ARE PROGRAM PARTICIPANTS GOOD EVALUATORS?

Jeffrey Smith
University of Michigan
UWO, IFS, NBER, IZA and PSI
econjeff@umich.edu

Alexander Whalley
University of Maryland
whalley@econ.umd.edu

Nathaniel Wilcox
University of Houston

Version of July 21, 2005

We thank the W.E. Upjohn Institute for funding this research. We are very grateful to Dan Black and audience members and discussants at University College London (especially Richard Blundell, Hidehiko Ichimura and Costas Meghir), the 2004 APPAM meetings in Atlanta and the 2004 SEA meetings in New Orleans for helpful comments. We thank participants in a 2002 OECD conference on evaluating local economic development programs for inspiration. Any and all scandalous mistakes and shocking opinions remain our own.

Abstract

Participants, like econometricians, may have difficulty in constructing the counterfactual outcome required to estimate the impact of a program. In this paper, we directly assess this question by examining the extent to which program participants are able to estimate their individual program impacts ex-post. Utilizing experimental data from the National Job Training Partnership Act (JTPA) Study (NJS) we compare experimentally estimated program impacts to individual self-reports of program effectiveness after the completion of the program. We implement two methods to estimate the individual experimental impacts base on: (1) subgroup variation; (2) the assumption of perfect rank correlation in impacts. Little evidence of a relationship between the experimentally estimated program impacts and self-reported program effectiveness is found. We do find evidence that cognitively inexpensive potential proxies for program impacts such as before-after differences in earnings, the type of training received, and labor market outcomes are correlated with self-reported program effectiveness.

1.0 Introduction

Systematic and rigorous program evaluation represents an important component of evidence-based policy. Recent developments in econometric evaluation methods, summarized in, e.g., Heckman, LaLonde and Smith (1999) and Abbring, Heckman and Vytlacil (2005), have been rapid and substantively important. Social experiments, virtually unknown before 1970, are now frequently used to evaluate a wide variety of economic, social and criminal justice policies; see the exhaustive list in Greenberg and Shroder (2004).

At the same time, participant evaluations have gained attention as a complement to, or substitute for, experimental or econometric evaluation of such programs. Participant evaluation builds on responses by program participants to survey questions about whether or not the program helped them. While the specific question wording, as well as the number and specificity of the questions, varies substantially among programs, the data from many, if not most, econometric and experimental evaluations that rely on survey data for their outcome measures (rather than, or in addition to, administrative data) include participant evaluations.

This paper compares econometric estimates of program impacts at the individual or subgroup level with individual participant evaluations using the rich data from the U.S. National Job Training Partnership Act (JTPA) Study (NJS). JTPA was the major employment and training program for the disadvantaged in the U.S. during the 1980s and 1990s. Section 2 describes the JTPA experimental evaluation and the data it generated in detail, including the specific structure and wording of the participant evaluation survey question.

We consider two big picture interpretations of our comparisons between econometric impact estimates and participant evaluations. The first interpretation assumes the consistency of both estimators, but views them as estimating different parameters. In this view, the econometric impact estimates consistently measure the treatment effect of a program on a specific outcome, such as earnings or employment, over a specific time period. The participant evaluation, in contrast, consistently estimates the treatment effect of the program on participant utility, with the reference period depending on the question wording (and perhaps on participant interpretation conditional on the wording). In this context, the relationship between the two measures provides information on the relative importance of impacts on earnings and employment in the period covered by the econometric impact estimate to participants' overall benefit (or lack thereof) from program participation.

The second interpretation presumes that participants, like econometricians, have difficulty constructing the counterfactual outcome required to estimate the impact of a program. In this view, because constructing counterfactuals constitutes a cognitively difficult task, participants may (implicitly) rely on crude, cognitively inexpensive methods of impact estimation in constructing their self-reported impacts. The cognitively inexpensive alternatives to consistent estimation of the counterfactual we consider include simple before-after comparisons, and the use of inputs (type of training received) and outcomes (what happened in the labor market without reference to a counterfactual) as proxies for impacts.

The second interpretation of the relationship between econometric impact estimates (including proxy measures) and participant evaluations has a number of empirical implications. First, consistent econometric estimates may have only a weak relationship with participant evaluations, even if participants care a lot about the outcome in question over the period covered by the econometric estimates. Second, comparisons of the strength of the relationship between participant evaluations and estimates produced by alternative impact estimators may shed light on the particular estimator implicitly employed by program participants in responding to survey questions on program benefits. Third, participant evaluations may exhibit a strong relationship with econometric impacts constructed using the same crude estimators that the participants implicitly use.

We also consider the relationship between the impact proxies commonly used in administrative performance standards systems for employment and training programs and participant self-evaluations. These measures, which consist of simple but poor proxies for long run impacts on earnings and employment, represent a bureaucratic solution to the difficulties associated with constructing proper impact estimates quickly and at low cost. Their simplicity and low cost suggest that participants may (implicitly) rely on them as well in constructing their survey responses. In addition, this analysis has independent interest in that it suggests the extent to which participant evaluations might substitute for these measures in administrative performance systems.

In addition to informing decisions about how best to evaluate policies, our research has broader implications. First, whether or not individuals can accurately assess their program impacts, and how they go wrong if they cannot, has implications for the interpretation of instrumental variables estimates in the context of the correlated random coefficient model, as in Angrist (2004), Heckman (1997a), Heckman and Vytlacil (2005) and Carniero, Heckman and Vytlacil (2005). In that model, complications arise when using instruments that correlated with the individual-specific component of impacts. Those problems go away if individuals do not know their impacts (that is, if they make decisions based on "noise"). Of course, if individuals use biased estimates of their impacts in making decisions, the problems may return in a different form depending on how the bias relates to the instrument. Second, and more broadly, the ability to individuals to accurately envision the outcomes associated with alternative choices lies at the heart of rational models of human behavior. We return to this broader issue in the conclusion and offer some thoughts regarding the meaning of our results in terms of this broader question.

