Curbing cream-skimming:
Evidence on enrolment incentives

Pascal Courty, Do Han Kim, and Gerald Marschke1

September 2010

Abstract: Using data from a large, U.S. federal job training program, we investigate whether enrolment incentives that exogenously vary the ‘shadow prices’ for serving different demographic subgroups of clients influence case workers’ intake decisions. We show that case workers enroll more clients from subgroups whose shadow prices increase but select at the margin weaker-performing members from those subgroups. We conclude that enrolment incentives curb cream-skimming across subgroups leaving a residual potential for cream-skimming within a subgroup.

JEL: H72, J33, L14.

Keywords: Performance measurement, cream-skimming, enrolment incentives, bureaucrat behavior, public organizations.

---

1Pascal Courty, Department of Economics, University of Victoria, Canada, email: pcourty@uvic.ca; Do Han Kim, Department of Public Administration, Kyung Hee University, 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Korea, email: dohankim@khu.ac.kr; and Gerald Marschke, Department of Economics, University at Albany, State University of New York, , Albany, NY, 12222, NBER and IZA, email: marschke@albany.edu. Marschke gratefully acknowledges the hospitality of the Kennedy School of Government and the Labor and Worklife Program at Harvard Law School, where he was visiting while this paper was completed. We thank participants at the 2008 “Vouchers, Contracting-out and Performance Standards” Conference in Nuremberg and the Fifth IZA Conference on Labor Market Policy Evaluation, October 2009, in Washington D.C.
1-Introduction

The recent introduction of performance incentives in several branches of the public service sector, such as job training, education, and health, has raised concerns as to their impact on who is chosen to receive services, and on the equity and efficiency of service delivery. At the center of this debate is whether incentives induce service providers to cream-skim, that is, to select recipients on the basis of performance on measured outcomes instead of value added according to the organization’s stated objectives (Anderson, Burkhauser, and Raymond, 1993, Iezonni et al. 2003, Clotfelter and Ladd, 1996). One proposal for reducing cream-skimming is to adjust the measures that are used to assess performance, effectively setting different ‘shadow prices’ for different distinguishable subgroups of recipients. Although such methods are used in practice, and their theoretical underpinning is uncontroversial, there is no evidence that adjustment models actually have an impact on service delivery.

Our evidence comes from the large U.S. federal training program created under the Job Training Partnership Act (JTPA), which, between 1982 and 2000, provided job training to economically disadvantaged Americans. A possible concern is that focusing on a single training program in a single country comes at the cost of generality. But several reasons justify this choice. The problem of cream-skimming is not specific to JTPA but has been documented in other training programs and also public sector programs in education and health. The JTPA response to cream-skimming (the adjustment method) raises the same question that has been discussed in other contexts, namely, ‘do agents respond to the large number of shadow prices that are required to address cream-skimming?’ In fact, the simple theoretical model we develop to capture
agents’ responses to adjustment models is general and its predictions could be investigated elsewhere. We further develop these points in the rest of this paper and establish connections between our case study and other applications of adjustment methods elsewhere.

From an empirical point of view, JTPA is a natural choice since it was one of the earliest and most extensive adopters of adjustment methods. Most importantly, it offers a unique opportunity to the researcher because it has changed the adjustment method used to assess performance three times during our sample period, using four different sets of shadow prices. We use this variation to produce the first econometric evidence on whether it is possible to influence job training case worker intake choices. Under JTPA, job training services were administered by over 620 semi-autonomous sub-state training agencies each evaluated according to a set of performance measures defined at the federal level. Specifically, a training agency’s yearly performance was adjusted upwards or downwards to account for the particular mix of persons the agency enrolled. To illustrate, consider the adjustment made to the employment at termination measure for enrolling adults who never received a high school degree.\(^2\) By enrolling more high school dropouts a training agency lowered the minimum performance (the minimum fraction of participants employed at termination) necessary to avoid sanctions and qualify for a performance award. We refer to this minimum performance threshold as the performance standard. The adjustment to the standard for enrolling high school dropouts varied over time. We test whether case workers respond to the changes in these

\(^2\) The employment at termination measure, the most important measure in the early years of JTPA, was defined as the fraction of program terminees who terminated with a job.
adjustments. We quantify the impact of the adjustment method both on intake populations and on performance outcomes.

There are good reasons to think that JTPA’s adjustment methods may not change enrolment decisions. First, case workers’ preferences may vary over socio-economic subgroups, or case workers may be subject to pressures by local influence groups that override the typically weak incentives backing the adjustments (Heckman, Smith, and Taber, 1996). Second, Heckman and Smith (2004) have shown that most of the selection occurs at the early stages of the participation process, such as between eligibility and awareness, over which the program staff has little or no control. Thus, even if case workers respond to changes in the shadow prices, their response may be negligible. Third, adjustment methods may have little impact in practice because they are complex. In our case study, for example, the adjustment model can potentially distinguish over 16 million different demographic subgroups for each of four different performance measures. It may be impossible, or not worthwhile, for a training agency to attempt to factor into its enrolment strategy so many ‘shadow prices’.

Our empirical approach leverages two changes to the adjustment weights in JTPA that are exogenous to the training agencies’ enrollment decisions. We establish two sets of results. First, we find that changes in the incentive for enrolling members of a subgroup significantly change the fraction of enrollees from the subgroup. Because this result holds for the average agency and the average subgroup, we can conclude that the adjustments had a significant impact on enrollment. Second, we demonstrate the existence of within-subgroup heterogeneity. Case workers increase the number of

---

24 different adjustment factors have been used during our sample period (see Table 1, and the later discussion). Each adjustment factor takes binary values implying $2^{24}$ different subgroups.
enrollees from a specific subgroup by enrolling at the margin applicants that perform worse on the measure. This finding is consistent with the cream-skimming hypothesis that case workers use their private information about the eligible population to select enrollees that perform well on the performance measures. In contrast with the literature, which focuses on the impact of incentives on overall enrolment at the training agency level (differences across subgroups), we demonstrate that private information carries through even within demographic subgroups.

**Literature**

Our results are of interest to economists and policy makers for two main reasons. First, this work contributes to the literature on the effectiveness of incentives in the public sector. Many policy analysts now believe that such systems can improve accountability and management (Osborne and Gaebler, 1992; Gore, 1993) and such systems have become policy in many developed countries. Our findings can help in the assessment and design of current and future adjustment systems. In the workforce area in the U.S., for example, the Workforce Investment Act (WIA), which supplanted JTPA in 2000, continued to reward training centers according to outcome-based measures of performance but eliminated regression-based adjustment of performance standards. In 2010, after policy-analysts had expressed concern that the absence of adjustments was encouraging cream skimming (Barnow and Smith, 2004; U.S. Government Accounting Office, 2004; Heinrich, 2004; Barnow and Heinrich, 2010), WIA began testing regression

---

4 Greater use of performance measurement systems in the public sector has also received support in academic circles (National Academy of Public Administration, 1991; Wholey and Hatry, 1992; Bouckaert, 1993; Kravchuk and Schack, 1996).

5 In the U.S., Government Performance and Results Act (GPRA) of 1993 which requires federal agencies to formulate measures of performance and set performance goals to improve public accountability and permit scrutiny by congressional oversight committees and the public.
models for its performance standards with the goal of fully implementing regression based adjustments beginning in July 2011 (U.S. Department of Labor, 2010).

Adjustment models have been used in job training outside the US and also in non-workforce environments such as education and health, when, for example, school performance and scorecard, respectively, have been used. Because the rationale for using adjustment methods, the type of adjustment methods used, and the debates around cream-skimming, are similar in JTPA and in these other applications, our results are relevant beyond JTPA. For example, Job Services Australia contracts with external vendors and judges their success obtaining employment for their enrollees using a regression-based model to adjust the provider's performance by the characteristics of persons served (Bruttel, 2005). In the UK, in Jobcentre Plus awards are allocated on the basis of a point system, whereby points are accumulated based on job placements, and the points received per job placement vary by subgroup. For example, the placement of a jobless single parent earns 12 points, compared to only 1 point for an already employed person (Burgess, Propper, Ratto, and Tominey, 2003).

In education, these schemes are used to account for differences across students in assessing teachers’ value-added. The purposes of these adjustments are to limit the penalties to schools which serve disadvantaged communities but also to reduce the distortions to the composition of students taking the test (Clotfelter and Ladd, 1996; Ladd, 1999), e.g., by increasing special education placements. In assessing the quality of hospitals they are also used to control for differences in patient populations (e.g., Iezonni, Ash, Shwartz, Daley, Hughes, and Mackiernan, 2003). In Germany, sickness funds (health insurance providers) are prohibited from charging premia based on the health
characteristics of the insured. To attenuate the incentive to cream skim, income is transferred across funds based on the characteristics of the insured and the average costs associated with these characteristics as predicted by historical data (Buchner and Wasem, 2003).

As prevalent as these adjustment models are, there’s scant work on their impacts. One contribution of this paper is to provide some evidence—the first that we know of—that bureaucrats respond to adjustment models and to quantify these responses. In fact the literature has repeatedly pointed out the difficulties in separating bureaucrat and applicant motives in explaining participation (Heckman and Smith, 2004). Our evidence circumvents this challenge by using a natural experiment that permits one to identify the relation between the adjustment model and enrolment choice.

