DISCUSSION PAPER SERIES

No. 3164

AN EMPIRICAL INVESTIGATION OF GAMING RESPONSES TO EXPLICIT PERFORMANCE INCENTIVES

Pascal Courty and Gerald Marschke

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP3164.asp

ISSN 0265-8003

AN EMPIRICAL INVESTIGATION OF GAMING RESPONSES TO EXPLICIT PERFORMANCE INCENTIVES

Pascal Courty, London Business School (LBS) and CEPR
Gerald Marschke, State University of New York

Discussion Paper No. 3164
January 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3164

January 2002

ABSTRACT

An Empirical Investigation of Gaming Responses to Explicit Performance Incentives*

This Paper studies a particular kind of gaming response to explicit incentives in a large government organization. The gaming responses we consider occur when agents strategically report their performance outcomes to maximize their awards. An important contribution of this work is to examine whether this behaviour diverts resources (eg agents' time) from productive activities or whether it simply reflects an accounting phenomenon. We evaluate the efficiency impact of the behaviour we identify and find that it has a negative impact on the true goal of the organization.

JEL Classification: J33 and L14
Keywords: contract theory, government incentives, moral hazard, multitasking and performance incentives

Pascal Courty
London Business School
Sussex Place
Regent's Park
LONDON
NW1 4SA
Tel: (44 20) 7262 5050 x3317
Fax: (44 20) 7402 0718
Email: pcourty@lbs.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=134679

Gerald Marschke
State University of New York, Albany
Department of Economics
BA-110, SUNY-Albany
Albany, NY 12222
USA
Tel: (1 518) 442 5274
Fax: (1 518) 442 5298
Email: marschke@csc.albany.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=153437

* Courty and Marschke 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 encouragement of James Heckman. We thank Michael Gibbs, Robert LaLonde, Canice Prendergast, Sherwin Rosen, Jeffrey Smith and seminar participants at Chicago, Rochester, Arizona, IESE, and Bristol for helpful suggestions. All errors are exclusively the responsibility of the authors.

Submitted 13 December 2001

1 Introduction

When performance awards depend non-linearly on performance outcomes, agents have an incentive to manipulate the timing of their performance. Under bonus-based contracts, for example, agents may time their performance so that they just meet the numerical standard to receive their bonus. Under this timing strategy, agents postpone excess performance above the standard from good to bad years thereby increasing their chances to receive bonuses in bad years. Although these timing responses to non-linear contracts have received some attention in the organization literature, there is no direct evidence that this behavior creates a welfare loss. (See Gibbons (1997) and Prendergast (1999) for a survey.) The main goal of this paper is to examine whether this behavior reflects a misallocation of real resources or simply an accounting phenomenon.

The distinction between timing responses based on an efficiency criterion plays a key role in this paper. Timing strategies that create no welfare loss while increasing the agents' chances of earning bonuses are accounting manipulations. Accounting responses do not create any welfare loss because the organization can neutralize this behavior by appropriately discounting the performance award. Alternatively, timing strategies may lead agents to "invest in the wrong tasks" or to "game" the performance incentive system (Holmstrom and Milgrom, 1991 and Baker, 1992).[1] Timing strategies are evidence of gaming if they do not only imply some kind of accounting manipulation, but also a costly misallocation of resources.

We investigate this distinction between accounting and gaming responses in a large federal job training program for the disadvantaged, created by the Job Training Partnership Act (JTPA) of 1982. JTPA is one of the first large-scale implementations of financial performance incentives in a federal bureaucracy. The JTPA incentive system rewards training agencies on the basis of their trainees' labor market achievements on the date they graduate from training. The incentive system leaves training agencies some discretion over the timing of the graduation date. The central focus of this paper is to study

[1]We take in this work a wide interpretation of the multi-tasking framework and consider the timing strategy as one of the agent's tasks even though it has no value to the principal.

how training centers use this discretion.

Some of the timing strategies presented here have already been identified in our previous work (Courty and Marschke, 1997). This paper builds on that work in two important ways. First, this paper offers an analytical description of the training agency's optimal graduation strategy that generates a much wider set of implications than were considered previously. Second and most importantly, this paper investigates some of the welfare implications of the timing strategies we identify. This constitutes the main contribution of this work.

The paper develops as follows. The following section reviews the organization literature on gaming responses to explicit incentives. Section 3 describes the JTPA organization, its performance incentive system, and our data. Section 4 shows that the program administrators maximize their performance awards by optimally timing the graduation date of program participants. Section 5 studies the efficiency of these timing strategies and Section 6 summarizes our findings.

2 Incentive Theory and Gaming Evidence

Bonus-based compensation schemes award bonuses when performance outcomes exceed preset numerical standards. Such schemes are common in some occupations: CEOs, sales people, and piece workers receive bonuses when firm earnings, yearly sales, and piece rates exceed predetermined targets, quotas or thresholds (Murphy, 1998). An important regularity that is emerging from the empirical literature on incentives is that nonlinear incentive contracts often encourage agents to manipulate the timing of their performance.

In an early contribution to this literature, Healy (1985) documents that managers who are compensated for meeting annual income thresholds use their discretion over the timing of income reporting to smooth their compensation across accounting years. More recent works report similar timing responses to threshold effects in other settings. For example, Asch (1990) showed that navy recruiters who receive awards for meeting year-end recruitment quotas respond by reallocating their work efforts over the year. Similarly, Oyer (1998) showed that there is more variability in firms' sales at the end of the fiscal years—when sales persons' bonuses are computed—than in the middle. The empirical evidence on threshold-motivated timing responses to incentives is surveyed in Gibbons and Prendergast.[2]

Four components common to many incentive systems create these timing responses. First, the incentive award is a non-linear function of the agent's performance. Second, the performance outcomes are aggregated over fixed periods and the agent is rewarded at the end of each period on the basis of her cumulated performance. Third, the performance outcomes vary for random reasons that are outside the agent's control. Finally, the agent is able to choose the period in which performance is reported. Although the argument applies generally to non-linear awards, the rationale behind these timing responses is best illustrated under a bonus-based award.

Consider a simple environment where random fluctuations imply that the agent would succeed in winning a bonus in good years but fail in bad years if she reported performance outcomes without delay. Now, consider the following strategy that takes advantage of the agent's discretion over performance reporting. In bad years, the agent delays performance below the standard until the next bonus year, while in good years the agent delays positive performance above the standard. In very bad years, it may be optimal to 'take a bath', that is, to report bad performances and carry over good performances until the next bonus year. Under this timing strategy, the agent is more likely to receive a bonus in bad years without compromising the chances of winning bonuses in good years.

In exposing this timing behavior, the literature shows the variety and creativity of agent responses to incentives but offers no direct evidence that such behavior creates a welfare loss. In fact, the literature does not make an explicit distinction between accounting and gaming responses as defined in the Introduction. Based on this distinction, a review of the timing literature suggests that agents manipulate performance accounting

[2]There is also a literature studying timing responses by tax-payers to minimize their tax bill (e.g. Dickert-Conlin and Chandra,1999, and Goolsbee 2000). Brown, Harlow and Starks (1996) and Chevalier and Ellison (1997) show that fund managers modify their investment strategies over the year to maximize investment flows into their funds.

but is inconclusive regarding whether this manipulation is gaming. For example, the managers in Healy's study may not have to consume resources in figuring out at the end of the accounting year how to optimally report their financial performance outcomes. Similarly, the military recruiters in Asch's study as well as the salespeople in Oyer's study may vary their effort supply over the contract year as a result of the incentives they face, but this by itself is not evidence of inefficiency. This work is the first attempt to estimate empirically the costs of timing activities in performance incentives.