We have identified little in the way of existing literature that tries to link objective impact estimates with subjective participant evaluations. The most directly related analyses are those of Heckman and Smith (1998) and Philipson and Hedges (1998), both of which use treatment group dropout as a crude indicator of participants' evaluations. More broadly, Jacob and Lefgren (2005) compare principals' subjective evaluations of teachers to econometric estimates of teacher value added, but do not consider the teachers' evaluations of their own value-added. Prendergast (1999) reviews the literature on subjective performance evaluation, but that literature primarily views subjective evaluations as a way to deal with situations in which agents have multiple tasks (the outputs from some but not all of which allow objective measurement), not as a potentially cost-saving alternative to objective evaluation. That literature is also focused mainly on performance evaluation of workers within firms, not evaluation of the effects of programs on participant labor market outcomes.

To foreshadow our main findings, the data indicate that the impact estimates we prefer – the ones using subgroup variation in experimental impacts and the ones based on quantile differences – have little relationship with self-reported impacts. At the same time, inputs, outcomes and simple before-after estimates all do predict self-reported impacts, often strongly so.

We organize the remainder of the paper as follows. Section 2 describes the data from the JTPA experiment and the basic structure of the JTPA program. Section 3 presents the conceptual framework that guides our econometric analysis and our interpretation of our results. Section 4 discusses the construction and interpretation of the alternative econometric estimates of program impact on employment and earnings that we construct using the experimental data. Section 5 presents our results on the relationship between participants' self-reported impacts and impacts estimated using the experimental data. Section 6 examines the relationship between self-reported impacts and before-after employment and earnings changes, as well as proxies such as inputs and outcomes while Section 7 examines the relationship between self-reported evaluations and performance measures. Finally, Section 8 lays out the conclusions that we draw from our analysis.

2.0 Data and institutions

2.1 The JTPA program

The U.S. Job Training Partnership Act program was the primary federal program providing employment and training services to the disadvantaged from 1982, when it replaced the Comprehensive Employment and Training Act (CETA) program, to 1998, when it was replaced by the Workforce Investment Act (WIA) program. All of these programs share more or less the same set of services (though the latter two omit the public sector jobs that led to scandal under CETA) and serve the same basic groups. They differ primarily in their organizational details (i.e. do cities or counties play the primary role) and in the emphasis on, and ordering of, the various services provided. Nonetheless, the commonalities dominate with the implication that our results for JTPA likely generalize to WIA (and CETA).[1]

The JTPA eligibility rules included categorical eligibility for individuals receiving means tested transfers such as Aid to Families with Dependent Children (AFDC) or its successor Temporary Aid to Needy Families (TANF) as well as food stamps. In addition, individuals were eligible if their family income in the preceding six months fell below a specific cutoff value. There were also special eligibility rules for a number of small groups and a 10 percent "audit" window that basically allowed local sites to enroll individuals at their own discretion. See Devine and Heckman (1996) for more details on the JTPA eligibility rules and Kemple, Doolittle and Wallace (1993) for detailed descriptive statistics on the experimental sample in the NJS. Heckman and Smith (1999, 2004) provide thorough analyses of the determinants of participation in JTPA conditional on eligibility.

The JTPA program provided five major services: classroom training in occupational skills (CT-OS), subsidized on-the-job training (OJT), job search assistance (JSA), adult basic education (ABE) and subsidized work experience (WE). Local sites

[1] One possible caveat is that potential participants may have better information to guide them in making participation decisions about a relatively old program, as JTPA was at the time of the experiment, then about a relatively new program. This reasoning suggests greater selection on impacts over time as a program matures.

had the flexibility to emphasize or de-emphasize particular services in response to the needs of the local population and the availability of local service providers. In general, CT-OS was the most expensive service, followed by OJT, ABE and WE. JSA costs a lot less. See Heinrich, Marschke and Zhang (1998) for a detailed study of costs in JTPA and Wood (1995) for information on costs at the NJS study sites.

Services get assigned to individuals by caseworkers, typically as the result of a decision process that incorporates the participant's abilities and desires. This process leads to clear patterns in terms of the observable characteristics of participants assigned to each service. The most job ready individuals typically get assigned to JSA or OJT, while less job ready individuals typically get assigned to CT-OS, BE or WE, where CT-OS often gets followed by JSA. See Kemple, Doolittle and Wallace (1993) for more about the service assignment process. This strongly non-random assignment process has implications for out analyses below in which we examine the relationship between the participant evaluations and types of services they receive.

2.2 The National JTPA Study data

The National JTPA Study (NJS) evaluated the JTPA program using a random assignment design. It was the first major social experiment to evaluate an ongoing program rather than a demonstration program brought into existence solely for the purposes of the experiment. Random assignment in the NJS took place at a non-random sample of 16 of the more than 600 JTPS Service Delivery Areas (SDAs). Each SDA had a local geographic monopoly on the provision of employment and training services funded under the JTPA. The exact period of random assignment varied among the sites, but in most cases random assignment ran from late 1987 or early 1988 until sometime in spring or summer of 1989. A total of 20,601 individuals were random assigned, usually but not always with the probability of assignment to the treatment group set at 0.67.

The NJS data come from multiple sources. First, respondents completed a Background Information Form (BIF) at the time of random assignment. The BIF collected basic demographic information along with information on past schooling and training and on labor market outcomes at the time of random assignment and earlier. Second, all experimental sample members were asked to complete the first follow-up survey around 18 months after random assignment. This survey collected information on employment and training services (and any formal schooling) received in the period since random assignment, as well as monthly information on employment, hours and wages, from which a monthly earnings measure was constructed. Third, a random subset (for budgetary reasons) of the experimental sample members was asked to complete a second follow-up survey around 32 months after random assignment. This survey collected similar information for the period since the completion of the first follow-up survey or, in the case of individuals who did not complete the first follow-up survey, over the period since random assignment. Response rates to both follow-up surveys were around 80 percent. Finally, administrative data on quarterly earnings and unemployment from state UI records in the states corresponding to the 16 NJS states were collected.[2] See Doolittle and Traeger (1990) on the design of the NJS, Orr et al. (1996) and Bloom et al. (1997) for the official impact reports and Heckman and Smith (2000) and Heckman, Hohmann,

[2] These data were collected twice, once for 12 of the 16 sites by Abt Associates, one of the prime contractors on the original experiment, and then for all 16 sites later on by Westat under a separate contract. We use the latter dataset in our analysis.