Second, our evidence sheds new light on the literature attempting to identify cream-skimming when adjustment models are in use. In the context of JTPA, most studies (Heckman, Smith, and Taber, 1996; Cragg, 1997; Heckman and Smith, 2011) show a modest degree of cream skimming at best, suggesting that incentives have at most a modest effect on enrolment decisions. But these studies have focused on the enrolment incentives due to performance measurement without factoring in the role of the adjustment weights in the performance standard. They test variants of the following hypothesis: Does rewarding (or sanctioning) a training agency based on the fraction of its clients who obtain employment dissuade it from serving high school dropouts and other persons with poor labor market prospects? But the JTPA adjustment model forces the training agency to consider how a person’s attributes not only affects the performance

---

6Indeed, Heckman et al. (1996) find that JTPA case workers were more likely to enroll the applicants with the lowest prospects for employment after training. Heckman et al. call this behavior “cream avoidance” which they attribute to a “social worker mentality” in training agency staff.
outcome but also the standard: the agency knows that enrolling a high school dropout lowers its employment outcome but it also lowers its standard. Thus, the absence of strong evidence of cream-skimming may be because the performance standard adjustment procedure was doing its job—that is, reducing the incentive to cream-skim—and not because JTPA case workers did not respond to incentives. Our study tells a more complex picture where both effects are at play: local agencies respond to enrolment incentives but cream-skimming still takes place due to unobservable characteristics within demographic subgroups.

The paper proceeds as follows. Section 2 describes how performance adjustment was implemented in JTPA. Section 3 derives our predictions on enrollee population and performance outcomes. Section 4 tests these predictions using micro-level data in JTPA, leveraging two exogenous changes in the PAW.

2-Performance Adjustment Weights: Background and Case Study

Much of the literature on PAW has focused on their use in correcting enrolment distortions due to the introduction of outcome based performance incentives.\(^7\) Performance incentives stimulate agency efforts to produce value added, but they may also change the population served. This problem has emerged with incentive schemes in education that measure school performance using standardized test scores (Haney, 2000; Jacob, 2005), in job training that evaluate performance using labor market outcomes of trainees (Heckman et al., 2002; Finn, 2009), in German health insurance markets (Buchner and Wasem, 2003) and in health care where doctors, hospitals, and nursing

\(^7\) PAW can also be used in the absence of outcome based incentives, and the point would then be to correct possible bias due to bureaucratic preferences. In fact, policy makers may reward bureaucrats for enrolling certain groups if they feel that these groups would be otherwise underserved.
homes are evaluated using “report cards” (Dranove, Kessler, McClellan, and Satterthwaite, 2003; Bevan and Hood, 2006; Mukamel et al., 2009).

The literature on PAW has been mostly conceptual or prescriptive in nature. Rubenstein, Schwartz, and Stiefel (2003) and Brooks (2002) lay out rationales for adjusting performance standards, compare and contrast different adjustment strategies, and offer recommendations to policy-makers on how to adjust standards. Courty, Heinrich, and Marschke (2005) situate the problem in the principal-agent framework, and discuss how performance outcome measures should be adjusted. Another strand of the literature documents how PAW have been used in the context of specific applications. Trott and Baj (1987), Barnow (1992), Heinrich (2004), and Barnow and Smith (2004) discuss applications to job training programs, Siedlecki and King (2005) to workforce development programs, Berne (1989), Stiefel, Rubenstein, and Schwartz (1999) and Stiefel, Schwartz, Rubenstein, and Zabel (2005) to education, and Iezzoni (2003) to healthcare. Yet another strand compares the effect of different adjustment techniques on the ranking of service providers (e.g., Koenig, Fields, Dale, Ameen, and Harwood, 2000).

Case study: JTPA

A large literature discusses various aspects of the JTPA program (e.g. Johnston, 1987), and offers descriptions of its incentive system (e.g. Courty and Marschke, 2003) and the bureaucratic responses they induce (e.g. Heckman, Heinrich, Courty, Marschke, and Smith, 2011). To reduce unnecessary repetition, we present here only those features of the organization that are essential to our analysis of PAW, and direct the reader to Appendix 1 for more comprehensive sources when required.
The JTPA program was highly decentralized: its 620 plus local training agencies administered the program with significant discretion over whom to enroll. While applicants had to meet an income test to be eligible to receive services, JTPA was not an entitlement. Given the JTPA annual budget (approximately $4.1 billion in 1993), and the large population that was eligible for training, agencies could serve only one to three percent of the eligibles (550,000 new participants were enrolled in 1993). The decision of which eligibles to enroll constitutes the focus of this paper.

The JTPA fiscal year, or *program year*, ran from July 1 to June 30 of the following calendar year. Our empirical analysis focuses on program years 1993-1998 and on the adult JTPA population. By 1993, the DOL had been using for several years four performance measures constructed from two labor market outcomes, employment and earnings, to evaluate training agencies. A training agency’s employment rate at follow-up (ER) for a particular program year was calculated as the fraction of enrollees terminated during that year who were employed 13 weeks after termination. The average weekly earnings (WE) was calculated as the average weekly earnings during the ninety days following termination for those enrollees who were employed 13 weeks after termination. From the ER and WE outcome, two performance measures were constructed: one based on the performance of all adult enrollees and another based only on the welfare-receiving subset of adult enrollees. Each measure had associated to it a separate standard. The DOL set lower standards for the welfare versions of the measures.

For each of the four standards, the DOL developed an adjustment model to account for the particular agency’s enrollee choices (demographic characteristics of the enrollee pool) and local labor market circumstances (socio-economic conditions outside the

---

8 See Dickenson et al. (1988) for a complete description of the JTPA eligibility rules.
control of the agency). For example, it was determined that training agencies that enrolled few high school dropouts should be handicapped relative to those that enrolled more, and that training agencies should not be penalized for operating in particularly adverse labor markets. In this study, we focus exclusively on the set of factors in the DOL adjustment models that are based on the demographic characteristics of the enrollee population, as only these factors can influence enrolment decisions.

**PAW in JTPA**

To illustrate how the adjustment methodology works in JTPA, assume two demographic factors, gender (female, male) and race (black, non-black). The training agency is rewarded on the basis of excess performance, that is, performance above the performance standard. The DOL model adjusts the performance standard around an exogenously given baseline level, that we denote \( m_0 \), according to the characteristics of the training agency’s enrollee population. Suppose an agency enrolled \( x_f \) percent of females, \( x_b \) percent of black and denote by \( \beta_j \) the adjustment weight for demographic characteristic \( j = f, b \). A stylized performance adjustment model can be written as

\[
M_0(x_f,x_b) = m_0 - (\beta_f x_f + \beta_b x_b)
\]  

(1)

where \( M_0 \) is the adjusted performance standard. The higher the standard, the greater is the difficulty obtaining an award. We define an adjustment factor as a socio-economic variable (e.g. \( x_f \)) that is used to correct the standard and an adjustment weight as the numerical value that is imputed to correct the standard (e.g. \( \beta_f \)). For example, if \( \beta_f \) is positive, then the agency is more likely to receive an award, ceteris paribus, if it enrolls more females.

The DOL chose different sets of factors for each performance measure based upon their availability, their statistical relation with the performance measure, and political
considerations. The first line in Table 1 presents the baseline level \((m_0)\), the first column in the bottom panel presents the set of adjustment factors \((x)\) for the adult ER and WE standards, and the core of the table reports the value of the adjustment weights \((\beta)\) corresponding to these factors.\(^9\) The adjustment weights are the regression coefficients from a regression of outcomes on demographic characteristics constructed before the beginning of a new two year cycle using data gathered in the previous cycle. The columns report the weights for different program years. The adjustment weights remain in force for two consecutive program years before they are updated. Thus for example in program years 1992 and 1993, the adjustment weights for the ER standard for females was .072; in program years 1994 and 1995 it was .056; and so on.

Table 1 shows that the DOL adjustment model can significantly impact the performance standard and therefore the agency’s likelihood to receive an award. For example, an agency in either 1992 or 1993 enrolling only applicants that embodied all of the characteristics associated with positive weights would face a ‘negative’ performance standard on the employment measure (the adjustment, \(100\Sigma \beta_i x_i\), is greater than the baseline level implying \(M_0<0\)), meaning that it would not be penalized even if none of its terminees were employed. Although this example is extreme, Table 1 reveals that many of the weights can lower the employment standard by 10 percent or more.

Table 2 focuses on the employment measure and presents summary statistics on the distribution across agencies of the actual adjustment to the baseline \((\Sigma \beta_i x_i)\) by program year. Line 1, for example, says that the ER standards in 1993 ranged from 37 percent (86-49) to 74 percent (86-12) suggesting that a training agency’s enrolment pool—which is a

---

choice variable—could greatly influence its standard. The adjusted performance standards for the earnings measure (not reported) show the same degree of variation.

The likely impact of PAW on enrolment is difficult to assess on theoretical grounds alone. On the one hand, the magnitude of the changes in the weights implies that the enrollee intake composition may have a significant impact on the standard. Meeting the standard was an issue in practice and had financial consequences. In fact, over the period 1993-1998, on average about 23% of training agencies failed to meet the employment rate standard, 6% failed to meet the earnings standard, and 6% failed to meet to meet both standards.\footnote{If the performance standards were set too high, so that all training agencies would fail no matter how they tried, then the ability to modify a standard using the enrolment composition would not matter much, and one would not expect to see enrolment choices affected by PAW. This is also true if performance standards were set too low so that all training agencies exceeded their standards whether they enrolled purposefully or not.} On the other hand, the number of demographic subgroups, and thus the number of implied shadow prices, increases exponentially with the number of factors. In addition, the PAW varied across performance measures. As a result, PAW introduced very complex trade-offs and may have had little consequence in practice. In the end, whether PAW influenced intake choices is an empirical issue.