Discovering whether timing strategies are inefficient has important economic implications. Evidence of welfare losses would contribute to a general argument, found in the organization literature, attributing the scarcity of explicit incentives to the difficulty of measuring performance outcomes. (See Jensen and Meckling (1992) for a review.) According to this argument, principals can specify explicit performance measures only approximately, which results in agents gaming the system by optimizing with respect to the actual measures instead of the intended unmeasurable objective (Holmstrom and Milgrom and Baker). Despite the popularity among organization theorists of the gaming argument as an explanation for the scarcity of explicit incentives, there is little formal evidence in its support. In fact, the gaming argument relies exclusively on anecdotes accumulated over the years by organizational behaviorists (e.g. Kerr (1975) and Lawler (1982)). Although these anecdotes suggest that explicit incentives do not always work as intended, they do not demonstrate that these unintended responses are evidence of inefficiency.

The set of threshold-driven timing responses sketched above seem to be a good candidate to conduct a formal test for costs of the unintended responses. We propose to conduct this test in the JTPA organization for two reasons. First, the program administrators in JTPA face a similar incentive environment as those studied in previous empirical work. Meeting a fixed numerical performance standard plays a key role in the JTPA award functions. Actually, several award functions in our sample are identical to a bonus-based formula. Second, a recent evaluation of JTPA has generated a wealth of data on its functioning, including separate accounts, by program administrators and by participants, of the training outcomes. Our ability to observe agent behavior that is normally hidden from the organization, at least in the short-run, makes this study especially compelling. In addition, these data are sufficiently rich to generate measures of agent productivity. In contrast with previous timing studies, these measures will allow us to conduct the efficiency study that is required to test the gaming hypothesis.

Our strategy will be to identify some timing responses and to sign their welfare impacts. Although it is possible to test for welfare costs, we cannot address more general efficiency questions (e.g. whether the JTPA incentive system is second or third best). Our work, however, presents findings that should be taken into account in a more complete evaluation of the overall efficiency of the JTPA incentive system.

3 JTPA: Organization, Incentive System and Data

This section starts with a few summary facts about job training under JTPA, then describes the JTPA incentive system emphasizing those features that are most relevant for this study and finally introduces the three data sources that will be used in this paper. The Job Training Partnership Act of 1982 created what is presently the largest federal employment and training program serving the disadvantaged. Its current annual budget is approximately $4 billion and it serves nearly one million people annually.[3] The JTPA bureaucracy is unusual for three reasons. First, JTPA is highly decentralized: job training is carried out by more than 620 semi-autonomous sub-state training agencies. Second, the Act gives these training agencies significant discretion over who is admitted to the program and how training is conducted. Third, and most important for this study, 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*.

[3]For a detailed description of JTPA, see Johnston (1987). For detailed descriptions of its incentive system see Barnow (1992), Courty and Marschke (2000), and Dickinson *et al.*, (1988). For empirical analyses of the effects of incentives on the population selected to receive training see Heckman, Smith and Taber (1996), Anderson, Burkhauser, and Raymond (1993), Cragg (1997) and on the service provided see Marschke (2000).

3.1 The Incentive System

In this section, we describe the main features of the incentive system which will be the focus of our empirical work. Congress intended the performance incentives to measure the training agency's success in developing participants' human capital (Job Training Partnership Act, Sec. 106(a)).[4] 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 DOL chose a set of short-term labor-market measures based on an enrollee's employment status, wage, earnings, and/or number of weeks worked at the end of job training. Table A.1 in the appendix defines the federal measures.

Along with selecting labor market measures, the DOL also decided that the labor market outcomes would be measured on the date participants graduate from the program.[5] The graduation date is the date the training agency officially closes an enrollee's case and removes her from its rolls. To receive additional training after the graduation date, the enrollee must be re-enrolled. Note that the same date corresponds to two concepts: *graduation*, that is, termination of the enrollment period, and *reporting* of the labor market outcome.[6] Training agencies' discretion over the graduation date is limited by the following rule: the JTPA regulations require training agencies to graduate a participant—at which time her labor market outcomes are recorded—within 90 days of her last day of training.

The individual states administer the incentive system. The JTPA fiscal year lasts from July 1 to June 30 of the next calendar year. At the end of each fiscal year, or *program* year, the state rewards (or sanctions) their training agencies on the basis of their year-end performance outcomes. In aggregating performance over a fixed period of 12 months, the JTPA incentive system is thus similar to many incentive systems. The award augments

[4]It should be noted that human capital development is not the exclusive goal of JTPA. The Act, the Act's amendments, and the U.S. Department of Labor in its role as interpreter of the Act mention other goals such as equity and special service to individuals in the bottom of the income distribution. See Heckman and Smith (1995) for a discussion of these goals and how they are expressed through the JTPA performance incentives.

[5]In the time period for our study, 1987-89, the date of performance measurement corresponds to the graduation date.

[6]We thank Canice Prendergast for suggesting this terminology.

the training agency's budget in the following year. Averaging over all JTPA training agencies, the award is the source of approximately seven percent of the operating budget.

3.2 The Data

This study relies on three data sources. The first two data sources are an administrative data source and a participant data source which were collected for the DOL-commissioned National JTPA Study (NJS). The NJS was an experimental study of the effectiveness of JTPA conducted between 1987 and 1989. The study was conducted using a classical experiment methodology according to which JTPA applicants were randomized into treatment and control groups. Members of the treatment group received training and members of the control group were denied JTPA training services for 18 months. Sixteen of the organization's roughly 620 job training agencies participated in the NJS.[7] The administrative data include enrollee-level information on training activities, enrollment and graduation dates, and the employment-at-graduation performance outcome. The participant data were produced from a questionnaire administered to participants at the time of their application and again 18 months later. As a result, this second data source includes detailed information on individuals' participation in job training, schooling, and job search and on employment during the 18 month period following random assignment.

The third data source describes the JTPA incentive policies. We obtained these policies from the states' governors offices.[8] We briefly summarize the main characteristics of the incentive policies that will be used in the paper. Although each state uses similar sets of performance measures, the policies show that the award formulas differ somewhat across states. To illustrate a stylized incentive contract, let S^k be the training agency's performance *outcome* on *measure* k, and $\bar{S}^k$ be the numerical *standard* for the same measure. The simplest type of policy would give an award ψ^k for each measure k only if the

[7] See Doolittle and Traeger (1990) for a description of the implementation of the NJS. See Orr *et al.* (1994) for a detailed description of the results of the NJS.

[8] We collected these data while research associates for the Center for Social Policy Evaluation, at the University of Chicago. A summary description of this data can be found in Courty and Marschke, 2000.

reported outcome exceeds its standard. Thus, the program year's incentive award is,

$$A = \sum_k \psi^k \, 1\{S^k > \bar{S}^k\}, \tag{1}$$

where $1\{\cdot\}$ equals one if the condition in brackets is true. Because the Act leaves the design of the incentive policies to the individual states, there is much variation in the precise functional form of the awards. In some states, training agencies receive additional awards for exceeding standards, with the performance-award relationship displaying many sorts of non-linearities, such as piece-wise linear forms and nested qualifying conditions.