Smith and Khoo (2000) for further interpretation. Appendix 1 describes the data used in this study in greater detail.

2.3 The self-evaluation questions

Exhibit 1 presents the two survey questions that, taken together, define the participant evaluation measure we use in this paper. The question appears on both the first follow-up survey and the second follow-up survey. In both cases, the skip pattern in the survey excluded control group members from both questions. Respondents in the treatment group were asked these questions in the second follow-up survey only if they did not complete the first follow-up survey.

The first question asks treatment group members whether or not they participated in JTPA. The question assumes application because it is implied by the respondent having been randomly assigned. The JTPA program had different names in the various sites participating in the evaluation; the interviewer included the appropriate local name in each site as indicated in the question.

In the second question, respondents who self-report having participated in the program get asked whether the program helped them get a job or perform better on the job. This is not the ideal question from our point of view, as it focuses more on a specific outcome than on an overall impact, but it is what we have to work with in the JTPA evaluation. However, to the extent that it focuses respondents' attention specifically on the effect of program participation on labor market outcomes, it should increase the strength of the relationship between the participant evaluations and the econometric estimates of labor market impacts, relative to a broader question that asked about generic program benefits.

We code the responses to both questions as dummy variables. The participant evaluation measure employed in our empirical work consists of the product of the two dummy variables. Put differently, our self-reported evaluation measure equals one if the respondent replies "YES" to question (D7), and "YES" to question (D9). Otherwise, it equals zero.

3.0 Conceptual framework

3.1 A simple model of participants' self-reported evaluations

In this section, we lay out a model of how individual participants might respond to a question regarding whether or not they benefited from a program. The discussion here is inspired by those in Manski (1990) and Dominitz and Manski (1994), who provide careful economic (and econometric) analyses of responses to questions about fertility intentions and returns to schooling, respectively. Our (very) simple model helps to structure the design and interpretation of our empirical work.

To begin, we suppose that respondents compare their observed utility given participation with the utility they would have experienced had they not participated. Let U_1 denote utility given participation, U_0 denote participation given non-participation and let $\Delta_{SR} \in \{0,1\}$ denote the response to the self-evaluation question. Then if respondents generate their answer by comparing the two utilities, we have

$$\Delta_{SR} = 1(U_1 > U_0).$$

This formulation ignores the timing of any affects of participation on utility relative to the survey response. Depending on the wording of the survey question and the respondents' interpretation thereof, respondents may focus on impacts during the period up to the survey response, after the survey response, or some combination of the two. In the JTPA context, we expect them to focus primarily on the effects of the program in the period leading up to the survey response. Expanding our notation, let the subscript "b" denote the period before the survey response and the subscript "a" denote the period following the survey response. We can then write

$$\Delta_{SR} = 1(U_1 > U_0) = 1(f(U_{1b}, E(U_{1a})) > f(U_{0b}, E(U_{0a}))),$$

where $f(\ )$ is an increasing function of both its arguments that maps the utility associated with participation or non-participation, both before and after the self-reported evaluation, into an overall valuation.

Next we consider what aspects of participation affect the utility levels of individuals. In particular, we can decompose the impacts that individuals experience into components related to earnings or employment and a residual component that includes other direct costs and benefits as well as psychic costs and benefits. Denote labor market impacts in the standard notation in the evaluation literature as

$$\Delta_Y = Y_1 - Y_0,$$

where Y_1 denotes the labor market outcome in the treated state and Y_0 denotes the labor market outcome in the untreated state. Similarly, denote the impact on all other determinants of participant utility by

$$\Delta_B = B_1 - B_0,$$

where B_1 and B_0 parallel Y_1 and Y_0 in their interpretation. In what follows, we will further distinguish between impacts realized before and after the survey response.

This decomposition into impacts on labor market outcomes and on all other outcomes that individuals care about corresponds to the components of the impacts that we can and cannot estimate econometrically using our data. The outcomes we (and hopefully the respondents) have in mind other than labor market outcomes include direct costs of participating in training, such as transportation and childcare expenses, leisure time as in Greenberg (1997), as well as any psychic costs and benefits from participating. Rewriting the survey response function in terms of this additional notation yields

$$\Delta_{SR} = 1(U_1 > U_0) = 1(U(Y_{1b}, B_{1b}, E(Y_{1a}), E(B_{1a})) > U(Y_{0b}, B_{0b}, E(Y_{0a}), E(B_{0a})),$$

or, alternatively

(1) $$\Delta_{SR} = 1(U_1 > U_0) = 1(g(Y_{1b} - Y_{0b}, B_{1b} - B_{0b}, E(Y_{1a} - Y_{0a}), E(B_{1a} - B_{0a})).$$

We estimate two variants of equation (1), one in cases where we have examine econometric estimates of $(Y_{1b} - Y_{0b})$ and another in cases where we examine simple proxies for $(Y_{1b} - Y_{0b})$. The next two subsections define these variants.

3.2 Econometric specification using econometric impact estimates

In the case of the econometric impact estimates, we begin by assuming additive separability of the $g(\ )$ function into components related to the labor market impact in the period prior to the survey and the remainder of the function.[3] Assuming that the utility function is monotonic in its arguments, we can then rewrite the relationship to put the labor market impact on the left hand side, yielding

(2) $$Y_{1b} - Y_{0b} = h(\Delta_{SR}, B_{1b} - B_{0b}, E(Y_{1a} - Y_{0a}), E(B_{1a} - B_{0a})).$$

We actually estimate linear versions of (2), given by

(3) $$Y_{1b} - Y_{0b} = \beta_0 + \beta_1 \Delta_{SR} + \varepsilon,$$

[3] Additive separability is not innocuous here; it implies no complementarities between the component of the impacts we estimate and the other components of the impacts.

where the hat on the impact denotes an estimate and where ε includes the idiosyncratic pieces of $B_{1b} - B_{0b}$, $E(Y_{1a} - Y_{0a})$ and $E(B_{1a} - B_{0a})$ (the means are captured in the intercept) as well as the estimation error in the impact estimate and any approximation error due to inappropriate linearization.