It is more convenient to work with demographic subgroups instead of demographic characteristics. There is a simple correspondence between subgroups and characteristics. In our example, the two factors determine four demographic subgroups (black female, black male, and so on). Denote $s=(s_{bf},s_{bm},s_{nf},s_{nm})$ the enrolment vector measured in percentage of overall population. $s_{bf}$, for example, represents the percentage of enrollees who are black and female. We can rewrite the performance standard as

$$M_0(n) = m_0 - \left( \omega_{bf}s_{bf} + \omega_{bm}s_{bm} + \omega_{nf}s_{nf} + \omega_{nm}s_{nm} \right)$$

(2)

where $\omega_{bf}=\beta_b+\beta_f$, captures the decrease in standard due to increasing the fraction of black
female by one percent, and can be interpreted as the ‘shadow price’ for that demographic subgroup. The other coefficients are similarly derived, ω_{bm}=β_b, ω_{nf}=β_f, ω_{nm}=0. In the rest of this paper, we also call the ω adjustment weights, keeping in mind the distinction between the β in (1) and ω in (2).

3-Theoretical Predictions

We present a general model of how an agency should respond to a generic adjustment method that introduces shadow prices for an exogenously given partition of population subgroups. For the sake of exposition, we cast the model in a job training context but one should keep in mind that the model could be applied in other contexts. We derive predictions on how enrolment decisions and performance outcomes change for different demographic subgroups when the adjustment method changes.

To simplify, we assume there is a single performance measure and I distinct demographic subgroups. Any applicant belongs to one and only one group. The training agency enrolls n_i applicants of demographic group i. The enrolment vector is denoted n=(n_1,...,n_I) and the total number of enrollees is N=Σn_i. The cost of training n_i enrollees c_i(n_i) is subgroup i specific and increasing and convex in n_i (c'_i(n_i)>0, c''_i(n_i)≥0). The average performance outcome for group i is m_i(n_i) with m_i(0)≥0, m'_i(n_i)>0, and m''_i(n_i)<0. These assumptions are consistent with the following interpretation. Applicants differ within a subgroup. Some applicants are easier to train and are more likely to achieve successful outcomes than others. It is optimal for the training agency to select first the most promising applicants. If the training agency enrolls more applicants of a given subgroup it will enroll those who cost more to serve (cost is increasing) and who are less likely to perform well (performance increases at a decreasing rate). These assumptions are reasonable if there is some heterogeneity within demographic
subgroups that is observed by the agency. Cream-skimming becomes possible because the agency enrolls those applicants within a demographic subgroup who are likely to perform well on the measure, irrespectively of how well they perform on the true objective of job training.

The aggregate performance outcome is the sum of performance outcomes over all groups, $M(n) = \sum_i m_i(n_i)$. The performance standard adjusts a baseline level $m_0$ for the enrollee composition

$$M_0(n) = m_0 - \sum_i \beta_i n_i$$

where $\beta_i$ is the adjustment weight for demographic group $i$. The training agency is rewarded on the basis of excess performance

$$\Delta(n) = M(n) - M_0(n).$$

The training agency chooses $n$ to maximize its utility subject to the budget constraint $\sum_i c_i(n_i) \leq B$. Utility is $U(n, \Delta(n))$ where the first argument captures agency preferences over enrollee choices. To simplify, we consider the following functional form,

$$U(n, \Delta) = \sum_i \alpha_i n_i + \Delta$$

where $\alpha_i$ is a real number that captures the marginal preference attributed to demographic group $i$. The overall level of $\alpha$ defines how the training agency is willing to compromise its own preferences over enrolment for higher performance award.

The designer may change one or more weights at a time. In general, the designer changes adjustment weight $i$ by $\delta_i$ where $\delta=(\delta_1, ..., \delta_I)$ is the vector of changes in weights. Denote

$$\overline{\delta} = 1/N \sum_i \delta_i$$

and the adjustment weight on measure $i$ by $\beta_i + \varepsilon \delta_i$ where $\varepsilon \in [0,1]$.

We first show that under general assumptions about the cost and performance outcome functions, the training agency responds to an increase in the adjustment weight of a single
demographic subgroup i, \( \delta = (0.0, \delta_i = 1.0, 0.0) \), by enrolling more applicants of subgroup i and fewer applicants of subgroup \( k \neq i \). The proposition holds independently of the training agency’s own preferences over enrollee choices. We then show that as the adjustment weight of subgroup i increases, the average performance outcome of subgroup i decreases. The reason is simply that to increase its enrolment of applicants of type i, the training agency has to enroll less attractive applicants. Marginal enrollees achieve lower performance outcomes than average ones.\(^{11}\)

**Proposition 1:** (a) \( \frac{d n_i}{d \beta_i} \geq 0 \) and \( \frac{d n_j}{d \beta_i} \leq 0 \) for \( j \neq i \) and these inequalities are strict for any interior solution \( (n_i > 0) \). (b) \( \frac{d [m_i(n_i)/n_i]}{d \beta_i} \leq 0 \) and the inequality is strict for any interior solution.

(The proofs of our propositions are in Appendix 2.)

Under additional assumptions on the model’s primitives, we can derive general predictions on the impact of simultaneous changes in multiple performance weights. Denote \( n_i(\varepsilon) \) the number of enrollees of group i as a function of \( \varepsilon \), \( \Delta n_i = n_i(1) - n_i(0) \) the change in the number of enrollees of group i, and \( \Delta (n_i/N) \) the same change measured in percentage terms. Similarly, we define \( \Delta [m_i(n_i)/n_i] \) as the change in average performance. Proposition 2 derives general predictions on the impact of any change in the performance weights.

**Proposition 2:** Assume \( c_i(n) = c_i' n = c \) and \( m_i''(n) = m_i''(n) = m \). (a) \( \Delta n_i > \Delta n_k \) iff \( \delta_i > \delta_k \). (a’) \( \Delta (n_i/N) > \Delta (n_k/N) \) iff \( \delta_i > \delta_k \). (b) \( \Delta [m_i(n_i)/n_i] < \Delta [m_k(n_k)/n_k] \leq 0 \) if \( \delta_i > \delta_k \). (c) If \( m_i(0) = 0 \) and \( \beta_i = \beta \) then \( \Delta [m_i(n_i)/n_i] < \Delta [m_k(n_k)/n_k] \leq 0 \) if \( \delta_i > \delta_k \).

Proposition 2 holds if the cost and performance measure functions have a linear and quadratic structure respectively, that is, \( c_i(n) = c_{0_i} + c_{1_i} n_i \) and \( m_i(n_i) = m_{0_i} + m_{1_i} n_i + m_{2_i} n_i^2 \). The

\(^{11}\) To simplify, we assume that the training treatment is constant across groups.
proposition makes a prediction on the relative change in enrollees. Proposition 2(a') says that the increase in the share of enrollees from subgroup i is greater than the increase in subgroup k if and only if subgroup i's adjustment weight increases by a larger amount (or decreases by a smaller amount) than subgroup k. This result constitutes the focus of our empirical investigation (Hypothesis H1). Proposition 2(b) says the change in average performance of subgroup i is lower than the change in average performance of subgroup k if and only if \( \delta_i > \delta_k \) (Hypothesis H2). The remainder of this paper tests hypothesis H1 and H2.

4-Data and Empirical Strategy

We use data from the Standardized Program Information Report which compiled by the Social Policy Research Associates for the DOL for the program years 1993 through 1998 and distributed by the W.E. Upjohn Institute of Employment Research. Appendix 3 explains in detail how we constructed our panel data of demographic subgroups. Our sample covers about two thirds of the agencies in about half the states. The JTPA adjustment model modified the standards for a program year based on the characteristics of the enrollees who terminated that year. Thus the subgroups for program year 1993, for example, include only those enrollees who terminated that year.

Using all 24 factors would generate more than 16 million subgroups. Since the JTPA enrollee population is much smaller (for the enrolment analysis, for example, we have information on 682,515 terminees over the 6 program years), we eliminate all subgroups for which we have no or few enrollees. In the end, we select 13 factors and construct 1,670 different subgroups for which we have information over all 6 years in at least one agency.

---

12 It does not say anything about the direction of the change in the number of enrollees of group i or k. The total number of enrollees of group i could increase or decrease and similarly for group k.
13 The assumptions stated in Proposition 2 are necessary and sufficient for claim (a).
This yields an average of 291 subgroups per agency-year. Table 3 presents descriptive statistics for our main variables (PAW, enrolment shares, and performance outcomes).

Tables 1 and 2 demonstrate that there is much variation across years in our explanatory variables. For example, to meet its employment standard in program years 1992 or 1993, an agency that enrolled no ‘high-school dropouts’ would have to achieve an employment rate 18.4 percent higher than an agency that enrolled only ‘high-school dropouts’ (assuming that all other characteristics are equal across the agencies). In program year 1998 or 1999, however, the difference drops to 6.6 percent, an order of magnitude of about three. In addition, some adjustment factors eventually disappear from the adjustment worksheets and new factors are introduced.\(^\text{14}\)

Table 4 shows that there is also much variation from year to year in the enrolment size of the demographic subgroups identified by the DOL adjustment factors. We investigate whether this variation in enrolment can be explained by the changes in adjustment weights as predicted by (H1).