What matters for this study is that the Act requires only that awards be contingent on achieving standards. As a consequence, the major portion of the potential award across all training agencies in our study is paid out for simply meeting the standard. For our analysis, we will assume that each training agency faces functional form (1) because it is the simplest formulation allowing us to test for incentive responses common to all incentive policies.

4 Timing Activities in JTPA

Courty and Marschke (1997) present evidence suggesting that training agencies time the graduation date to maximize their awards. This section as well as the appendix extend the argument sketched in Courty and Marschke (1997). A shortcoming of our previous work was that it provided no systematic statistical modelling. In this section, we develop and test econometric models that more carefully measure the behavior previously identified. Also, our previous work considered only the impact of the employment-based performance measure on the optimal graduation strategy. The appendix extends the argument of Courty and Marschke to examine additional performance measures and shows that although these measures matter their influences are second order.

4.1 Graduation Timing Before the Program Year-End

Because labor market outcomes vary over time naturally on their own, training agencies have an incentive to choose the date they report enrollees' employment outcomes strategically. At the end of an enrollee's training, training agencies face a decision: to graduate the enrollee and report her labor market outcomes, or to postpone graduation in hopes that the outcome improves. The optimal graduation strategy leads the training agency to graduate enrollees who are employed within the 90 day period following training either on the last day of training or on the first day of employment, whichever occurs first, and all others on the 90th day following training end.

Figure 1 shows the order of events for a particular enrollment in the National JTPA Study. Application, acceptance into the experiment, random assignment, and enrollment (for those randomized into the treatment group) occur in rapid succession. A period of training then follows, with most enrollees finishing training in eight months or less. Training concludes with or without the enrollee in a job. At the conclusion of training, the training agency then has 90 days to graduate the enrollee. Approximately eighteen months after random assignment enrollees are interviewed and their labor earnings and employment histories over the previous 18 months are taken.[9]

We introduce some notation to model the optimal graduation strategy. Let g_i and f_i be enrollee i's graduation date and training end date respectively, and let e_i be the earliest date following f_i that individual i is employed (see Figure 1). Finally, let u_i be an indicator variable, equal to one if enrollee i failed to find a job within ninety days after f_i, and zero otherwise (note that u_i equals zero if enrollee i was employed as the 90 day grace period began). Using this notation, the optimal graduation strategy states that training agencies graduate (1) enrollees who are employed within the ninety day period on their first day of employment after training ends (that is, $g_i = e_i$ if $u_i = 0$) and (2) enrollees who were never employed within the 90 day period on the 90th day following the end of training (that is, $g_i = f_i + 90$ if $u_i = 1$). Thus, we have

$$g_i = (1 - u_i)e_i + u_i(f_i + 90). \tag{2}$$

To test this identity, we assume that participants report their training end date and employment date with the same error, and that the training agency reports the graduation

[9] We delete from the analysis a very small number of enrollee-observations whose follow-up interviews preceded graduation.

date with a random error that is stochastically independent of the participant's reporting error. The observed variables are u_i, $\tilde{g}_i = g_i + \eta_{g_i}$, $\tilde{f}_i = f_i + \eta_i$, and $\tilde{e}_i = e_i + \eta_i$, with η_{g_i} and η_i assumed to be stochastically independent. Taking the graduation date as a reference point, (2) can be rewritten in terms of the observed variables,

$$\tilde{g}_i - \tilde{f}_i = (1 - u_i)(\tilde{e}_i - \tilde{f}_i) + u_i 90 + \nu_i, \tag{3}$$

where $\nu_i = \eta_i - \eta_{g_i}$. To test the identity (3) we add coefficients and estimate

$$\tilde{g}_i - \tilde{f}_i = \beta_1(1 - u_i)(\tilde{e}_i - \tilde{f}_i) + \beta_2 u_i + \nu_i. \tag{4}$$

Because the residual is stochastically independent of the independent variables, we estimate (4) using least squares. Under the assumption that the error term is normally distributed, we can test whether the coefficients corresponding to $(1 - u_i)(\tilde{e}_i - \tilde{f}_i)$ and u_i, that is, β_1 and β_2, equal 1 and 90, respectively.

The estimate of Model I in Table 1 offers a simple comparison of the mean delay for enrollees who are employed sometime between training end and the end of the ninety day period (enrollees for whom $u = 0$) and enrollees who never become employed ($u = 1$).[10] (Model I is a regression of graduation delay on the variables $(1 - u)$ and u, without an intercept.) Model I shows that participants who are never employed in the ninety days following training end graduated 67 days later than those participants who obtain employment sometime during the ninety day period. Panel C of Table 1 shows the p value of the test of significance of this difference to be almost zero.

Model II in Table 1 corresponds to equation (4). The estimated coefficient on $(1 - u_i)(\tilde{e}_i - \tilde{f}_i)$, at .99, is close and statistically identical to one (the p value of the test of identity to one is .888). The coefficient on u_i equals 101.8 and is significant. Moreover, the

[10] The NJS data files contain 6444 adults with valid training agency-supplied enrollment and graduation dates who graduated in program years 1987 through 1989. Nevertheless, many of these participants failed to report in the NJS's participant survey having experienced a job training spell. Moreover, many who reported training spells supplied invalid beginning and ending dates. For some of these spells, however, we were able to impute dates. The subsample used in our graduation delay study described here contains, after imputing some training spell dates, 2327 persons. A detailed description of the construction of this subsample, including the imputation procedure, is available upon request. Table A.2 in the appendix reports the means and standard deviations of selected variables in the samples analyzed in this section.

coefficient estimate is statistically different from 90. Thus, on average, training agencies wait 101.8 days to graduate those enrollees who remain unemployed after completing their training. The observation that training agencies wait a little longer than 90 days to graduate some unemployed enrollees suggests that the 90 day constraint is imperfectly enforced.

One measure of the private gains from delaying the graduation of unemployed enrollees is the difference between the fraction of enrollees who are employed on the day they finish their training and the fraction employed on the day the training agency officially reports them. We find that the overall employment rate at graduation increases by 11.3 percentage points between these dates, from 47.0 percent to 58.3 percent. Stated differently, training agencies in this study would produce an employment rate outcome 20 percent lower if they were required to graduate enrollees (and report their performance outcomes) on the date they actually finish training.

4.2 Graduation Timing Toward Program Year-End

We generalize the simple graduation strategy presented above to take into account the training agency's discretion over the program year in which it graduates participants. Consider a stylized two program year incentive system where the training agency receives a fixed payment if the yearly labor market-based performance outcome exceeds a fixed performance standard. The training agency does not know its final aggregate performance outcome until the end of the program year because the labor market outcomes depend upon random factors, such as the state of the local economy, which are outside its control. Because of the graduation strategy described above, the training agency reaches the end of the year with an inventory of enrollees who have finished training within the previous 90 days but are unemployed. At the end of the first program year, the training agency chooses how many from this inventory to graduate in the present program year, the remainder to be graduated in the following program year. Assume there are n such persons, of whom n_1 will be graduated in the first program year and $n_2 = n - n_1$ in the next one. The training agency chooses n_1 to maximize the present value of the sum of the two awards.