We adopt this formulation in the case of the econometric impact estimates for two reasons. First, because the econometric impact estimate includes estimation error, we want to put it on the left hand side for the usual reasons associated with measurement error. In contrast, the participant evaluation has no measurement error; the variable is defined as the response to the survey question.[4] We include no additional covariates on the right hand side because one of our two econometric estimates (described in detail in Section 4.1) consists of predicted subgroup experimental impact estimates. To include both these predicted impacts and a set of observables would require excluding at least one observable from this equation, but including it among the observables used to construct the subgroup impacts. The observables available to us lack an obvious candidate for exclusion.

Under the first interpretation of our analysis, a weak estimated relationship in equation (3) indicates that participants care primarily about something other than earnings or employment impacts in the period prior to the survey. This conclusion requires the qualification that we should not forget what lies in the error term. Among the items in the error term, we would expect long term impacts to correlate positively with short term impacts; in contrast, impacts on leisure likely correlate negatively with impacts on labor market outcomes prior to the survey. A weak relationship in (3) could

[4] The counter-argument in favor of making the participant evaluation the dependent variable despite the estimation error in our impact estimates relies on the econometric impacts having the larger variance of the two variables.

thus also result from a combination of a positive direct effect of impacts on employment or earnings and a negative indirect effect on leisure, working through the correlation between the omitted impact on leisure and the included impact on employment or earnings. Finally, in a common effect world in which the program has the same impact on everyone, or in which the impact varies but the idiosyncratic component is unknown even to participants, the true coefficient on the econometric estimate in (3) equals zero.

Under the second interpretation of our analysis, the absence of a relationship between the participant impacts and our econometric estimates has an additional possible meaning, namely that the respondents have used some other, less cognitively taxing, estimator to (implicitly) construct their own impact estimates. In this case, our econometric estimates will have a weak relationship with the participants' survey responses, but variables related to the chosen alternative estimator, including crude proxy variables if that is what respondents rely on, should display a strong relationship with the participants' self-reported evaluations.

Finally, under either interpretation, large estimated standard errors suggest that our econometric impact estimates embody substantial estimation error.

3.3. Econometric specification: impact proxies and performance measures

In the case of the proxy variables and the simple performance measures that act as proxies for impacts, we adopt a more direct analog to equation (1) as our econometric specification. In particular, we assume that

(4) $$\Delta_{SR} = 1(\beta_0 + \beta_1(proxy(Y_{1b} - Y_{0b})) + \beta_X X + \varepsilon > 0),$$

where ε has a logistic distribution and X is a vector of observable characteristics with corresponding coefficients β_X. This is, of course, a standard logit model, which means that we can identify the coefficients only up to scale; we report estimates of mean derivatives below.

We employ (4) rather than (3) in this case because the proxies for impacts that we examine, such as labor market outcomes and the types of services received, unlike our econometric impact estimates, do not contain any measurement or estimation error. In addition, because we measure these variables directly, rather than predicting them as a linear combination of the X, we can include conditioning variables X. These conditioning variables soak up residual variance and thus make our estimates more precise. They may also proxy, in part, for $B_{1b} - B_{0b}$, $E(Y_{1a} - Y_{0a})$ and $E(B_{1a} - B_{0a})$, thus clarifying the interpretation of our estimates.

4.0 Econometric impact estimators

This section describes the two econometric estimators that we apply to the experimental data to obtain impact estimates that vary among participants.

4.1 Experimental impacts at the subgroup level

The first method we employ for generating impact estimates that vary among participants takes advantage of the experimental data and the fact that random assignment remains valid for subgroups defined based on characteristics measured at or before random assignment, as discussed in, e.g. Heckman (1997b).

We estimate regressions of the form

(5) $$Y_i = \beta_0 + \beta_D D_i + \beta_X X_i + \beta_I D_i X_i + \varepsilon_i ,$$

where Y_i is some outcome measure, D_i denotes assignment to the experimental treatment group, X_i denotes a vector of characteristics measured at or before random assignment and $D_i X_i$ represent interactions between the characteristics and the treatment indicator. It is these terms that yield variation in predicted impacts among individuals at the subgroup level. The predicted impacts based on (5) for the treatment group members are given by

(6) $$Y_{1i} - Y_{0i} = \hat{\beta}_D + \hat{\beta}_I X_i .$$

Though quite straightforward conceptually, our experimental subgroup impact estimates do raise a few important issues, which we now discuss. The first issue concerns the choice of variables to interact with the treatment indicator. We address this issue by presenting two sets of estimates based on vectors of characteristics selected in very different ways. One set of estimates simply borrows the vector of characteristics employed by Heckman, Heinrich and Smith (2002) in their analysis of the JTPA data. The notes to Table 3 list these variables. The second set of estimates utilizes a set of characteristics selected using the somewhat unsavory method of stepwise regression. While economists typically shun stepwise procedures as atheoretic, for our purposes here that bug becomes a feature, as it makes the variable selection procedure completely mechanical. Thus, we can be assured of not having stacked the deck in one direction or another. In both cases, we restrict our attention to main effects in order to keep the problem manageable.

We implement the stepwise procedure using essentially all of the variables from the BIF including variables measuring participant demographics, site, receipt of means-tested monetary and in-kind transfers, labor force status and work history. We include a missing indicator for each variable (to avoid losing a large fraction of the sample due to item non-response), and interact both the variables and the missing indicators with the treatment group indicator. The stepwise procedure has to keep or drop each variable along with the missing indicator and interactions with the treatment indicator as a group. The stepwise procedure, which we perform separately for each of the four demographic groups, iteratively drops variables with coefficients not statistically different from zero in a regression with self-reported earnings in the 18 months after random assignment as the dependent variable.[5]

The second issue concerns the amount of subgroup variation in impacts in the NJS data within the four demographic groups – adult males and females ages 22 and older and male and female out-of-school youth ages 16-21 – for which both we and the official reports conduct separate analyses. Although the NJS impact estimates differ substantially between youth and adults (and between male and female youth when considering the full samples), the experimental evaluation reports – see Exhibits 4.15, 5.14, 6.6 and 6.5 in Bloom et al. (1993) for the 18 month impacts and Exhibits 5.8, 5.9, 5.19 and 5.20 in Orr, et al. (1994) for the 30 month impacts – do not reveal a huge amount of statistically significant variation in impacts among subgroups defined by the observables available on the BIF. If the impact does in fact vary a lot among individuals, but not in a way that is correlated with the characteristics we use in our model, then we