**Exogeneity of the changes in the PAW**

The PAW were changed three times in our sample period, at the end of 93, 95, and 97 (see Table 1). In the empirical analysis, we assume that these changes are exogenous to contemporaneous enrolment decisions. Several arguments support this assumption. The PAW were computed as coefficient estimates of a regression of performance outcomes on demographic factors using performance data from all training agencies in the previous

---

\(^\text{14}\) This lifecycle phenomenon of adjustment weights was observed earlier in JTPA’s history by Barnow and Constantine (1988) who attribute it to increased proficiency due to learning by the training agencies in selecting enrollees on the basis of factors not included in the model.
two year cycle. Thus while agency choices affect future standards, a single agency has only a negligible effect on its own standards, as its outcomes and enrollment choices are only a minuscule portion of the data used to compute future weights. Thus we can assume an individual agency maximizes its current period award myopically.

The changes in the PAW could also be driven by changes in labor market conditions. This is not an important issue for our empirical exercise, and if anything, can only create an under-estimation of the agency responses. To see this, assume that these changes are conditionally uncorrelated (e.g. random walk or permanent changes). Such changes would influence the PAW (through the regression model) but this would not introduce an endogeneity problem since the change in next period labor market conditions is uncorrelated with the current change in PAW. The only concern is trends in labor market conditions. Such trends could bias the inference against our hypothesis. Assume for example that the labor market potential of a subgroup starts to degrade. This increases that subgroup’s PAW but the increase under-compensates for the continuing degradation in the subgroup’s potential so we would under-estimate the enrolment response relative to the response that would take place with a truly exogenous change in PAW.16

**Empirical strategy**

Denote $s_{iat}$ as the share of enrollees of demographic subgroup $i$ in agency $a$ in year $t$ and $w_{it}$ the adjustment weight, common to all agencies, for subgroup $i$ and year $t$.17 $H1$ implies an increasing relation between changes in relative weights and changes in relative

---


16 In this last scenario, $H2$ may hold for obvious reasons: the performance outcomes will decrease due to the negative trend.

17 $s_{iat}$ is defined as $\frac{n_{iat}}{\sum_k n_{kat}}$ where $n_{iat}$ is the number of enrollees of group $i$ in agency $a$ and year $t$. 

18
shares of subgroups. We test this relation using the three changes in adjustment weights that took place in our sample period.

We propose different specifications to test H1 that are variations around the following approach. Assuming that the increasing relation implied by H1 is linear and does not vary across subgroups, agencies, or years gives

\[(s_{iat}-s_{iat'})-(s_{kat}-s_{kat'})=\gamma[(w_{it}-w_{it'})-(w_{kt}-w_{kt'})]\]

for all i,k,a,t,t'  \((H1')\)

where the parameter of interest \(\gamma\) is positive. Instead of comparing pairs of demographic subgroup-years, which does not naturally fit a regression framework, we aggregate this hypothesis to obtain a relation that can be estimated using a fixed-effect regression framework.\(^{18}\) Formally, H1' is averaged over subgroups k and years t' to obtain

\[s_{iat}=-(s_a-\gamma w_a)+(s_{ai}-\gamma w_{ai})+(s_{at}-\gamma w_{at})+\gamma w_{it}\] for all i,a,t

where \(s_a\) denotes the average share in agency a across all years and subgroups, \(s_{ai}\) the average i share in agency a across all years and similarly for \(s_{at}\) and the w averages. The observed shares could vary randomly because they are measured with error (which is the case in our application since only a representative sample of 62 percent of total population is included in our dataset).\(^{19}\) We obtain the following empirical model

\[s_{iat}=\alpha+\alpha_{ai}+\alpha_{at}+\gamma w_{it}+\epsilon_{ait}\]  \((3)\)

where \(\alpha\) is a constant, \(\alpha_{ai}\) is a subgroup-agency fixed effect and \(\alpha_{at}\) is an agency-time fixed effect. We assume \(\epsilon_{iat}\) is normal, mean zero, and distributed independently across training agencies. We cluster the errors at the training agency level to permit arbitrary forms of autocorrelation and heteroscedasticity within training agency panels.

\(^{18}\) Alternatively we could choose a subgroup to serve as the reference subgroup against which we compare all other subgroups but the choice of the reference subgroup is arbitrary.

\(^{19}\) Alternatively, we could derive the econometric model following a random utility approach, assuming that agencies have group preferences that vary randomly over time.
The theory makes no prediction on \((\alpha,\alpha_{ai},\alpha_{at})\) but predicts that \(\gamma\) should be positive.

Specification (3) tests an averaged version of H1. Ignoring the subscript \(a\), \(\gamma\) measures the average impact of a change in a subgroup’s relative PAW on the share of enrollees relative to the average share. Loosely speaking, we are comparing across the three program cycles each enrollee share relative to an hypothetical representative group. \(\gamma\) measures the average impact across all groups and years, and since we pool all agencies (subscript \(a\)), the impact is also averaged over agencies.

To test H2, we follow a similar procedure. We estimate the performance of each subgroup holding constant agency-subgroup and agency-time effects

\[ m_{iat} = \theta + \theta_{ai} + \theta_{at} + \nu_{ijt} \text{ for all } i,a,t \]  

where as before \(\theta_{ai}\) and \(\theta_{at}\) allows for agency-subgroup and agency-time fixed effects. The parameter of interest is \(\theta\), which our model predicts is negative. As with (3), we assume \(\nu_{ijt}\) is normal, mean zero, and distributed independently across training agencies and we cluster the errors at the training agency level.

In all specifications reported to test H1 and H2, we weight each subgroup-agency-year observation by the subgroup-agency share of the entire terminee population.\(^{20}\) We have also considered two variant specifications and the results were not affected (not reported): one with equal weights and another with weights proportional to the subgroup’s share relative to its agency population. In addition, we have considered specifications where in constructing the subgroups we exclude those enrollees who are terminated in the first four months of each two year cycle. Our reasoning is that enrollees entering a new two year cycle may have been

\(^{20}\)That is, we weight each observation by \(\sum_{t} n_{iat} / \sum_{a} \sum_{k} \sum_{t} n_{kat}\) where \(n_{iat}\) is the number of enrollees in subgroup \(i\) in agency \(a\) in year \(t\).
enrolled to optimize the previous cycle’s weights.  

5-Results

Each of the four standards had its own adjustment weights that could potentially influence the enrollee intake choice (H1) and also the performance outcomes (H2). We initially test H1 and H2 using the adjustment weights on the adult ER performance measure. In focusing on the employment measure, we follow the policy evaluation literature (e.g., Anderson et al. 1993) and this is justified by the fact that this measure has received a disproportionate emphasis in the incentive system (Courty and Marschke 2003).

5-1 Tests of H1 and H2 for the ER Adjustment Weights

Table 5 reports the results from our estimation of the enrolment decision model, equation (3). In all specifications the dependent variable is the subgroup’s termination share. The right-hand side of the regression includes the subgroup’s weight for the employment standard (ER) in addition to the \( \alpha_{ai} \) and \( \alpha_{at} \).

Model 1 produces a positive and statistically significant estimate of the ER weight coefficient, a finding that is consistent with H1. To give the reader an idea of the magnitude of the impact of the weight change on enrollee choice, we include the standardized coefficients. Literally interpreted, our result says that a one standard deviation in a subgroup’s performance weight relative to the average ER weight increases the subgroup’s enrolment share by about .1 percent relative to the average agency subgroup. This response, however, is measured at the subgroup level which is the correct unit of analysis to understand agency behavior, but is of limited relevance from a policy point of view. To assess the economic significance of this response, consider the following thought experiment. Assume a

\[ \text{We chose four months, because the average enrolment duration is between four and five months long. Five months into the new cycle, we reason, enrollees will be terminating in the cycle in which they were intended to be terminated.} \]
coefficient on a demographic characteristic is increased by a standard deviation relative to the average coefficient. For the sake of exposition, imagine a change in the coefficient on female share, noting that is the choice of characteristic is arbitrary because, as mentioned earlier, γ measures an average effect across all groups. The enrolment share of all female subgroups will increase by about .1 percent relative to the average subgroup. Since there are on average 291 subgroups per agency in our sample (see Appendix 3), the overall increase in the share of females will be 14.7 percent (0.00101*291/2, because half of the subgroups are female on average). The female PAW alone can have a large impact on the composition of the enrollee population. But there are approximately a dozen demographic characteristics in play in the analysis suggesting that changes in PAW do have a significant influence on enrolment. To put this figure into perspective, consider Cragg’s (1997) analysis of the impact of an adjustment policy distinct from PAW on the enrollment of persons with little work experience and thus likely to perform poorly on employment-based performance measures (Table 2, p. 156). He finds that the share of enrollees with little work experience increased by 56 percent in those states that used the adjustment policy. In light of that number, our 14.7 percent figure does not appear unrealistic.