The optimal graduation strategy on the last day of the first program year depends on the difference between the performance outcome and the standard as the last day arrives. Let $N = N_e + N_u$ be the number of persons who were graduated during the year (excluding the year's last day), where N_e and N_u are the numbers of such persons graduated employed and unemployed, respectively. Let $\bar{S}$ be the performance standard. Three cases can be distinguished (see Figure 2). In case one, on the last day of the year the cumulative performance outcome exceeds the standard by so much that the training agency can graduate all unemployed enrollees. In case one, because $\frac{N_e}{N+n} \geq \bar{S}$, $n_1 = n$. In case two, the cumulative performance outcome exceeds the standard, but not by much. In case two, because graduating all unemployed enrollees would push the outcome below the standard, it pays the training agency to graduate persons from its inventory only until the standard is bound. That is, the training agency chooses n_1 such that $\frac{N_e}{N+n_1} = \bar{S}$. Rearranging yields

$$n_1 = \frac{N_e}{\bar{S}} - N. \tag{5}$$

Equation (5) implies that n_1 lies between 0 and n, approaching zero when the training agency just meets the standard and n when the training agency outperforms the standard by n/N percentage points or more. In case three, the training agency fails to meet the standard at the end of the year, $(\frac{N_e}{N} < \bar{S})$. In this case, because it cannot win an award this year, the training agency 'takes a bath', graduating all n persons from its inventory to maximize the probability of an award next year.[11]

The two-period model outlined above suggests the following modification of (4).

$$\begin{aligned} \tilde{g}_i - \tilde{f}_i \;=\; & \beta_1(1-u_i)(\tilde{e}_i - \tilde{f}_i) + \beta_3 u_i(1-j_i) + \beta_5 u_i j_i LOW_i \\ & + \beta_6 u_i j_i MED_i + \beta_7 u_i j_i HIGH_i + \nu_i, \end{aligned} \tag{6}$$

where j_i is a dummy, equal to one if June 30th falls within the 90 day period following training and zero otherwise, and LOW_i, MED_i and $HIGH_i$ are three dummies, respec-

[11] Because the risks of not meeting a standard are substantial, the return on the strategy described above is likely to be high. For example, according to Department of Labor data on the entire JTPA system in 1988, 13.9 percent of training agencies did not achieve the standards for the employment rate at graduation.

tively equal to one if the performance outcome at the end of the program year lies in the low, medium and high regions defined by the horizontal axis in Figure 2, and zero otherwise.

We start by testing a very elementary prediction of the model. The model predicts that we should find that unemployed enrollees whose training ends in the last three months of the program year should be less likely to be delayed the full 90 days. As a consequence, we should find that the length of delay is longer on average for persons who finish training between July and March than for persons who finish between April and June. We test whether the length of delay differs in this way by imposing the constraint $\beta_5 = \beta_6 = \beta_7 = \beta$, estimating (6), and testing whether $\beta_3 > \beta$.

Models III in Table 1 tests this simple prediction and provide evidence consistent with the two-period model of the graduation decision. Model III reruns Model II, splitting enrollees who are not employed throughout the 90 day grace period (enrollees for whom $u = 1$) into two groups: one group contains enrollees whose grace period includes June 30, $j = 1$, and the other group contains enrollees whose grace period does not include June 30, $j = 0$. As predicted, we find the estimate of graduation delay is greater for the average person who finished training between July and March, compared to the average person who finished between April and June; the point estimates of graduation delay past the end of training is 113 days for the former group, and 79 days for the latter group. This difference is statistically different (the p value of the test of equality is .02). In addition, the R^2 rises slightly, from .150 to .153.

Confident that this simple implication is consistent with the data, we move to a more sophisticated prediction of the model. Among unemployed persons whose training ends within 90 days of June 30th, enrollees in LOW_i and $HIGH_i$ categories should be delayed for a shorter period on average than individuals in MED_i. In terms of (6), this corresponds to testing whether $\beta_6 > \beta_5$ and $\beta_6 > \beta_7$. In addition, individuals in LOW_i and $HIGH_i$ should be delayed approximately the same length of time, that is, β_5 should be equal to β_7.

To test this additional prediction, we create a subsample based on the graduation decisions for the subsample of training agencies in our data for which we had reliable estimates of year-end performance.[12] The model predicts that training agencies will delay graduating the average person whose training ends in MED for a longer period of time than the average person whose training ends in either HIGH or LOW. The estimates of Model IV are consistent with this prediction. That is, the estimated coefficient for the variable $u \times j \times MED$—we are omitting the i subscript to simply the notation—is greater than the estimated coefficients for the variables $u \times j \times HIGH$ and $u \times j \times LOW$. These differences are statistically significant (the p values for two-tail tests of equality are .030 and .024, respectively). Model IV shows that training agencies apparently exploit the ninety day grace period fully: the coefficient estimates for $(1-u) \times (e-f)$ and $u \times (1-j)$ are very close and statistically identical to one and ninety, respectively.

To summarize, Table 1 shows that a training agency (1) delays graduating idle, unemployed enrollees longer than idle, employed ones, (2) graduates idle, unemployed enrollees sooner if they finish in the last three months of the program year than if they finish within the first nine months of the program year, and (3) graduates unemployed enrollees who finish training in the last three months of the program year sooner if the training agency is doing either very well or poorly relative to the employment standard. These findings are consistent with the two period graduation model. We have also found timing behavior associated with wage-based performance measures. Because this behavior occurs on a much smaller scale than the behavior documented above, we describe this evidence in Appendix B.

[12] Although training agencies provided us with the complete incentives in place during the years 1987 through 1989, they were generally unable or unwilling to provide reliable estimates of their year end performance. Creating reliable estimates of year-end performance from our data is challenging. Because of the design of the NJS, a training agency's experimental population generally represented only a portion of the year's participants. To develop reliable estimates of year-end performance, we omitted training agencies from the estimation for which we did not have a substantial portion of the total year's participants. The estimates of Model IV are based on graduation decisions in training agencies for which 70 percent or more of participants graduated during the program year were represented among the NJS participants. For this reason, Model IV is estimated on a subsample of the data used to estimate Models II and III. We also ran Model IV on samples created using training agencies for which we had 60, 50, 40 and 30 percent of the year's population of graduates. We found that these samples generally produced statistically insignificant relationships, presumably because we were admitting sets of observations with ever more noisy measures of year-end excess performance.

5 Analysis of Gaming

The above results demonstrate that training agencies time the graduation date to increase their performance outcomes and to maximize their intertemporal award streams. In this Section, we test whether these timing activities reduce program efficiency.

A simple two-tasks model captures the main idea. Training agencies can invest in training and timing activities. Define a_1 and a_2, respectively, as investment in training and timing activities where investment will be interpreted broadly as including effort (e.g. caseworker time and attention) and the services themselves. We distinguish training activities, which include enrollment, training assignment, and any other activities directly related to the normal delivery of training, from timing activities, interpreted as year-end and graduation timing, which occur in addition to normal training.

Graduation timing activities include, among other things, the assessment of the employment state of enrollees who have completed their training and also the provision of 'quick fix' job referrals and placements, which training agencies offer to those individuals who after the main body of training ends are still not employed. A training agency's year-end timing activities include calculating how different graduation strategies affect its current year performance relative to each of the standards it faces, as well as forecasting how different graduation strategies influence its award prospects in the subsequent year. The important thing to note about timing activities is that, although we cannot exactly quantify these activities, we observe when they take place. Graduation timing occurs only during the 90 day window after training ends while year-end timing occurs only in June.

Training agencies maximize the rewarded objective, $M(a_1, a_2, s)$, which aggregates many dimensions of performance. The rewarded objective depends both on training and on timing activities, as well as on an exogenous incentive shifter s whose role will be explained shortly. We can restate the findings of Section 4 using this notation: training agencies invest in timing activities, that is, $a_2 > 0$, both for graduation and year-end timing, and that the measured objective responds positively to investment in timing activities, i.e. $\frac{\partial M}{\partial a_2} > 0$.