[5] We employ the "step up" stepwise procedure as it has more power than the "step down" and "single step" procedures. See Dunnett and Tamhane (1992) and Lui (1997) for details. We set the p-value for choosing variables in the final specification at 0.05.

may reach the wrong conclusions about participants' ability to construct consistent estimates of earnings impacts. This case has more than academic interest given that Heckman, Smith and Clements (1997) bound the variance of the impacts in the JTPA data away from zero for adult women; their lower bound on the standard deviation of the impacts equals \$674.50 with a standard error of \$137.53 (see their Table 3).[6] In addition to simply keeping it in mind, we attempt to address this concern in part by examining the quantile treatment effect estimates described in the next section, which do vary a lot among participants, and by looking, in other work, at data from other experimental evaluations with more in the way of subgroup variation in impacts.

The third issue concerns an additional assumption that we must make in order to interpret out results in the way that we have described here. A simple example illustrates the need for this assumption. Consider two subgroups and suppose that participants care only about earnings impacts, and give a positive survey evaluation when they have an earnings impact greater than zero. In group one, suppose that 10 percent of the individuals have an impact of 1000 while 90 percent have an impact of zero. The mean impact for subgroup one thus equals 100, while the fraction of positive participant evaluations equals 0.1. In contrast, in group two, 20 percent of the individuals have an impact of 400 while 80 percent have an impact of zero. The mean impact for subgroup two thus equals 80 while the fraction of positive participant evaluations equals 0.2. In this example, when comparing across the two subgroups the mean impact varies inversely with the fraction with a positive impact. In interpreting our results below, we

[6] Our subgroup impacts have standard deviations that range from \$840 to \$2600 depending on the demographic group and set of covariates. The quantile treatment effects have lower standard deviations; they range between \$257 and \$477.

assume that this case does not hold in the data. Put differently, we assume that mean impacts and the fraction with a positive impact positively co-vary at the subgroup level.

4.2 Quantile treatment effects

The second econometric method we use to derive individual level treatment effect estimates relies on an additional non-experimental assumption. In particular, we make the assumption of a perfect positive rank correlation between the outcomes in the treated and untreated states described in Heckman, Smith and Clements (1997). Intuitively, we assume that the expected counterfactual for an individual at a given percentile of the treatment group outcome distribution consists of the mean outcome at the same percentile of the control group outcome distribution. One way to think about this assumption is that expected labor market outcomes depend on a single factor, so that individuals who do well in the treatment state also do well in the control state. This represents a very different view of the world than, for example, the classic model of Roy (1951), but may represent a reasonable approximation for treatments, such as those offered by JTPA, that have small average impacts relative to the mean of the outcomes in question.

Using this method, we estimate the impact for treated individual "*i*" with an outcome at percentile "*j*" of the treatment group outcome distribution as

$$(7) \qquad Y_{1i} - Y_{i0} = \hat{Y}_1^{(j)} - \hat{Y}_0^{(j)},$$

where the superscript "(j)" denotes the percentile. This estimator underlies the literature on quantile treatment effects, as in Abadie, Angrist and Imbens (2002) and Bitler, Gelbach and Hoynes (2004), with the difference that rather than interpreting the estimates as the effect of treatment on the quantiles of the outcome distribution, we make the additional rank correlation assumption. As discussed in Heckman, Smith and Clements (1997), the rank correlation assumption pins down the joint distribution of outcomes, which in turn pins down which quantile of the control outcome distribution provides the counterfactual for each quantile of the treatment outcome distribution and allows us to assign impact estimates to specific individuals.

5.0 The relationship between econometric impact estimates and participant evaluations

5.1 Bivariate relationships

We begin our analysis of the data from the NJS with simple bivariate relationships between mean experimental impacts for a variety of labor market outcomes and the fraction of participants with a positive self-reported evaluation for the four demographic groups in the experiment. This analysis, presented in Table 1, extends that presented in Table 8.11 of Heckman and Smith (1998). It represents a very basic application of the methodology outlined in Section 4.1.

The first four rows in Table 1 correspond to the four demographic groups described above. The first column presents the fraction of the experimental treatment group with a positive self-reported evaluation. The remaining columns report mean impacts on eight different earnings and employment outcomes. The first two measures consist of self-reported earnings and any self-reported employment in the first 18 months after random assignment, which roughly corresponds to the period prior to the survey response for most sample members. The second two measures consist of self-reported earnings and employment in month 18 after random assignment, thus focusing on the respondent's status right around the time of the survey, rather than over the entire period since random assignment. The remaining four measures repeat the first four, but now using the quarterly data from the matched UI earnings records rather than the self-reported outcome data. We include both sets because they appear substantially different at both the individual and aggregate levels – see the discussions in Smith (1997a,b) and Kornfeld and Bloom (1999). The final row of Table 1 reports the correlation between the percent with a positive self-reported evaluation and the impact estimates in each column and the p-value from a test of the null hypothesis that the population correlation coefficient equals zero.

Table 1 reveals that the subgroup with the worst impact on all eight of the outcome measures, namely male youth, has the second highest fraction with a positive self-reported evaluation. The group with the highest fraction with a positive self-reported evaluation, namely female youth, often has the second-worst impact estimate. Consistent with this basic pattern, the correlations reported in the last row end up negative six out of eight times, though we can never reject the null of a zero correlation (which does not surprise given we have only four groups). Thus, at this crude level, we find very little in the way of an association between the participant evaluations and the econometric estimates that rely on subgroup variation; as noted above, this may mean that earnings and employment do not figure much in respondents' evaluations of JTPA, or it may mean that respondents do not do a very good job of constructing the relevant counterfactual.

Looking specifically at the employment impacts, we might expect improved performance given the focus of the actual question on help in finding a job. However, the signs of the correlation differ between employment measures for both data sets and between data sets for each measure. More broadly, if we make the rank correlation assumption described in Section 4.2, the experimental impacts show that the program improved the employment situation of at most a few percent of the respondents, yet well over half self-report a positive impact on the survey.

