B)]. In addition, training centers receive all or most of their award for simply meeting the state-defined standards. Consider a stylized incentive system where the training center receives a fixed payment if the yearly performance outcome, computed as the average performance over all the participants who have been reported in that year, exceeds a fixed performance standard.

Because the labor-market outcomes depend upon random factors, such as the state of the local economy, which are outside the training center's control, the training center does not know its final aggregate performance outcome until the end of the program year. As shown previously, the training center postpones reporting outcomes for unemployed enrollees. In the last month, if the training center finds itself above or hopelessly below its standards, it should terminate unemployed enrollees who are no longer actively training even before the 90-day grace period expires.

Figure 2 presents evidence that training centers follow this strategy. Figure 2 shows the

<sup>5</sup> The risk that an employed enrollee will lose her job deters training centers from postponing termination (see Courty and Marschke, 1996).

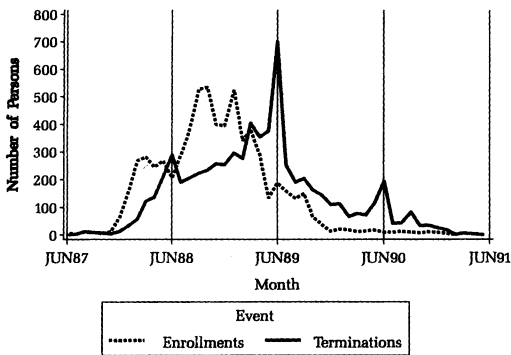


FIGURE 2. MONTHLY ENROLLMENT AND TERMINATION COUNTS

Notes: There are 6,972 usable observations represented by the graph. Reprinted from Courty and Marschke (1996) with the permission of the publisher.

aggregate monthly enrollment and termination (or performance report) counts for a set of 16 experimental sites (see footnote 4). Notice that a large fraction of the year's reports occur in June.<sup>6</sup> The program year-end pile-up is not explained by the enrollment patterns of the JTPA enrollees in our data set.<sup>7</sup> Moreover, in CM we report substantially lower performance outcomes among persons reported in June, as compared to enrollees reported in other months. In addition, the plot of the employment outcome versus the standard in Figure 3 shows that the employment-rate measure drops sharply from above the standard to just meet the standard at the end of the year. This sharp movement in the employment outcome is consistent with a practice of holding unemployed enrollees in an inventory, drawing down the inventory only when achievement of the standard is assured, and then only until the overall outcome attains the standard.

<sup>6</sup> When we examine training centers individually, we find a program year-end pile-up in at least one program year in 14 of the 16 sites.

<sup>7</sup> Many enrollees are enrolled in classroom training services in community colleges. In CM we investigate and rule out a link between the June pile-up and the academic cycle.

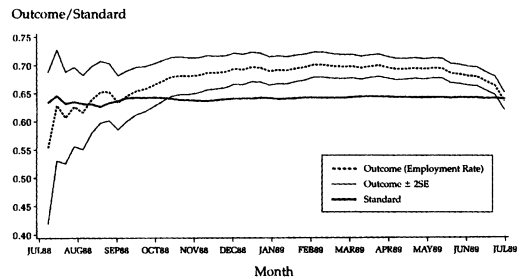


FIGURE 3. EMPLOYMENT RATE AT TERMINATION IN PROGRAM YEAR 1988

Notes: There are 4,321 usable observations represented by the graph.

### C. Discussion

We identified a fundamental problem specific to the use of labor-market performance measures. This fundamental problem is the consequence of three circumstances: (i) the labor-market outcomes are influenced not only by the training services but also by other unrelated factors; (ii) training centers privately observe the labor-market outcomes; and (iii) training centers observe valuable local information that can be effectively used only if they have some discretion over the training history. Bureaucratic discretion and performance measurement conflict because case managers privately observe participants' labor-market outcomes and can take difficult-to-monitor actions which influence these outcomes. This conflict creates a moral-hazard problem where case managers have an incentive to use their private information to maximize their private reward rather than social return.

The costs imposed by gaming may doom such systems in many principal-agent contexts, which may be why such high-powered incentives in the private sector are rare. We suspect that where such incentives are successful, principals have been able to limit gaming costs by vigilant monitoring and adjustment. But because the principal may lack the incentives to monitor and fix them, the adjustment costs may be especially high for government bureaucracies. We discuss these issues and present additional evidence in the following section.

### III. Perfection and Maintenance of Performance-Incentive Systems

Most modern corporations and many other organizations spend much to develop, maintain, and perfect systems for evaluating or appraising performances.

—Bruns (1992 p. 4)

Case studies suggest that, in order for performance-incentive systems to be effective, they must continuously be adapted to address unforeseen consequences and dysfunctional responses. This continuous change is less likely to occur in government organizations than in private organizations because these former organizations differ from the latter in two fundamental ways: (i) government bureaucracies lack a residual claimant, which in market bureaucracies provides monitoring and motivation; and (ii) government bureaucracies often lack a well-defined goal, which hampers their ability to identify dysfunctional behavior.