In this section we test whether timing activities are evidence of gaming. To do that, we need to define the principal's true objective. JTPA's stated mission is to increase the human capital stock of participants through job training (Section 106(a)). Following the job training evaluation literature, we measure job training's human capital impact by its impact on the enrollees' earnings. The principal wants training agencies to maximize the earnings impact, $I(a_1, a_2)$. This function depends on training activities and may or may not depend on timing activities.

Using performance incentives creates inefficient responses if the allocation that maximizes the true objective, call it (a_1^I, a_2^I), is different from the allocation that maximizes the rewarded objective, call it (a_1^M, a_2^M). When this happens, the agent will choose allocation (a_1^M, a_2^M) and the efficiency cost will be $I(a_1^I, a_2^I) - I(a_1^M, a_2^M)$. Ideally, one would want to estimate this difference but this is difficult in practice because one does not observe independent variations in the investment strategies (a_1, a_2) that would permit reconstruction of the function I.

The timing strategies allow us to approach the efficiency question differently. We propose two approaches to test for evidence of inefficiency. The first approach uses the concept of an incentive shifter introduced above. Assume we can identify an incentive shifter s that enters the rewarded objective M but not the earnings impact function I. This shifter changes the marginal products of a_i (for $i = 1, 2$), leading the training agency to change the resources allocation, i.e. $\frac{\partial a_i^M}{\partial s} \neq 0$ for $i = 1, 2$. The efficient allocation, however, should not depend on the incentive shifter, i.e. $\frac{\partial a_i^I}{\partial s} = 0$ for $i = 1, 2$. To demonstrate inefficiency, it is enough to show that $\frac{\partial a_i^M}{\partial s} \neq 0$ and $\frac{\partial I}{\partial a_i} \neq 0$ for some i. This test asks the question, does the incentive system distort the investment allocation?

While our first approach focuses on input distortions, our second approach focuses on output distortions. We suspect that the timing strategies lower the true objective. Evidence of this is evidence that timing activities constitute gaming. We ask the question: Are earnings impacts lower when the training agency invests more effort in timing activities? Our strategy uses the observed variation in the level of effort invested in timing activities. Assume that we can observe different regimes of investment in timing activities that are indexed by s where a higher value of s corresponds to a greater investment in timing activities. For example, the June month corresponds to a higher level of s than non-June months. We will conclude that timing activities are evidence of gaming if they lower the true objective, $\frac{d}{ds}I(a_1^M(s), a_2^M(s)) < 0$, while they are evidence of accounting manipulation if $\frac{d}{ds}I(a_1^M(s), a_2^M(s)) = 0$.

The rest of this section maps the two approaches outlined above. The next subsection tests if training is distorted towards the end of the program year. The following subsection tests whether graduation and year-end timing strategies lower the organizational output.

5.1 Does the Incentive System Distort the Training Allocation?

We focus on the distortion in the training allocation that occurs in June. June distorts the marginal products of training and timing activities. We showed in Section 4 that training agencies respond to this shift by graduating a disproportionate fraction of the enrollees in June. The training agencies graduate some idle enrollees who have finished training and are in their 90 day window. At the end of a program year, the training agency may also find that it is optimal to graduate a poorly performing enrollee in the current year even if doing so would mean cutting short the training plan.

We ask whether the training agency prematurely ends some training plans in June. Our identifying assumption is that under an efficiency hypothesis the training plan should not be truncated in June while it may be optimal to truncate a training plan in June under an award maximizing hypothesis. In terms of the model sketched out above, we use June as a proxy for the incentive shifter s and training length as a measure of investment in training activity a_1, and we ask whether June changes the training length, that is, whether $\frac{\partial a_1}{\partial s} \neq 0$. Finding that this is the case will be interpreted as evidence of inefficiency.[13]

We estimate a hazard rate model for the duration of training with a dummy variable indicating whether the current month is June as an independent variable. Our model of training posits that on the date of enrollment the enrollee's case worker assigns her

[13] This interpretation is correct as long as $\frac{\partial I}{\partial a_1} \neq 0$ holds where a_1 corresponds to the actual number of days in training.

to a training stream based on her characteristics. The duration of the training stream is function of its type, the characteristics of the enrollee, and state of the local labor market. Normally, the case worker would allow the training spell to take its course. In June, however, the case worker may decide to truncate the enrollee's training spell for the purposes of graduating her if her employment prospects are low and the training agency's year-to-date performance is either high or low relative to the standard. It is important to note that we run a hazard rate for the duration of training—not the duration of enrollment. The efficient training duration should not depend on the incentive shifter.

To test whether training agencies strategically truncate training spells we estimate a Cox proportional hazards model. The survival time in training for each enrollee i is assumed to follow its own hazard function, $h_i(t) = h(X_i, Z_i, T_i, JUNE_i(t), t)$, written as

$$h_i(t) = h_0(t)exp(\theta_0 + \theta_1 X_i + \theta_2 Z_i + \theta_3 T_i + \theta_4 JUNE_i(t)) \tag{7}$$

where $h_0(t)$ and θ are unknowns, and X_i is a vector containing the enrollee's age, education level, labor market experience, race, gender, marital status, number of children living at home, and welfare program participation, all measured upon enrollment. X_i also contains dummies indicating the enrollee's training agency. The state of the labor market over the post-random assignment period is described in the vector Z_i, which contains the unemployment rate, labor force participation rate, and relative size of the service sector, all measured locally and for the year(s) of enrollment. The enrollee's training type—whether it is on-the-job training, classroom training, or job search assistance—is denoted by T_i. $JUNE_i(t)$ is an indicator variable, equal to 1 if t is the month June, and equal to zero otherwise. Equation (7) is estimated using individual training spells, measured in months. $h_i(t)$ therefore is the probability that a training spell will end in month t given that it has lasted until month t. The baseline hazard, $h_0(t)$, corresponds to the case where all explanatory variables are equal to zero. The regressors, contained in the exponential component, scale the baseline hazard. Cox's semiparametric model is often used to analyze survival data when explanatory variables are time-varying, as is the case with our variable $JUNE_i(t)$.[14]

We estimate (7) using Cox's partial likelihood method, which eliminates the unknown baseline hazard and yields consistent estimates of the parameters. In the Cox proportional hazards model the coefficient estimates are the derivatives of the log of the hazard with respect to the explanatory variables. The coefficient estimates should be interpreted as the constant proportional effect of the explanatory variables on the conditional probability of the training spell ending. Finding that the estimated θ_4 coefficient is positive would be consistent with the hypothesis that those enrollees graduated in June have been prematurely graduated (that is, $\frac{\partial a_1}{\partial s} \neq 0$.)

Table 2 reports our results.[15] The first column contains our estimates of the coefficients in (7), assuming a constant baseline hazard across person types. The coefficient estimate is positive (.389) and significant. An important assumption of the Cox model is that the proportional effect of a regressor on the conditional probability of ending a spell does not depend on duration. Thus, for example, if the conditional probability of a black enrollee ending her training in the fifth month is twice that of a white enrollee, it is also twice the white enrollee's at ten months, and so on. The chi-squared test of the null hypothesis of constant proportionality is presented for each model estimated. The test for Model I shows that the proportionality assumption does not hold. Therefore, Model II estimates a stratified Cox proportional hazards model in which the baseline hazard, $h_0(t)$, is allowed to vary across groups of enrollees. In particular, this second specification allows the $h_0(t)$ to vary by training agency and training type, but constrains the coefficient estimates, θ, to be the same across training agency and training type. The results of this estimation are presented in the second column. Note that the coefficient estimates of Models I and II are qualitatively similar and that Model II does not reject the proportionality assumption. The coefficient estimate from Model II implies that the hazard rate in the month of June is about 46 percent greater than in other months.[16]

[14]See Kalbfleisch and Prentice (1980) or Kiefer (1988).