Tables 2A, 2B, 2C and 2D report the results of a similar bivariate analysis using variation in the experimental impacts among subgroups of the four demographic groups. Each table corresponds to one of the four demographic groups. Within each table, the rows correspond to the variables used to define the subgroups and the columns refer to the same eight labor market outcome variables considered in Table 1. The variables we use are race/ethnicity, years of schooling categories, marital status, time since last employment categories, site and age categories, where we omit the age variable for youth as the group is limited to individuals from 16 to 21 in any case. Each entry in the table gives the estimated correlation coefficient between the fraction with a positive self-reported evaluation and the experimental impact estimate for the outcome variable for the column and for the subgroups defined by the row variable.

The bottom of each table also presents some summary statistics. In particular, we present the number of positive and negative correlations in the table and, within each of these categories, the fraction statistically significant at the five and ten percent levels. For the adults, we would expect random variation to lead to 4 or 5 estimates statistically significant at the 10 percent level and 2 or 3 at the five percent level. These vote counts provide a useful but imperfect summary of the 48 (or 40 for the youth) entries in each table. In particular, the vote counts ignore the lack of independence among the estimates in each table and do not make any attempt to weight or value the estimates based on their precision.

The results in Table 2 paint a picture that looks a lot like the one from Table 1. No clear patterns emerge in terms of coefficient signs and the number of statistically significant correlations looks about like what you would expect if the population coefficients all equal zero. So far, our findings strongly suggest either that participants weight labor market outcomes over the period prior to the survey very little in evaluating JTPA, or that participants do care about impacts but do a very bad job of estimating them, or that the experimental impact estimates based on subgroup variation have only a weak correlation with actual individual impacts.

5.2 Regression results for experimental subgroup estimates

We now to turn evidence from regressions of estimated impacts on each of the labor market outcomes considered in Tables 1 and 2 on the indicator variable for a positive self-reported evaluation. In terms of our earlier discussion, we report estimates of β_1 from equation (3), where the dependent variable consists of an experimental impact estimate based on subgroup variation in impacts as in equations (5) and (6). These estimates appear in Table 3, where each entry in the table represents a separate regression. The rows correspond to particular labor market outcomes. Each of the four demographic groups has two columns of estimates, one for each of the two sets of covariates used in estimating equation (5). The columns headed by (1) contain the estimates using the covariates from Heckman, Heinrich and Smith (2002), while the columns headed by (2) contain the estimates using the covariate set chosen by the stepwise procedure. The final two rows of the table summarize the evidence in each column; in particular, they give the numbers of positive and negative estimates and, within each category, the number of statistically significant estimates at the five and ten percent levels.

The evidence in Table 3 suggests little, if any, relationship between the experimental impact estimates based on subgroup variation and the self-reported evaluations. While the estimates lean negative in the aggregate, only a handful of the estimates reach conventional levels of statistical significance (and not all of those fall on the negative side of the ledger). The regression evidence thus compounds the evidence from the simple bivariate relationships examined in Tables 1 and 2. Either the participants do not weigh labor market impacts very heavily in their response, or else their impact estimates (or ours) do not do a very good job of capturing the actual impacts.

5.3 Results based on quantile treatment effect estimates

This section presents evidence on the relationship between impact estimates constructed under the perfect positive rank correlation assumption described in Section 4.2. We present these results in both graphical and tabular form. Figures 1A to 1D present the evidence in graphical form, with one figure for each demographic group. The horizontal axis in each figure corresponds to percentiles of the untreated outcome distribution. The solid line in each graph presents impact estimates at every fifth percentile (5, 10, 15, … , 95) constructed as in equation (7). The broken line in each graph represents an estimate of the fraction with a positive self-reported evaluation at every fifth percentile. For percentile "j", this estimate consists of the fraction of the treatment group sample members in the interval between percentile "j-2.5" and percentile "j+2.5" with a positive self-reported evaluation. If the assumptions underlying the percentile difference estimator hold, if participants care about labor market outcomes in answering the survey question, and if participants consistently estimate their own impacts, then the two lines should move together in the figures.

Several features of the figures merit notice. First, in the lower percentiles in each figure the econometric impact estimate equals zero. This results from the fact that the lowest percentiles in both the treated and untreated outcome distributions have zero earnings in the 18 months after random assignment; the difference between the two then equals zero as well. Surprisingly, a substantial fraction (over half in all four demographic groups) of treatment group members at these percentiles respond positively to the survey evaluation question, even though they have zero earnings in the 18 months after random assignment. This suggests that respondents view the question as asking about longer term labor market impacts and not solely as a narrow question about finding a job immediately after participation in the program.

Second, the fraction with a positive self-reported evaluation has remarkably little variation across percentiles of the outcome distribution. For all four demographic groups, it remains within a band from 0.6 to 0.8. For the adults, the mean increases with the percentile; for the youth, the data fail to reveal a clear pattern.

Third, no obvious relationship between the two variables emerges from the figures for three of the four demographic groups. Adult women constitute the exception; for them, both variables increase with the percentile of the outcome distribution. More specifically, for adult women, both variables have a higher level for percentiles where the impact estimate exceeds zero. Within the two intervals defined by this point, both variables remain more or less constant.

Table 4 presents some of the numbers underlying the figures. In particular, the first five rows present the values for the 5th, 25th, 50th, 75th and 95th percentiles. The last two rows of the table give the correlation between the quantile treatment effects and the fraction with a positive self-reported evaluation for each group (and the corresponding p-value from a test of the null that the correlation equals zero) along with the estimated coefficient from a regression of the quantile treatment effects on the fraction with a positive self-reported evaluation (and the corresponding standard error). The correlation and regression estimates quantify and confirm what the figures indicate: a strong positive relationship for adult women, a weak and statistically insignificant positive relationship for adult men, and moderately strong and negative relationship for male youth and a similar, but not statistically significant, relationship for female youth. Although we find a bit more here than in the estimates that rely on subgroup variation, once again the data do not suggest a strong, consistent relationship between the econometric impact estimates and the self-reported evaluations.