It is useful to compare the magnitude of the response we find with the estimates of enrollment distortion found in the cream-skimming literature. Anderson et al. (1993), for example, report that females make up 62 percent of the eligible sample, but only 57 percent of the participant sample in Tennessee in the late 1980s (Table 1, p. 618). The difference in

---

22 Table 2 in Cragg reports that low (high) experience workers are 107 (41) percent more likely to be enrolled in states with the adjustment policy. Assuming that about 30 percent of the enrollees is experienced (Table A2) implies the 56 percent figure reported in the text.
23 As mentioned in footnote 6, Heckman, Smith and Taber (1996) find evidence of cream avoidance. This evidence, however, is based on the behavior of a single training center in Corpus Christi, Texas in the late 1980s. There may exist much heterogeneity in the agencies’ objective function, and Corpus Christi may have strong preferences over the intake population that cancel the incentive to cream skim.
shares for females is 5 percent and the highest difference in shares across all characteristics considered is 22 percent.\textsuperscript{24} Our results suggest that raising the female PAW by a third of a standard deviation would bring the female share of the participant population in line with its share in the eligible population. One could correct the highest distortion in share with an increase in PAW of one and half standard deviations.

Model 3 includes both the subgroup’s ER and WE adjustment weights. The coefficient estimate on the employment weight (in raw and standardized form) changes little from column 1. This finding supports our assumption that there is little interaction between the different measures of the incentive system.

Table 6, column 1, reports the results of the employment outcome estimation (equation 4). We find a statistically significant and negative coefficient estimate on the employment weight, as predicted under H2. The magnitude of the estimate suggests that a one standard deviation increase in a subgroup’s ER weight relative to the average ER weight decreases the subgroup’s relative employment rate by about 2 points. This result remains when we add the earnings weight as an explanatory variable. Note our interpretation of the change in outcome here: we are treating the change in outcome as resulting from the change in the type or characteristics of the persons enrolled (the raw material with which case workers work), not as a change in labor market outcomes produced by case workers (e.g., by the change in training effort or quality). Because the regressions in Table 6 contain subgroup-training agency fixed effects, we interpret the coefficient estimate as indicating the added effect on the subgroup's performance outcome from the PAW.

To assess the economic implication of this result, consider the thought experiment discussed above. Increasing the female weight by one standard deviation decreases the

\textsuperscript{24} The socio-economic variables considered in that study do not exactly match those covered in the PAW.
performance of females relative to the average subgroup performance by 2 points. This seems reasonable to us considering that this change in weight is associated with a 14.7 percent increase in the relative share of females. Changing multiple PAWs simultaneously could have large impacts on the employment outcome. In contrast, Anderson et al. (1993) find that selective enrolment increased the probability of employment relative to random enrolment by 9.1 percent (p. 620). This suggests that agencies are willing to forgo significant gains in performance outcomes---of the similar magnitude as what non-random enrollment would add---to take advantage of the adjustment policy.

The quality of the marginal enrollee within a subgroup decreases as more enrollees are drawn from this subgroup, which is consistent with the hypothesis that agencies cream-skim the best enrollees within each subgroup. Still, this figure is small relative to the variation in performance across demographic subgroups. Table 3 shows that the standard deviation in the employment outcome across all subgroups and years is 43 points. Therefore the potential to cream-skim across subgroups (which can be curbed with the PAW) is of an order of magnitude greater than within subgroups (which is unaffected by the PAW). The PAW can eliminate the incentive to cream-skim across subgroups leaving a residual incentive to cream-skim within subgroup which is second order.

5-2 Additional Measures and Robustness

The previous analysis is valid under the assumption that the ER has received the most emphasis in the incentive scheme, as has been argued in the literature. It is also valid if the

---

25 What is relevant for cream-skimming across subgroups in the absence of PAW are the predictable differences across subgroups at the agency level. To capture this, we first take a year average of the subgroup performance at the agency level (this eliminates the unpredictable component of performance that is irrelevant in cream-skimming), then compute the standard deviation in subgroup performance at the agency level, and finally take the average across all agencies. We obtain an average standard deviation of 43.6 which is very similar to the above figure.
variation in the different enrolment incentives associated with each set of weights are orthogonal to one another. Still, multi-dimensional incentives may matter. We address this issue in two ways. First, we test H1 and H2 for the adjustment weights on the WE measure. Second, we consider the impact of the weights on both the ER and WE measures on the sub sample of non-welfare recipients and the logic is that the weights on the two welfare measures should not influence the enrolment incentives among the non-welfare sub populations.

Adjustment Weights on the Earnings Performance Measure (WE)

Table 5, columns 2 and 3, show the impact of the WE adjustment weights on the enrolment decision. Both columns show no impact which goes against H1. Table 7 reports the results of the earnings outcome specification (model (4) applied to WE). Whether we estimate the model with just the earnings weight or both the earnings and employment weights, the estimated coefficient on the earnings weight is statistically insignificant against H2. These two results could be because the WE measure plays a lesser role in the incentive system or because the agencies have less discretion to select enrollees who are likely to perform well on the WE measure.

Interestingly, the coefficient estimate corresponding to the employment weight in table 7, column 2, is positive and significant (p value 0.001). While there are many potential explanations for this finding, it is consistent with the existence of a trade-off under multi-dimensional incentives: the kinds of enrollees within a subgroup that produce higher employment outcomes, reduce earning outcomes.

Non-Welfare Recipients

26 The number of observations used in this analysis is smaller than in the employment analysis because, consistent with the JTPA definition of the earnings measure, we use only the enrollees who are employed (by the employment measure definition) in the calculation of the earnings outcome.
The training agency’s decision to enroll adult non-welfare recipients is less complicated than the decision to enroll welfare ones because non-welfare recipients’ characteristics enter into the determination of only the two standards that have been the focus of this analysis. If the welfare measures play an important role, we should obtain a cleaner test of H1 and H2 when we limit the analysis to the non-welfare adults. Therefore, Table 8 shows the results of the previous analyses excluding the welfare recipients.

Two points should be made. First, the power of the significance tests is smaller after we exclude welfare recipients, which constitute about 40 percent of the adult population. This is partly responsible for why we observe that the coefficient estimates on the employment weight in the enrolment share (model 1) and outcome regressions (models 2 and 3) are insignificant. Second, the impact of the earnings weight in the regressions is greater when we exclude welfare recipients. In the enrolment share regression, though the coefficient estimate on earnings weight remains insignificant (by conventional significance standards) it is positive (as predicted under H1). The coefficient estimate on earnings weight in the earnings outcome regression (model 3) is now both negative and significant as predicted under H2. The standardized coefficient corresponding to this estimate is about -6 suggesting that a one standard deviation increase in the WE weight relative to the average WE weight reduces the relative subgroup earnings per week by about $6.

Taken together, these two new sets of results suggest that although H2 does not hold for the entire sample, it does hold for the subset of non-welfare recipients. An explanation is that agencies have much more discretion to select applicants who are likely to perform well on the earnings measure, when they have to choose among non-welfare recipients, than they do for welfare recipients, who have on average lower levels of human capital. Also the earnings
measure is calculated only off employed terminees. Because welfare recipients are less likely to be employed their prospective earnings might not be of such concern in the enrolment decision.

5-Summary and Conclusions

The introduction of performance incentives in several branches of the public service sector, such as in job training, education, and health, has raised concerns as to their impact on who receives services. In particular, rewarding public agencies based on measurable outcomes such as test scores, employment or health outcomes may lead job training programs to neglect persons with significant employment barriers, schools to route students out of the mainstream and into special education programs, and hospitals to turn away the chronically ill. To retain control over the recipient population, some policy-makers have proposed adjusting the measures that are used to assess performance, effectively setting different ‘shadow prices’ for different subgroups of service recipients, but little evidence exists about the effectiveness of these methods.

In the context of a large public sector job training program, we investigate the influence of enrolment incentives on case workers’ choice of intake population. Job training agencies in the program we study are rewarded for improving the labor market performance of the clients they serve but the reward function also depends on the enrolment choice. The main objective of the enrolment incentives is to level the playing field by relaxing the standards of performance for training agencies that enrolled less able applicants.

Our empirical analysis establishes two sets of results. First, we find that changes in the incentive for enrolling members of a subgroup significantly change the fraction of
enrollees from this subgroup. Second, we demonstrate the existence of within-subgroup heterogeneity. Case workers increase the number of enrollees from a specific subgroup by enrolling at the margin applicants that perform worse on the measure. That is, case-workers appear to be cream-skimming: they use their private information about applicant heterogeneity within subgroups. In contrast with the literature on cream skimming in job training programs which focuses on the impact of incentives on enrolment at the training agency level, we demonstrate that private information carries through even within the demographic subgroups defined by PAW. We show, however, that the potential for cream-skimming within subgroups is second order relative to across subgroup cream-skimming.

Our finding that PAW in JTPA has predictable and economically meaningful effects on enrollment suggest PAW will be effective in other settings where, as in JTPA, incentives are financial and direct (even if low-powered) and where shadow prices are at least of the magnitude of their simple correlations with performance outcomes. Thus we suspect PAW would influence enrolment decisions in programs as diverse as Jobcentre Plus in the UK and WIA in the U.S., as well as non-workforce related programs such as the health insurance industry (the sick fund) in Germany. While Job Services Australia does not explicitly back performance with financial incentives, it makes contract awards contingent on past performance and this may be a powerful motivator of agency enrollment and strong enrollment responses to adjustments. An intriguing question is whether adjustments to "report cards" which attempt to influence only the reputation of an organization (a hospital or school for example) has the same effect on enrollment as adjustments to performance that has direct, financial consequences.