[15]Table A.3, columns 2 and 3, in the appendix reports the means and standard deviations of selected variables for the sample used in our analysis.

[16]Note that the ratio of the hazard in the month of June to any other month is simply $exp(\theta_4)$. Substituting $\hat{\theta_4} = .377$ for θ_4 yields our estimate of the impact of June on the hazard.

To refine our test that the incentive system generates input distortions, we cross the June variable with our measure of performance at the end of the year—LOW, MED, $HIGH$—constructed as in Section 4. As in Section 4, we view year-end performance as an exogenous shock outside the control of the agency, and thus a valid instrument for identifying the effect of performance incentives on training duration. The hypothesis we test is that enrollees are more likely to be graduated under LOW and $HIGH$. We are able to estimate year-end performance for a subset of June months (see Section 4.2 and footnote 12). Thus, for a subset of enrollees the propensity to truncate in June can be compared across the different states of performance. Define the variable $\tau_i(t)$, an indicator variable, equal to one if t is a June month for which year-end performance is observed, i.e., for which the variables LOW, MED, and $HIGH$ are non-missing. We assume $\tau_i(t)$ is exogenous because it is determined by the fraction of the training agency's total enrollment contained within the experimental sample. We estimate the following modification of (7)

$$\begin{aligned} h_i(t) &= h_0(t) exp(\theta_0 + \theta_1 X_i + \theta_2 Z_i + \theta_3 T_i + \theta_4 (1 - \tau_i) JUNE_i(t) \\ &\quad + \theta_5 \tau_i JUNE_i(t) LOW_i + \theta_6 \tau_i JUNE_i(t) MED_i + \theta_7 \tau_i JUNE_i(t) HIGH_i). \quad (8) \end{aligned}$$

These results are reported in the third column of Table 2. $JUNE_i(t)$ $HIGH_i$, $JUNE_i(t)$ MED_i, and $JUNE_i(t)$ LOW_i involve a comparison of the subsample of enrollees—much smaller that the samples used to estimate θ_4 in Models I and II—in training centers and years for which year-end performance could be reliably measured.[17] Note that $JUNE_i(t)$ $HIGH_i$ is significant and positive as predicted, while the coefficient estimate for $JUNE_i(t)$ MED_i is not significantly different from zero, again, as predicted. The implication of these estimates is that high performance makes an enrollee twice as likely to exit training in June, relative to performance just above the performance standard. The coefficient estimate for $JUNE_i(t)$ LOW_i is not significant but this may be due to the small subsample of enrollees receiving training in LOW training centers.

[17]The sample contains 133 June enrollment months, with 26 occuring in the LOW state, 38 in the MED state, and 61 in the HIGH state.

Thus, the evidence suggests that the incentives encourage training agencies to truncate training spells. We argue that this finding is not wholly due to unmeasured differences among enrollees whose training carries into June because we can show that whether training agencies truncate training or not depends in a predictable way on their performance in June.

5.2 Efficiency Cost of Timing Activities

While the previous subsection demonstrates that timing activities distort the input allocation it does not shed much light on the efficiency cost of these distortions. In this section, we present additional evidence of efficiency costs and we also attempt to measure some of these costs. We will follow the approach presented in the model sketched out above, estimating $\frac{d}{ds}I(a_1^M(s), a_2^M(s))$ both for graduation and year-end timing.

Is graduation timing evidence of gaming?

To test whether graduation timing is evidence of gaming, we use the fact that graduation-timing activities occur during the 90 day window after training ends. We can measure for each enrollee the amount of time between training end and graduation. This amount of time is a proxy for the investment in graduation timing. We hypothesize that earnings impacts should be lower the greater the training agency's investment in graduation timing, which we measure at the training agency level. To be more precise, we construct for each training agency the variable $DELAY$, the number of days graduation is delayed for the median enrollee in the training agency.[18] Across training agencies, the median number of days of delay ranges from 0 to 106. The variable $DELAY$ varies across training agencies because different training agencies face different incentive schemes, applicant pools, labor market opportunities and training technologies. We hypothesize that training agencies respond to differences in their external environment by making different trade-offs in allocating effort across training and timing activities. We use the variable $DELAY$ as a proxy for the investment in graduation timing.

[18] We use median instead of mean because the median is less sensitive to outliers.

Consider the typical production function for an enrollee's earnings used in the evaluation literature,

$$Y = f(X, Z, T, D, DELAY, \nu), \tag{9}$$

where Y measures the enrollee's labor market earnings for the eighteen months following random assignment, which corresponds closely to the start of the enrollee's enrollment spell. (To keep the notation simple, we have suppressed the subscript identifying these variables as measures for a single enrollee.) It is important to note that the starting and ending dates for the measurement of the enrollee's earning are independent of the graduation decision (see Figure 1). This independence rules out the possibility of endogeneity bias where the graduation decision determines earnings impacts indirectly through the measurement procedure.

The enrollee's earnings over the eighteen month period are assumed a function of her demographic characteristics, X, the state of the labor market over the post-random assignment period, captured in Z, the type of training she receives, T, as well as its duration, denoted D.[19] If gaming activities are costly, an enrollee's earnings is also a function of the degree to which training staff devotes time to gaming, captured by the variable $DELAY$. Finally, ν is an error term representing the unobservable determinants of an enrollee's earnings. We assume that $f(\cdot)$ is linear, that is

$$Y = \gamma_0 + \gamma_1 X + \gamma_2 Z + \gamma_3 T + \gamma_4 D + \gamma_5 DELAY + \nu, \tag{10}$$

and we estimate the parameters, γ, of (10) using least squares and the data on all the adults in all training agencies, assuming that ν is distributed i.i.d. normal. Under the gaming hypothesis we conjecture that the coefficient on $DELAY$ is negative ($\gamma_5 = \frac{d}{ds} I(a_1^M(s), a_2^M(s)) < 0$).

[19]Earnings might rise with the training duration because the longer she is in training the more skills she acquires. Earnings may also fall with the length of training: for some enrollees, longer training spells mean that they enter the labor market and begin earning wages late in the eighteen month post-random assignment period. Because the training she receives is non-randomly assigned, her training type and its duration may also reflect non-measured human capital. For example, enrollees may have been assigned to especially lengthy training spells because they enter the program with few marketable skills. Thus, the length of training and the earnings may be negatively related, holding all other measurable variables constant.

The estimates in Table 3 support this hypothesis.[20] Model I in Table 3 corresponds to equation (10). The estimated coefficient on $DELAY$ is statistically significant and implies that a one day increase in the number of days of delay for the median enrollee reduces the average earnings gain by about $47; a one standard deviation (29.9) increase in delay reduces the gain by about $1406. Because a training agency can enroll thousands of enrollees, a thirty day increase in the delay experienced by the median enrollee can represent a substantial increase in delay activities. Nevertheless, the estimated impact of delay is quite large, larger than most estimates of the average impact of training.[21]

While these results are consistent with our hypothesis that delay strategies take resources away from training efforts, they are also consistent with another interpretation based on agency heterogeneity. According to that story, training agencies have idiosyncratic preferences for awards and these preferences ultimately determine both earnings impacts and graduation timing.[22] In this case, $DELAY$ may be correlated with unobserved behavior that makes earnings impacts low. For example, training agencies that pursue delay strategies may also *cream-skim*, that is, enroll persons whom they expect will perform well on the JTPA performance measures, even though they experience low earnings gains from job training.