6.0 Relationship between positive self-evaluation and proxies for impacts

6.1 Motivation

In this section, we present evidence on the extent to which simple proxies which respondents might use in constructing their impact estimates predict a positive self-reported evaluation. The proxies we examine include input (training type) measures, labor market outcome (employment and earnings) measures and simple before-after differences in labor market outcomes. For each of these proxies, we present estimates of equation (4) in Section 3.3. If the proxy variables drive the self-reported evaluations, this suggests that participants rely on these readily accessible variables in answering the survey evaluation question instead of thinking hard about their counterfactual outcome. If respondents really do know the impacts, then proxies poorly correlated with the actual impacts should have little explanatory power.

Two important caveats weaken this argument. First, if we have really imprecise (or inconsistent, in the case of the quantile treatment effects) econometric estimates, then we have no way of knowing the extent to which the proxies correlate with actual impacts; our view that they do not relies solely on our priors. Second, particularly in the case of the input measures, the proxies may correlate with both the psychic and the direct costs and benefits of participation not captured in the labor market outcomes we examine. For example, classroom training may be more fun than, say, job search assistance. Alternatively, classroom training may have higher direct costs if takes place at a distant community college while job search assistance takes place at a local neighborhood organization. With both the big picture and these caveats in mind, we now turn to our results.

6.2 Results with input and outcome measures

Tables 5 presents logit estimates of equation (4) that include not one but two measures of the training received by JTPA treatment group members. We interpret these inputs as potential proxies for impacts. Respondents receiving only inexpensive services (or no services at all) might reason that as a result the program can have had little if any impact, while participants who receive expensive services such as CT-OS or OJT may draw the opposite conclusion.

The two measures of service receipt derive from self-reports collected in the NJS follow-up surveys and from administrative data from the individual sites participating in the experiment.[7] As shown in Smith and Whalley (2005), these two measures differ substantially; as a result we do not run into collinearity problems when including them both. The two data sources code the service types somewhat differently; for comparability and ease of interpretation, we employ just five service types: CT-OS, OJT/WE (which is almost all OJT), JSA, ABE and "other". We code a dummy variable for each service type in each data source indicating whether or not the respondent received it; a respondent who received more than one service type in a given data source gets coded based on the training type they receive in their first spell.

The logit models presented in Table 5 also include a variety of background variables. These variables play two roles. First, we expect them to pick up parts of the overall impact of participation unrelated to the labor market outcomes we examine. For example, the site dummies will pick up differences in the friendliness and efficiency of site operation as perceived by the respondents. The variable "work for pay", which is an indicator variable for whether or not the respondent has ever worked for pay, relates to the opportunity cost of participation, as does the variable for having a young child. The The AFDC receipt at random assignment variable captures variation in the cost of classroom training due to the availability of an income source not tied to employment. In order to avoid losing a large fraction of the sample due to item non-response, we recode missing values to zero and include indicator variables for missing values of each variable.

[7] In fact, two versions of the administrative data on service receipt exist, one created by MDRC and one created by Abt Associates. Both rely on the original MIS files from the 16 sites in the experiment. Our experience with both files, described in detail in Smith and Whalley (2005), leads us to employ the Abt version in this paper.

Each column in Table 5 corresponds to one of the four demographic groups. The table presents mean derivatives, estimated standard errors for the mean derivatives in parentheses and the p-value of a test of the null hypothesis that the mean derivative equals zero in square brackets. Table 6 summarizes the results in Table 5 by presenting test statistics and p-values from tests of the joint null that the mean derivatives for groups of related covariates (e.g. all of the self-reported training type variables) equal zero.

Consider the variables other than the training type variables first. Although they are not shown in the table, the site variables have a strong effect on the probability of a positive self-reported evaluation. The magnitudes vary a lot as well; for example, for adult males the coefficients on the site dummies range from -0.257 to 0.093. Moreover, Table 6 shows that these variables are strongly statistically significant as a group. We interpret this as indicating that respondents take account of non-pecuniary aspects of their JTPA experience, such as the friendless and efficiency of the staff and the attractiveness and ease of access of the JTPA office and the local service providers. However, these variables may also proxy for site differences in program impacts (though this seems unlikely given the findings in Section 5.2) or for other features of the local environment, such as the state of the economy, that affect respondent's experiences.

With the exception of age for adults, race for youth, and age and education for female youth, the other demographic variables play surprisingly little role in determining the probably of a positive self-reported evaluation. Among adults, age has a strong negative effect on the probability of a positive self-evaluation, while black male youth and Hispanic male and female youth have higher probabilities of a positive response. The limited role played by background characteristics in the analysis surprised us.

In contrast to the background characteristics, the training type variables play a major role in determining individual self-reported evaluations. Table 6 shows that, taken together, both the self-reported and administrative training type variables achieve high levels of statistical significance for adults, and the administrative measures do so for youth.

Smith and Whalley (2005) show that the self-reported and administrative measures of receipt of classroom training in occupational skill tend to agree; as such, we can (as a crude approximation) simply add their coefficients. Doing so reveals that CT-OS has a large positive effect on the probability of a positive self-reported evaluation for all four groups. Subsidized on-the-job training reports often do not coincide in the two data sources; here we find that self-reported OJT has a strong (and usually statistically significant) positive effect, as does administratively reported OJT! CT-OS and OJT generally represent the largest resource investment in the JTPA participant; the participants appear to recognize this and use it as a proxy for impacts in responding to the self-reported evaluation questions.

Job search assistance, the cheapest of the services, elicits less of a positive effect. This training type also tends to get reported differently in the two data sources. Here, except for self-reported JSA for adult males and administratively reported JSA for adult females, we find modest and statistically insignificant effects. Adult Basic Education (ABE) and “other” training do not yield precise estimates, except for “other” training in the administrative data, which has, somewhat puzzlingly, a negative and statistically significant effect for adult women and a positive and statistically effect for female youth.

These two training types have, in general, smaller sample sizes than the others, which may account for the imprecision of the estimates.

Overall, the training type variables matter in predicting a positive self-reported evaluation in a manner consistent with the view that respondents use the intensity of the services they receive – the inputs – as a proxy for the impact the services have upon them.

Tables 7 and 8 describe a set of logit estimates of equation (4) in which we include the same background variables as in the models of Table 5 but add various versions of Y_1, the labor market outcome in the treated state. Respondents may reason that if they have done well in the labor market over the period between random assignment and the survey or if they are doing well around the time of the survey, then the program must have benefited them. Respondents who have not done well may draw the opposite conclusion. As with the training type variables, the outcomes represent an easily observable proxy for impacts that respondents may rely in when determining their responses to the survey evaluation question.