We leave several additional questions for future research. Our result that training
centers exploit heterogeneity within subgroups suggest more fine-grained adjustment schemes that define more subgroups and more shadow prices will offer less within group heterogeneity to exploit and thus less cream skimming. The increased complexity in the adjustments may come at a price, however. As Barnow (1992, pp. 299-300) points out the greater the number of variables and the greater the number of standards adjusted by these variables the more difficult it is for training agencies to understand the incentives they face and therefore the less effective the PAW. Our finding that case workers in JTPA do appear to navigate the complexity created by hundreds of shadow prices bodes well for other programs' complex adjustment schemes, such as Germany's health insurance system which has 600 shadow prices, and the sophisticated risk adjustment models used for assessing hospital and physician performance that are inspiration for the original JTPA regression adjustment models. We leave for future work the determination of the optimal level of granularity in the definition of subgroups.

We also leave for future research an evaluation of the regression-based adjustment model generally and DOL's model specifically for determining PAW. Did the DOL methodology achieve a reduction in cream-skimming or some other objective (e.g. channeling resources toward the subgroups with the highest earning impact)?

27 Also, how do regression model adjustments fare compared to other strategies for dealing with cream skimming, such as the more standard monitoring of training agencies activities as in the pre-JTPA US, the Netherlands (Bruttel, 2005), and other European countries?

---

27 Heckman, Heinrich, and Smith (2002) show little relationship between outcome-based measures such as in JTPA and earnings impacts suggesting that the efficiency effects of the regression model are likely to be small.
References


Burgess, Simon; Propper, Carol; Ratto, Marisa; and Emma Tominey. 2003. "Incentives in the Public Sector: Some Preliminary Evidence from a UK Government Agency," CEPR Discussion Papers 4010,


Heckman, James; Jeffrey Smith; and Christopher Taber. 1996. “What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance” in the JTPA Program.” In *Advances in the Study of Entrepreneurship, Innovation, and Growth, Volume 7, Reinventing Government and the Problem of Bureaucracy*, ed. Gary Libecap,


### Table 1: Baseline Levels and Adjustment Factors and Weights, Program Years 1992-1999

<table>
<thead>
<tr>
<th></th>
<th>Follow-up Entered Employment Rate (ER, %)</th>
<th>Follow-up Weekly Earnings (WE, $/week)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline level ($m_0$)</td>
<td>86 88 88 77</td>
<td>266 301 361 376</td>
</tr>
<tr>
<td>Adjustment Factors (x)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*</td>
<td>0.072 0.056 0.052 0.050</td>
<td>0.425 0.443 0.602 0.683</td>
</tr>
<tr>
<td>Age ≥55 years</td>
<td>-- 0.118 0.105 0.130</td>
<td>1.126 0.774 0.484 0.61</td>
</tr>
<tr>
<td>Age 30 to 54 years*</td>
<td>-- 0.098 -- --</td>
<td>-- -- -- --</td>
</tr>
<tr>
<td>Black*</td>
<td>0.064 0.086 0.035 0.027</td>
<td>0.270 0.325 0.226 0.177</td>
</tr>
<tr>
<td>Other minority*</td>
<td>-- -- -- --</td>
<td>-- 0.100 0.042 0.065</td>
</tr>
<tr>
<td>Minority male*</td>
<td>-- -- -- 0.026</td>
<td>-- -- 0.279 0.306</td>
</tr>
<tr>
<td>High school dropout*</td>
<td>0.184 0.084 0.073 0.066</td>
<td>0.271 0.276 0.24 0.145</td>
</tr>
<tr>
<td>Post high school attendees*</td>
<td>-- 0.066 0.032 0.008</td>
<td>-- -0.415 -0.659 -0.235 -0.334</td>
</tr>
<tr>
<td>HS dropout &lt; 30 yrs old</td>
<td>-- -- 0.02 0.015</td>
<td>-- -- -- 0.088</td>
</tr>
<tr>
<td>Handicapped</td>
<td>0.083 0.09 0.075 0.096</td>
<td>0.367 0.558 0.28 0.315</td>
</tr>
<tr>
<td>UI or UC claimant</td>
<td>-- -- -- -0.037 -0.022</td>
<td>-1.062 -0.361 -0.127 -0.081</td>
</tr>
<tr>
<td>Long-term AFDC recip.*</td>
<td>0.151 0.234 0.025 0.018</td>
<td>-- -- -- 0.086</td>
</tr>
<tr>
<td>Cash welfare recipient*</td>
<td>-- -- 0.054 0.031</td>
<td>-- -- 0.093 0.072</td>
</tr>
<tr>
<td>SSI recipient</td>
<td>-- -- 0.091 0.133</td>
<td>-- -- 0.027 0.265</td>
</tr>
<tr>
<td>Offender*</td>
<td>-- 0.057 -- --</td>
<td>-- -- -- --</td>
</tr>
<tr>
<td>Limited English speaking</td>
<td>-- -- -- --</td>
<td>-- -- -- 0.259 0.251</td>
</tr>
<tr>
<td>Basic skills deficient</td>
<td>-- -- 0.034 0.037</td>
<td>-- -- 0.193 0.286</td>
</tr>
<tr>
<td>Reading skills &lt; 7th grade</td>
<td>-- 0.032 -- --</td>
<td>0.148 0.344 -- --</td>
</tr>
<tr>
<td>No signific’t work history*</td>
<td>0.074 0.059 0.050 0.055</td>
<td>0.292 0.144 0.150 0.098</td>
</tr>
<tr>
<td>Unemployed ≥ 15 weeks*</td>
<td>0.111 0.103 0.086 0.073</td>
<td>-- 0.242 0.091 0.076</td>
</tr>
<tr>
<td>Not in the labor force*</td>
<td>0.122 0.113 0.103 0.108</td>
<td>-- -- -- 0.044</td>
</tr>
<tr>
<td>GA/RCA recipient</td>
<td>0.137 0.05 -- --</td>
<td>-- -- -- --</td>
</tr>
<tr>
<td>Veteran (Vietnam era)</td>
<td>0.160 0.135 0.030 0.081</td>
<td>-- -- -- --</td>
</tr>
<tr>
<td>Homeless</td>
<td>-- -- -- 0.043</td>
<td>0.595 0.602 -- 0.136</td>
</tr>
</tbody>
</table>

*An adjustment factor that is included in our analysis.

Notes:
1. A factor was excluded from out analysis if either (1) the factor described hardly any or almost all individuals (i.e, its mean fell outside the [.1,.9] interval) or (2) factor information was missing for more than 10% of the observations (16% of observations lacked information on the variable Basic skills deficient, and 11% of observations lacked information on the variable Reading skills below 7th grade).
### Table 2: Adjustment to the Baseline Level ($\sum \beta_i x_i$) for the Employment Rate Standard*

<table>
<thead>
<tr>
<th>Program Year</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max</th>
<th>Number of Agencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>25.41</td>
<td>4.79</td>
<td>11.76</td>
<td>49.43</td>
<td>639</td>
</tr>
<tr>
<td>1994</td>
<td>27.62</td>
<td>4.83</td>
<td>14.10</td>
<td>49.36</td>
<td>627</td>
</tr>
<tr>
<td>1995</td>
<td>27.25</td>
<td>5.05</td>
<td>13.99</td>
<td>50.29</td>
<td>665</td>
</tr>
<tr>
<td>1996</td>
<td>18.06</td>
<td>3.75</td>
<td>9.25</td>
<td>39.60</td>
<td>634</td>
</tr>
<tr>
<td>1997</td>
<td>17.42</td>
<td>3.78</td>
<td>5.40</td>
<td>36.07</td>
<td>663</td>
</tr>
<tr>
<td>1998</td>
<td>17.43</td>
<td>3.63</td>
<td>4.05</td>
<td>31.91</td>
<td>610</td>
</tr>
<tr>
<td>1993-1998</td>
<td>22.23</td>
<td>6.34</td>
<td>4.05</td>
<td>50.29</td>
<td>3838</td>
</tr>
</tbody>
</table>

*We compute for each agency the adjustment to the baseline level $\sum \beta_i x_i$ using the agency’s actual enrollee population. The summary statistics reported here are based on the distribution of $\sum \beta_i x_i$ across agencies.