Our empirical models are estimated with enrollee and training characteristics on right hand side to control for differences in agency strategies. These control variables, however, may not soak up all variation in agency preferences. Research in the policy field has emphasized the importance of the job training agency's mission in determining its behavior (see Grubb and McDonnell, 1996, and Heinrich, 1999). In JTPA, some training agencies are organized as private non-profits, others are part of local government, and still others are run by the local Private Industry Council (PIC), a JTPA-mandated overseeing board composed of local business, labor, and government leaders. These organizational differ-

[20] Table A.3, column 2, in the appendix reports sample statistics for key variables used in this analysis.

[21] Large estimated effects on per capita earnings of features of the JTPA incentive system have been reported before. Cragg (1997) reports that states that allow their training agencies to appeal awards—an indicator of a philosophy of leniency—show per capita (lifetime) earnings among JTPA enrollees between $2000 and $2200 less than states that do not, *ceteris paribus*.

[22] Note that we cannot control for agency idiosyncratic preferences in specification I and II by adding agency dummies because these dummies would be co-linear with $DELAY$ and the labor market variables.

ences, it is argued, affect the types of training strategies pursued and may also therefore determine the taste for performance awards.[23] Among the sixteen training agencies in our study, 8 are parts of government agencies, 2 are private non-profit organizations, 3 are run by PICs, and 3 fall under the category "other". Model II in Table 3 includes on the right hand side indicators for organizational type.[24] Note that including the organizational type of the training agency leaves the estimated effect of $DELAY$ on earnings little changed.

Thus, training agencies that show high levels of graduation timing show lower earnings gains. This finding remains even after we control for heterogeneity in training, enrollment strategies, and organizational types. This is consistent with the hypothesis that graduation timing is evidence of gaming.

Is year-end timing evidence of gaming?

The month of June modifies the training agency's training strategy in two ways. First, training agencies invest more in timing activities in June than in other months. Training agencies substitute time and effort away from training toward timing. Using the two-task model, we have $a_2^{s_1} \geq 0$ and $a_2^{s_0} = 0$ where a_2 is interpreted as the investment in year-end timing and s_0 (s_1) indicates whether the variable is measured in June (a month other than June). Under the hypothesis that timing is evidence of gaming, we should find that training that takes place in June is less effective than in other months. Second, and as demonstrated in the previous subsection, training agencies truncate some training spells in June. We call these two effects the substitution effect and the truncation effect.

To measure the substitution effect, we construct $JUNFRAC$, which is equal to the fraction of the enrollee's training spell spent in the month of June. For example, for an enrollee who started training on April 1, 1987 and finished on June 30, 1987, $JUNFRAC$

[23]For example, a non-profit may be a community-based organization or one that is closely allied with a community based organization, and therefore narrowly focused on serving a particular subgroup of the local eligible population. PICs, because local businesses make up half of their membership, may be more interested in using JTPA to identify promising workers or to subsidize their workforce.

[24]The source of this information is a data set collected by SRI, Inc to evaluate for the National Council for Employment Policy the efficacy of performance standards in JTPA. See Dickenson, *et al.* for a description of these data.

is one third. $JUNFRAC$, therefore, measures the extent to which enrollees may have suffered from year-end timing. To distinguish the impact of sudden truncation in June on earnings impact, we construct the dummy $JUNTRUNC$ which is equal to one if the enrollee is trained and graduated in June, and equal to zero otherwise. We estimate the following modification of (10)

$$Y = \gamma_0 + \gamma_1 X + \gamma_2 Z + \gamma_3 T + \gamma_4 D + \gamma_5 DELAY + \gamma_6 JUNFRAC + \gamma_7 JUNTRUNC + \nu \quad (11)$$

using least squares.[25] The coefficient on the variable $JUNFRAC$ captures the change in earnings that results from a different investment in year-end timing independent of a truncation effect, that is, $\gamma_6 = I(a_1^{s_1}, a_2^{s_1}) - I(a_1^{s_0}, a_2^{s_0})$, measured for a representative enrollee. Under the gaming hypothesis, we conjecture that those enrollees trained in June have lower earnings, that is $\gamma_6 < 0$. The spirit of this test is that the investment in year-end timing activities occurs only in the month of June. Because year-end timing occurs only in the month of June, it is more likely to lower the human capital impact of those enrollees whose training spells overlap with June.

It is more difficult to identify the truncation effect than the substitution effect. The reason is that training agencies truncate in June those enrollees who have little chance to find an employment and they base this decision on a much finer set of variables than the set that is available to us. One could argue, for example, that the variable $JUNTRUNC$ is subject to a selection bias because those individuals graduated in June are more likely to be poor performers.[26] Although the variable $JUNTRUNC$ only imperfectly captures the truncation effect, it permits us to control for potential selectivity bias in our estimation of the substitution effect.[27]

[25] We use only the enrollees who report receiving training over the enrollment spell to estimate (11). See footnote 10.

[26] Because $JUNTRUNC$ identifies persons whose training and enrollment spells both end in June, the estimation of γ_7 is subject to another kind of bias. A large fraction of this subset of enrollees have had their training spells curtailed. The training spells of the remainder, however, ended naturally. The presence of non-truncated spells among the truncated spells will bias towards zero the coefficient γ_7.

[27] One concern with the $JUNFRAC$ variable is that it oversamples enrollees who are graduated in June. Recall that June graduates perform worse on the performance measures. This suggests that the $JUNFRAC$ variable may pick up unobserved skill differences across individuals that are correlated with earnings impacts. Controlling for $JUNTRUNC$ takes care of this selection bias.

The results of our estimation of equation (11) are reported as Model III in Table 3. As in models I and II, the estimates of the coefficients on $DELAY$ are negative and significant. The estimate of the substitution effect (γ_6) is also negative and significant. This is consistent with the substitution hypothesis. The estimate of the γ_7 is negative and insignificant. The finding that γ_7 is insignificant could be explained by the fact it may difficult to find earnings differences due to the truncation effect within 18 months after random assignment.[28]

The negative year-end effect shows that those enrollees who are trained in June do not gain as much from JTPA training. This is consistent with our hypothesis that year-end timing diverts effort and resources from human capital activities. The $JUNFRAC$ coefficient estimate, however, may represent, partly or wholly, a seasonal effect. In June, July, and August the labor supply in low-skill labor markets temporarily shifts out as high school and college students seek temporary, full-time employment. In 2000, a typical year, the labor force among persons aged 16-19 rose from 7.1 million in May to 8.3 million in June. In July and August, the labor force stood at 8.6 million and 8.1 million, respectively, before settling back to 6.8 million in September, the beginning of the new school year.[29] The increased competition for jobs in the summer months may depress the wages and/or employment of JTPA enrollees who leave training for the low-skill labor market. In our sample, of those enrollees whose training spell included June ($JUNFRAC > 0$), 40 percent terminated in one of the summer months, compared to just 23 percent of those enrollees whose training spell did not include June ($JUNFRAC = 0$). The June variable oversamples enrollees whose enrollment concludes in the summer months.