JTPA incentive-designers have attempted to adapt the incentive system in response to new information. Nevertheless, the incentive-designers' responses have not completely eliminated inefficient training decisions. Consider for example the 90-day reporting constraint described above in Section I and analyzed in Section II. The 90-day reporting constraint was not present in the first years of the incentive system. The incentive-designers included this constraint to address the problem of training centers never reporting and terminating poorly performing enrollees. As shown in the previous section, however, training centers responded to the 90-day constraint by reporting employed enrollees on their training end date and by postponing the reporting of unemployed ones as long as possible. Therefore, the 90-day constraint is at most a second-best solution to the measurement problems that arise when training centers have discretion over the reporting date.

A second piece of evidence illustrates even more strikingly the difficulties in trying to change behavior by changing the incentives. A number of DOL investigations found that training centers instituted "quick fixes" such as job-placement services and other services that had no long-term impacts on enrollees'

skills. In response, the DOL formulated a number of follow-up measures based on outcomes measured three months after the termination date. The DOL introduced follow-up measures "in order to encourage placement of adults in longer-term, more stable jobs" and because termination-based measures "merely emphasizes getting participants jobs" (*Federal Register*, 14 January 1986, 51(9), p. 1877). Eventually, the DOL abolished altogether the measures of outcomes on the date of termination.

In a 1994 telephone survey we administered to 11 of the 16 training centers represented in our data set, we asked administrators how they responded to the newly instated follow-up measures. Most administrators indicated that, upon the introduction of the follow-up measures, case managers began tracking terminees until the follow-up period expired. To increase the chances that an employment match lasted until the third month, some training centers reported that they offered special services between termination and follow-up, such as child-care, transportation, and clothing allowances. Case managers also attempted to influence employers to keep their clients until the third month. If the client lost her job, case managers scheduled job-counseling appointments and offered placement services. All training administrators reported that after the third month they did not contact the client again, nor did they offer more placement or counseling services. Although these activities may produce higher employment rates in the 13th week after termination, we believe that they are not likely to have a significant long-term effect on enrollees' human-capital stocks.<sup>8</sup>

### IV. Conclusion

This paper examines the provision of incentives in a federal job-training program. We find that training-center bureaucrats respond to

<sup>8</sup> The behavior described here may explain why Heckman and Smith (1995) found little correlation between follow-up performance measures and the earnings impacts of JTPA.

the incentive system by maximizing their private rewards, possibly at the expense of social welfare. This is to our knowledge the first systematic evidence of gaming in an incentive system providing "low-powered" incentives. These results suggest that making the reward small may not eliminate dysfunctional behavior because bureaucrats maximize what they are told to maximize, especially when this constitutes the only precisely defined and quantifiable objective.

We identify a cost of using market-based measures when the agent has private information about these measures and some discretion over measurement. We show that job-training bureaucrats use their discretion over the training plan to report the labor-market outcomes when they are good and to report the bad ones in good years only. This conflict between bureaucratic discretion and objective measurement imposes moral-hazard costs and reduces any efficiency gains from implementing market-based measures.

In the private sector, owners or managers devote substantial effort and resources to maintaining and perfecting performance-incentive systems. We find that the incentive-designers adjust and perfect the incentive system as they receive information about dysfunctional behavior. Nevertheless, this process is likely to be slower in a government bureaucracy because of the absence of a well-defined goal and because these bureaucracies are controlled by multiple principals with possibly different perceptions of the impact of incentives and conflicting interests.

## REFERENCES

- Asch, Beth J.** "Do Incentives Matter? The Case of Navy Recruiters." *Industrial and Labor Relations Review*, February 1990, 43(3), pp. 89S-106S.
- Bruns, William J.** "Introduction," in William J. Bruns, ed., *Performance measurement, evaluation, and incentives*. Boston, MA: Harvard Business School Press, 1992, pp. 1-14.
- Courty, Pascal and Marschke, Gerald R.** "The JTPA Incentive System." Unpublished manuscript, University of Chicago, 1995.
- . "Moral Hazard Under Incentive Systems: The Case of a Federal Bureaucracy," in Gary D. Libecap, ed., *Advances in the study of entrepreneurship, innovation, and economic growth*. Greenwich, CT: JAI Press, 1996, pp. 157-90.
- Doolittle, Fred and Traeger, Linda.** *Implementing the National JTPA Study*. New York: Manpower Demonstration Research Corporation, 1990.
- Healy, Paul.** "The Effect of Bonus Schemes on Accounting Decisions." *Journal of Accounting and Economics*, April 1985, 7(1-3), pp. 159-81.
- Heckman, James J. and Smith, Jeffrey A.** "The Performance of Performance Standards: The Effects of JTPA Performance Standards on Efficiency, Equity, and Participant Outcomes." Unpublished manuscript, University of Chicago, 1995.
- Oyer, Paul.** "The Effect of Sales Incentives on Business Seasonality." Unpublished manuscript, Princeton University, 1995.