### Table 3: Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>25 Pctl</th>
<th>50 Pctl</th>
<th>75 Pctl</th>
<th>Max</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrollee Share ($s_{it}$)</td>
<td>0.003</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>1.000</td>
<td>808530</td>
</tr>
<tr>
<td>Employ. Rate (%) ($m_{it}$ -- ER)</td>
<td>67.7</td>
<td>42.8</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>168304</td>
</tr>
<tr>
<td>Earnings ($/week) ($m_{it}$ -- WE)</td>
<td>313.4</td>
<td>164.7</td>
<td>0.0</td>
<td>220.0</td>
<td>280.2</td>
<td>365.5</td>
<td>6997.9</td>
<td>125486</td>
</tr>
<tr>
<td><strong>Independent Variables ($w_{it}$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment Weight</td>
<td>0.22</td>
<td>0.13</td>
<td>-0.07</td>
<td>0.13</td>
<td>0.20</td>
<td>0.28</td>
<td>0.79</td>
<td>808530</td>
</tr>
<tr>
<td>Earnings Weight</td>
<td>0.53</td>
<td>0.42</td>
<td>-0.66</td>
<td>0.25</td>
<td>0.58</td>
<td>0.84</td>
<td>1.43</td>
<td>808530</td>
</tr>
</tbody>
</table>

Observation is a subgroup. Total shows the number of subgroups. Subgroups constructed from 1993-1998 SPIR data (see text and Appendix 3 for construction details).
Table 4: Summary Statistics for Adjustment Factors

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female*</td>
<td>63.03</td>
<td>65.80</td>
<td>66.50</td>
<td>67.56</td>
<td>66.88</td>
<td>65.64</td>
<td>65.81</td>
</tr>
<tr>
<td>55 years old &amp; over</td>
<td>2.12</td>
<td>2.05</td>
<td>1.99</td>
<td>1.78</td>
<td>1.86</td>
<td>2.16</td>
<td>2.00</td>
</tr>
<tr>
<td>age 30 to 54*</td>
<td>56.45</td>
<td>56.32</td>
<td>56.52</td>
<td>57.25</td>
<td>57.67</td>
<td>57.65</td>
<td>56.91</td>
</tr>
<tr>
<td>Black*</td>
<td>29.52</td>
<td>30.04</td>
<td>31.85</td>
<td>33.26</td>
<td>33.55</td>
<td>34.61</td>
<td>31.91</td>
</tr>
<tr>
<td>Other minority*</td>
<td>18.58</td>
<td>19.49</td>
<td>20.58</td>
<td>22.16</td>
<td>21.03</td>
<td>23.08</td>
<td>20.64</td>
</tr>
<tr>
<td>Minority male*</td>
<td>17.49</td>
<td>16.99</td>
<td>17.64</td>
<td>17.95</td>
<td>17.91</td>
<td>19.67</td>
<td>17.85</td>
</tr>
<tr>
<td>High school dropout*</td>
<td>21.05</td>
<td>22.08</td>
<td>20.66</td>
<td>20.51</td>
<td>19.59</td>
<td>20.51</td>
<td>20.79</td>
</tr>
<tr>
<td>Post high school attendees*</td>
<td>21.26</td>
<td>22.28</td>
<td>23.69</td>
<td>24.26</td>
<td>24.51</td>
<td>23.81</td>
<td>23.21</td>
</tr>
<tr>
<td>High school dropout under 30</td>
<td>9.50</td>
<td>9.80</td>
<td>9.23</td>
<td>9.03</td>
<td>8.32</td>
<td>8.56</td>
<td>9.12</td>
</tr>
<tr>
<td>Handicapped</td>
<td>13.66</td>
<td>8.22</td>
<td>7.69</td>
<td>7.23</td>
<td>7.12</td>
<td>6.54</td>
<td>8.57</td>
</tr>
<tr>
<td>Long-term AFDC recipient*</td>
<td>15.87</td>
<td>15.81</td>
<td>16.47</td>
<td>16.16</td>
<td>15.55</td>
<td>13.43</td>
<td>15.63</td>
</tr>
<tr>
<td>Cash welfare recipient*</td>
<td>39.61</td>
<td>42.91</td>
<td>42.25</td>
<td>40.08</td>
<td>37.28</td>
<td>32.33</td>
<td>39.41</td>
</tr>
<tr>
<td>SSI recipient</td>
<td>3.00</td>
<td>3.37</td>
<td>3.30</td>
<td>3.34</td>
<td>3.58</td>
<td>3.49</td>
<td>3.33</td>
</tr>
<tr>
<td>Offender*</td>
<td>13.05</td>
<td>13.46</td>
<td>13.57</td>
<td>14.82</td>
<td>16.06</td>
<td>17.26</td>
<td>14.52</td>
</tr>
<tr>
<td>Limited English speaking</td>
<td>5.10</td>
<td>4.68</td>
<td>4.18</td>
<td>4.48</td>
<td>4.10</td>
<td>4.84</td>
<td>4.58</td>
</tr>
<tr>
<td>Basic skills deficient</td>
<td>57.50</td>
<td>58.69</td>
<td>55.59</td>
<td>54.68</td>
<td>54.61</td>
<td>56.42</td>
<td>56.28</td>
</tr>
<tr>
<td>Reading skills below 7th grade</td>
<td>14.31</td>
<td>15.62</td>
<td>13.60</td>
<td>13.41</td>
<td>12.50</td>
<td>13.78</td>
<td>13.95</td>
</tr>
<tr>
<td>Lacking significant work history*</td>
<td>35.06</td>
<td>34.81</td>
<td>35.40</td>
<td>36.61</td>
<td>34.98</td>
<td>34.17</td>
<td>35.18</td>
</tr>
<tr>
<td>Unemployed 15 weeks or more</td>
<td>41.49</td>
<td>36.54</td>
<td>33.36</td>
<td>31.26</td>
<td>31.26</td>
<td>31.16</td>
<td>34.62</td>
</tr>
<tr>
<td>Not in the labor force*</td>
<td>29.85</td>
<td>32.63</td>
<td>35.45</td>
<td>37.16</td>
<td>33.00</td>
<td>29.89</td>
<td>32.98</td>
</tr>
<tr>
<td>GA/RCA recipient</td>
<td>5.52</td>
<td>5.93</td>
<td>4.95</td>
<td>4.19</td>
<td>3.40</td>
<td>2.98</td>
<td>4.63</td>
</tr>
<tr>
<td>Veteran (Vietnam era veteran)</td>
<td>9.52</td>
<td>7.83</td>
<td>7.44</td>
<td>7.00</td>
<td>6.98</td>
<td>6.39</td>
<td>7.64</td>
</tr>
<tr>
<td>Homeless</td>
<td>4.38</td>
<td>2.22</td>
<td>2.42</td>
<td>2.63</td>
<td>2.57</td>
<td>2.40</td>
<td>2.81</td>
</tr>
<tr>
<td>Total Number of Terminees</td>
<td>138533</td>
<td>136834</td>
<td>121086</td>
<td>112822</td>
<td>110648</td>
<td>95653</td>
<td>715576</td>
</tr>
</tbody>
</table>

*Thirteen adjustment factors included in the analysis (see Table 1, Note 1)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment Weight</td>
<td>0.00759</td>
<td>0.00101</td>
<td></td>
<td></td>
<td>0.00773</td>
<td>0.00103</td>
</tr>
<tr>
<td></td>
<td>(0.00219)</td>
<td></td>
<td></td>
<td></td>
<td>(0.00222)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>Earnings Weight</td>
<td></td>
<td>0.00007</td>
<td>0.00003</td>
<td>-0.00042</td>
<td>-0.00018</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00059)</td>
<td>(0.00059)</td>
<td></td>
<td>0.900</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00000</td>
<td>(0.00001)</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.256</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.00799</td>
<td>0.00777</td>
<td>0.00798</td>
<td></td>
<td>0.00006</td>
<td>(0.00006)</td>
</tr>
<tr>
<td></td>
<td>(0.00006)</td>
<td></td>
<td>(0.00000)</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td></td>
<td>0.000</td>
<td></td>
<td>0.256</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>738689</td>
<td>738689</td>
<td>738689</td>
<td></td>
<td>738689</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td></td>
<td>0.49</td>
<td></td>
</tr>
</tbody>
</table>

Notes:
2. P values in italics.
3. Errors clustered on training agencies.
4. All models include fixed effects (see equation 3 in text).
5. All models are weighted by the subgroup-agency share of all terminees.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(3.42407)</td>
<td>0.000</td>
<td>(3.44840)</td>
<td>0.000</td>
</tr>
<tr>
<td>Earnings Weight</td>
<td></td>
<td></td>
<td>1.36506</td>
<td>0.56749</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.34891)</td>
<td>0.312</td>
</tr>
<tr>
<td>Constant</td>
<td>1.84150</td>
<td></td>
<td>1.86583</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.16014)</td>
<td>0.000</td>
<td>(0.16243)</td>
<td>0.000</td>
</tr>
<tr>
<td>Observations</td>
<td>164488</td>
<td></td>
<td>164488</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.46</td>
<td></td>
<td>0.46</td>
<td></td>
</tr>
</tbody>
</table>

Notes:
2. P values in italics.
3. Errors clustered on training agencies.
4. All models include fixed effects (see equation 4 in text).
5. All models are weighted by the subgroup-agency share of all terminees.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment Weight</td>
<td>52.52690</td>
<td>6.92964</td>
<td>(15.31844)</td>
<td>0.001</td>
</tr>
<tr>
<td>Earnings Weight</td>
<td>3.02610</td>
<td>1.25155</td>
<td>-0.58678</td>
<td>-0.24268</td>
</tr>
<tr>
<td>Constant</td>
<td>3.93647</td>
<td>6.66870</td>
<td>(0.35794)</td>
<td>(0.93327)</td>
</tr>
<tr>
<td>Observations</td>
<td>122467</td>
<td>122467</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.52</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
2. P values in italics.
3. Errors clustered on training agencies.
4. All models include fixed effects (see equation 4 in text).
5. All models are weighted by the subgroup-agency share of all terminees.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment Weight</td>
<td>0.00493</td>
<td>0.00052</td>
<td>-3.51101</td>
<td>-0.36899</td>
<td>37.22452</td>
<td>3.79134</td>
</tr>
<tr>
<td></td>
<td>(0.00459)</td>
<td>0.284</td>
<td>(5.44246)</td>
<td>0.519</td>
<td>(23.48833)</td>
<td>0.114</td>
</tr>
<tr>
<td>Earnings Weight</td>
<td>0.00315</td>
<td>0.00129</td>
<td>-1.35605</td>
<td>-0.55022</td>
<td>-14.19245</td>
<td>-5.70765</td>
</tr>
<tr>
<td></td>
<td>(0.00196)</td>
<td>0.109</td>
<td>(1.75199)</td>
<td>0.439</td>
<td>(8.09520)</td>
<td>0.080</td>
</tr>
<tr>
<td>Constant</td>
<td>0.01473</td>
<td>2.20412</td>
<td>1.18350</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00015)</td>
<td>0.000</td>
<td>(0.22983)</td>
<td>0.000</td>
<td>(1.16103)</td>
<td>0.309</td>
</tr>
<tr>
<td>Observations</td>
<td>366682</td>
<td>92187</td>
<td>71800</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.45</td>
<td>0.42</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
2. P values in italics.
3. Errors clustered on training agencies.
4. All models include fixed effects (see equations 3 and 4 in the text).
5. All models are weighted by the subgroup-agency share of all non-welfare receiving terminatees.
Appendix 1: The JTPA Incentive System