To control for the summer recess effect, we re-estimate equation (11) including on the right hand side $SUMMFIN$ (it replaces $JUNTRUNC$), an indicator variable equal to one if the enrollee finishes training in June, July, or August, and equal to zero otherwise.

[28]The finding that truncation does not change earning impacts does not mean that truncation is not inefficient. It is correct to conclude that truncation is inefficient because truncation changes the cost of training. Stated differently, the condition in footnote 13 holds once the function I is redefined to include the cost of training.

[29]These figures represent labor force participation rates among 16-19 year-olds of .51, .60, .62., .58, and .49 for the months May, June, July, August, and September, respectively. The source of these figures are the Current Population Survey.

$SUMMFIN$ controls for the negative impact on earnings of the summer recess. The results of the estimation of equation (11) are reported as Model IV in Table 3. Our estimate of γ_6, now uncontaminated by the summer recess effect, remains negative and is now marginally significant.[30] Note that the calendar effect while negative, is small and statistically insignificant. The implication of the more conservative estimate of γ_6, -3553.9, is illustrated by the example of an enrollee whose training lasts five months (roughly the average length of a training spell) and ends on June 15. The coefficient estimate implies that if the enrollee had begun her 5 month training spell 15 days earlier, so that her training ended not in the middle of June but on May 31, her 18 month earnings gain would have been $355 greater. To put this in perspective note that the point estimate of the 18 month earnings impact from the National JTPA Study ranges from $543 to $898.[31] That is, by beginning her training earlier so that it ended before June, this enrollee would likely enjoy a 36 to 60 percent increase in her 18-month earnings gain from training.

To summarize, the evidence suggests that timing strategies are evidence of gaming. The evidence that earnings impacts are lower in those training agencies that engage more in graduation timing is consistent with the hypothesis that graduation timing is inefficient. In addition, we find that year-end timing is inefficient on two counts. First, training agencies are more likely to suddenly truncate training in June; an input distortion which is a direct consequence of year-end timing. Second, we find that earnings impact are lower for those enrollees who receive training in June. We interpret this finding as evidence that training agencies substitute time and effort away from training toward the end of the program year.

6 Summary

This paper studies timing responses to incentives. We find that JTPA training agencies time the reporting of the trainee's performance outcomes to maximize their incentive

[30] We also ran a regression with both $JUNTRUNC$ and $SUMMFIN$ and the results were similar.

[31] These numbers are based on the figures in Exhibit S.1 of Bloom, et al (1993). The 18-month earnings gain is the difference in mean earnings in the 18 months following the date of random assignment between persons in the control and treatment groups. See Bloom, et al for details. See also footnote 21.

awards. We show that training agencies report good training outcomes promptly but wait on bad ones in the hope that they improve. In those cases where bad outcomes are unlikely to improve, training agencies report these outcomes only in good years where they do not risk losing their awards. In bad years, they postpone reporting bad performances until the next year to secure the minimum level of performance required to qualify for an award.

We formally test for the impact of these timing responses on the efficiency of the organization. We find that the timing responses in JTPA lower the effectiveness of job training, as measured by the impact of training on enrollee earnings. This efficiency test suggests that this timing behavior is evidence that training agencies game the incentive system. This evidence of gaming demonstrates that there are costs to using explicit performance incentives.

We conclude with a note on our perspective as to where our work fits within the incentive literature. This paper presents evidence that is consistent with the hypothesis that performance incentives in organizations lead to costly distortions in agent behavior. One could argue, however, that the costs we identify are specific to a poorly designed incentive system, and that these costs could be easily reduced or even eliminated under alternative designs. Whether these costs can easily be eliminated, however, is entirely an empirical issue that cannot be addressed until one observes and studies the responses to these alternative designs. Only by studying the responses to alternative designs can we tackle the fundamental issue of whether incentives are second-best efficient. Unfortunately, our data do not allow us to address this second-best question. Keeping this broader perspective in mind, this paper constitutes the first step toward a more ambitious efficiency analysis of the use of incentives in organizations.

References

[1] 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.

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

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

[4] 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.

[5] Brown, K., Harlow, W., and Starks, L. (1996). Of tournaments and temptation: An analysis of managerial incentives in the mutual fund industry. *Journal of Finance*, pages 85–110.

[6] Chevalier, J. and Ellison, G. (1997). Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy*, 105(6), pages 1167–1200.

[7] Courty, P. and Marschke, G. (1997). Measuring Government Performance: Lessons from a Federal Job-Training Program. *American Economic Review*, 87(2):383–388.

[8] Courty, P. and Marschke, G. (2000). The JTPA Incentive System: Program Years 1987-1989. In *Performance Standards in a Government Bureaucracy: Analytic Essays on the JTPA Performance Standards System*, (J. Heckman, ed.), Upjohn Monograph Series (forthcoming).

[9] 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.

[10] Dickert-Conlin, S. and A. Chandra. (1999) Taxes and the Timing of Births. *Journal of Political Economy*. 107, (1): 161-177.

[11] 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. Manuscript, National Commission for Employment Policy, Research Report No 88-16.

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

[13] Gibbons, R. (l997). Incentives and Careers in Organizations. In, Kreps and Wallis editors, *Advances in economics and econometrics: Theory and applications: Seventh World Congress* Volume 2. Econometric Society Monographs, no. 27, pages 1-37. Cambridge University Press.

[14] Goolsbee, A. (2000). What Happens When You Tax the Rich? Evidence from Executive Compensation, *Journal of Political Economy*, 108(2): 352-78.

[15] Grubb, W. N. and McDonnell, L. M. (1996). Combatting Program Fragmentation: Local Systems of Vocational Education and Job Training, *Journal of Policy Analysis and Management*, 15(2): 252–270.

[16] Healy, P. (1985). The Effect of Bonus Schemes on Accounting Decisions. *Journal of Accounting and Economics*, 7:85–107.

[17] Heckman, J. J. and Smith, J. A. (July, 1995). The Performance of Performance Standards: The Effects of JTPA Performance Standards on Efficiency, Equity and Participant Outcomes. University of Western Ontario, Papers in Political Economy: 73.

[18] 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. NBER Working Paper: 5535.

[19] Heinrich, C. (July 1999) Do Government Bureaucrats Make Effective Use of Performance Management Information? *Journal of Public Administration Research and Theory*, 9(3): 363–393.

[20] 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.

[21] Jensen, M. C. and Meckling, W. H. (1992). Specific and General Knowledge and Organizational Structure. In Werin, L. and Wijkander, H., editors, *Contract Economics*, chapter 9, pages 252–71. Cambridge, Mass. and Oxford: Blackwell.

[22] Johnston, J. W. (1987). *The Job Training Partnership Act: A Report by the National Commission for Employment Policy.*

[23] Kerr, S. (1975). On the folly of rewarding a, while hoping for b. *Academy of Management Journal*, 18:769–83.

[24] Lawler, E. E. (1982). *Pay and Organizational Strategies and Pay Systems.* Jossey-Bass, San Francisco.

[25] Marschke, G. (2000). Performance Incentives and Bureaucratic Behavior: Evidence from a Federal Bureaucracy. Manuscript, University at Albany.

[26] Murphy, K.J. (1998) Performance Standards in Incentive Contracts. Manuscript, Marschall School of Business, University of California.

[27] 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.

[28] Oyer, P. (1998). The Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality *Quarterly Journal of Economics*; 113(1), pages 149–85.

[29] Prendergast, C. (March, 1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1).