The Act called for financially-backed performance incentives that would measure and reward training agency’s success in developing participants’ human capital, the primary goal of the program according to the Act (JTPA, section 106(a)). Congress gave the Department of Labor (DOL) the responsibility of developing a workable set of performance measures that would reflect the Act’s mission. At the end of each program year, training agencies were rewarded (or sanctioned) on the basis of their performance relative to these DOL measures.

The award for the successful training agency averaged about seven percent of its budget. In some states, the highest awards amounted to about sixty percent of the training agency’s budget. The reader who is interested in the details of the incentives confronting JTPA training agencies should see Courty and Marschke (2003).

Meeting the performance standards was a condition for receiving an award and in many states most of the award a training agency was eligible for was paid out for simply meeting the standard. Thus, the structure of the incentives under JTPA meant that a training agency interested in avoiding sanctions and maximizing its award, should focus on meeting its standards.
Appendix 2: Proofs of Propositions

We first derive a general result that is used in the proofs of both propositions. Denote $\lambda$ the Lagrange multiplier on the budget constraint. In any interior solution ($n_i>0$), the first order condition says
\[ m'(n_i) + \alpha_i + \varepsilon \delta_i + \beta_i = \lambda c_i'(n_i). \]

Take derivative of the first order condition and budget constraint with respect to $\varepsilon$.
\[
\begin{cases}
  m_i'' \frac{dn_i}{d\varepsilon} + \delta_i = \lambda c_i'' \frac{dn_i}{d\varepsilon} + c_i' \frac{d\lambda}{d\varepsilon} \\
  \sum_j c_j' \frac{dn_j}{d\varepsilon} = 0
\end{cases}
\]

Compute the value of $\frac{d\lambda}{d\varepsilon}$ as
\[
\frac{d\lambda}{d\varepsilon} = \frac{\sum_j \delta_j c_j'}{\lambda c_j'' - m_j''} - \frac{\sum_j c_j^2}{\lambda c_j'' - m_j''}.
\]

Plugging back into the first order condition gives
\[
\frac{dn_i}{d\varepsilon} = \frac{\sum_j c_j'}{(\lambda c_j'' - m_j'') \sum_j c_j^2} \left( (\delta_i - \delta_j) c_j - c_j' \right). \quad (A)
\]

**Proof of Proposition 1**
(a) Set $\delta_i=1$ and $\delta_j=0$ for $j \neq i$ in expression (A) and conclude using the identity $dn_j/d\alpha_i = dn_j/d\varepsilon$.
(b) Taking derivative of the average performance of group $i$ with respect to $\alpha_i$
\[
\frac{d}{d\alpha_i} \left[ \frac{m_i(n_i)}{n_i} \right] = \frac{(m_i' n_i - m_i)}{n_i^2} \frac{d}{d\alpha_i} n_i.
\]
But since $m_i$ is concave, we have $m'(n)n \leq m(n)-m(0)$, and the assumption $m(0) \geq 0$ implies $m_i'n_i-m_i<0$. We conclude that $d[m_i(n_i)/n_i]/d\alpha_i<0$. QED

**Proof of Proposition 2**
(a) Under the assumptions stated in proposition 2, expression (A) becomes
\[
\frac{dn_i}{d\varepsilon} = \frac{-1}{m''} \left( \delta_i - \bar{\delta} \right).
\]

$\delta_i>\delta_k$ implies $\frac{dn_i}{d\varepsilon} > \frac{dn_k}{d\varepsilon}$ and since $\frac{dn_i}{d\varepsilon}$ is linear in $\delta_i$ we have $\Delta n_i/\Delta n_k>1$.
(b) We have, $(d/d\varepsilon)[m_i(n_i)/n_i]= - K_i(n_i) \left( \delta_i - \bar{\delta} \right)$ where $K_i$ is a positive function.
(c) We have, $(d/d\varepsilon)[m_i(n_i)/n_i]= - K \left( \delta_i - \bar{\delta} \right)$ where $K$ is a positive constant. QED
Appendix 3: Demographic Subgroup and Variable Construction

The use of the DOL PAW methodology was not mandated in JTPA and states could opt out of using them. We contacted all state agencies that had been in charge of administering JTPA and asked them whether they had used the PAW methodology during the time period of our study, program years 1993 through 1998. Of the 33 states that supplied this information, 29 indicated they used the methodology and 4 indicated they did not. We include in our analysis only the 463 training agencies residing in the 29 states that used the PAW.

Construction of terminee sample for analysis: The shares of subgroup terminees and the average performance outcomes were computed for each subgroup-agency-program year cell using data from the Standardized Program Information Report (SPIR). The SPIR dataset reports information for 715,576 terminees. In the 463 agencies where the PAW were used, we have complete demographic information for our 13 selected factors for 95.4 percent of the SPIR sample or 682,515 adult terminees, which accounts for 63% of the entire JTPA adult population during the 1993-1998 period. This subsample appears to be representative of the entire enrollee population, however. For example, our sample includes 66% female, 32% black, 21% high school dropouts, and 40% welfare recipients and these figures are almost identical to the ones corresponding to the JTPA population (66% female, 32% black, 22% high school dropouts, and 37% welfare recipients).

Construction of the demographic subgroups and subgroup shares: For our empirical analysis we use demographic subgroups rather than demographic characteristics. There were 24 adjustment factors used in the DOL’s adjustment model during our time period. All factors are binary (e.g. male/female). We omitted from our analysis the very small demographic subgroups by dropping the nine factors for which the factor’s minority realization represented 10 percent or fewer JTPA participants at any program year during our study time period (for example, because only 3 percent of participants were SSI recipient for six program years, we omitted the “SSI recipient” factor). We omitted two more factors due to missing demographic information on program participants for these factors. In the end, we used 13 adjustment factors in the analysis. These factors are marked with a start (*) in Table 1. These 13 adjustment factors yield 8,192 (=2^13) demographic subgroups for each of 463 training agencies, and thus 3,792,896 (=8,129×463) possible subgroup-agency combinations for each program year. Many of these 3,792,896 subgroup-agency cells were empty, prompting us to further limit the data. We excluded from the analysis all subgroup-

---

28The SPIR data were compiled by Social Policy Research Associates for the Office of Policy and Research Employment and Training Administration, U.S. Department of Labor. The data and documentation can be obtained from the W.E. Upjohn Institute for Employment Research.
29The percentages of terminees included in our analysis (relative to the entire population) for program years 1993, 1994, 1995, 1996, and 1997 were 53%, 65%, 65%, 67%, 65%, and 66%, respectively.
30Nine factors excluded from the analysis are 55 years old & over, High school dropout under 30, Handicapped, UI or UC claimant, SSI recipient, Limited English speaking, GA/RCA recipient, Veteran, and Homeless.
31Our cut-off for inclusion in the analysis was a 90 percent data availability rate. Thus because information about being “basic skills deficient” was reported for only 84% of participants and possessing “reading skills below 7th grade” was reported for only 89 percent of participants, those factors were omitted from our analysis.
agency combinations that had zero terminees in each of the six program years. Applying this criterion led us to drop 96% of the 3,792,896 possible subgroup-agency combinations. The final panel data includes 1,670 different subgroups for 463 agencies. The number of subgroups vary across agencies (Min=2, Max=1073) and there are on average 291.04 subgroups per agency. There are 134,755 (=463 × 291.04) subgroup-agency observations by program year. The final analysis for enrolment share used 738,689 subgroup-agency observations which is less than the total number of observations (808,530=134,755 × 6 program years) due to missing data on local economic conditions and demographic information. The enrolment subgroup shares were computed based on our sample of terminees (SPIR minus 4.6%) and thus sum to one.

Construction of the performance outcomes: Since under JTPA the follow-up performance outcomes were measured for only a subset of all terminees, the samples for the employment and earnings outcomes analysis (H2) are smaller than the sample for the enrolment analysis (H1). SPIR reports a follow-up employment outcome for 44% of terminees (N=297,352) and a follow-up weekly earnings outcome for 72% of the follow-up employment outcome sample (N=213,176). We construct the subgroups for the outcome analysis using the same method as above and obtain 164,488 subgroups-agency-year observations for the employment outcome and 122,467 subgroups-agency-year observations for the earnings outcome.

---

32 To save money, JTPA administrators estimated each training agency’s overall performance from the performance of a sample of terminees drawn randomly from the training agency’s terminee population.