The top panel of Table 7 reports estimates from a specification in which we divide self-reported earnings in the 18 months after random assignment into five categories: zero, and four quartiles of the distribution of positive earnings, and then include dummy variables for four of the five categories, with the highest category as the excluded category. The second panel of Table 7 corresponds to a similar specification but using earnings in the six calendar quarters after random assignment from the UI administrative data. The last two lines report estimates from a specification that includes a dummy variable for any employment in the 18 months after random assignment, again measured first based on the survey data and then based on the UI administrative data. Table 8 summarizes the evidence in Table 7 as well as evidence from a series of alternative specifications not reported here for reasons of space. As in Table 6, the summary takes the form of p-values from tests of the null hypothesis that the coefficients or coefficients corresponding to a specific labor market outcome measure equal zero.

The broad picture from Table 7 is that labor market outcomes appear to predict self-reported evaluations. This is particularly true for adults. For both adult males and adult females and for both earnings measures, all of the estimated coefficients are negative (as expected when compared to the highest earnings quintile) and most are statistically significant. Also, broadly speaking, the estimated coefficients decrease as earnings increase, as expected. For youth, the coefficients also turn out largely negative (indicating a positive relationship) but rather imprecisely estimated. The self-reported employment measure also has a strong positive and statistically significant relationship to the self-reported evaluations, but the UI employment outcome measure does not. This latter finding may result from measurement error in UI employment due to its omission of government jobs and informal jobs.

Turning to Table 8, which summarizes a large number of specifications via chi-square tests and associated p-values (including both the specifications reported in Table 7 and many not reported there for reasons of space), we find similar patterns. The relationships tend to be statistically stronger for adults than for youth, and stronger for the measures based on the self-reported data than on the UI data. Also, the earnings measures tend to yield more statistically significant relationships than the employment measures, especially for measures that consider outcomes just around the time of the survey.

Overall, this section presents compelling evidence that simple proxies for impacts in the form of inputs and outcomes predict self-reported evaluations. This pattern of findings lends support to the view that respondents adopt cognitively simple alternatives to the difficult task of trying to construct a counterfactual outcome when answering the self-evaluation question.

At the same time, the role of the input measures may also reflect, in part, various non-pecuniary aspects of the services received. For example, in addition to representing a larger financial investment, respondents who receive classroom training may have more fun in their JTPA experience than those who receive, say, job search assistance. Also, outcomes may proxy in part for impacts (and with less measurement error than the impact measures we consider earlier in the paper); in a world where the counterfactual is zero earnings – the correct world for some fraction of the treatment group – outcomes and impacts coincide, which muddies the interpretation of our findings.

6.3. Results with before-after comparisons of labor market outcomes

In addition to inputs and outputs, survey respondents may employ simple before-after comparisons of outcomes as a (cognitively as well as conceptually) simple estimator. Of course, before-after comparisons, despite their simplicity, provide consistent estimates of program impacts under the condition that the "before" period outcome consistently estimates the outcome that would have been realized in the absence of treatment. Heckman and Smith (1999) show that this condition fails rather dramatically in the NJS data, with the result that, due to "Ashenfelter's dip" in earnings in the pre-program period, before-after impact estimates tend to have a strong upward bias. In this section, we relate several before-after impact estimates on employment and earnings to our self-reported evaluation measure.

Tables 9 and 10 present estimates of logit models with the self-reported evaluation measure as the dependent variable and three different measures of before-after earnings changes as independent variables, along with all of the variables in Column 1 of Table 5. The first measure, for which the estimates appear in Table 9, consists of the difference in average monthly earnings between the 12 months before random assignment and the 18 months after random assignment. We can use only the 12 months before random assignment due to the limitations of the survey data on pre-random assignment earnings for the treatment group.[8] The second and third measures, for which the estimates appear in Table 10, rely on the UI earnings data. The first measure, in the top panel of Table 10, consists of mean monthly earnings in the six calendar quarters after random assignment minus mean monthly earnings in the six calendar quarters after random assignment. The second measure, denoted UI(2) in the lower panel of Table 10, consists of the difference in mean monthly earnings between just the sixth quarter after and the sixth quarter before random assignment. In each case, we let the data speak to the functional form by including indicator variables for quintiles of the before-after difference.

We find strong evidence that before-after differences in labor market outcomes predict self-reported impacts. For the self-reported measure, the relationship is clearest for the adult females and the male youth, where the estimated coefficients increase

[8] In particular, we only have the response to a question about earnings in the previous year from the BIF.

monotonically (or almost so) and are statistically and substantively significant for the upper quintiles. Even stronger findings appear for the UI earnings difference measure in the top panel of Table 10, with large, and almost always statistically significant, coefficients for all four groups for the two upper quintiles. For male youth, the key difference seems to be between the lowest quintile and the other four; the four coefficients are all relatively large, all about equal and all statistically significant. For the other three groups, there is a general pattern of increasing coefficients as you move down the table. The results for the UI(2) measure in the bottom panel of Table 10 are weaker, in both a substantive and a statistical sense, than those in the top panel; this suggests that respondents use outcomes over the entire pre- and post-random assignment periods in constructing their implicit before-after estimates of program impact.

Given that the second of the two survey questions that compose our self-reported evaluation measure asks directly about finding a job, in Table 11 we consider its relationship to before-after employment changes. We coded an employment status difference variable based on employment at the date of random assignment and 18 months after random assignment. This yields four patterns. We include dummy variables for three of the four patterns, with employed at both points in time as the omitted pattern. The findings here are, perhaps, less strong than expected. In general, relative to the always employed, those who are never employed or who lose a job tend to have less positive self-reported evaluations. For the adults, those who gain a job tend to be somewhat more positive. However, only a handful of the differences achieve statistical significance. Measurement error in the "after" employment status may account for our weak results. By looking at employment around the time of the survey, we have given the respondents plenty of time to lose jobs that JTPA helped them find and, in the case of dropouts, to find jobs without the help of JTPA.

Finally, Table 12 presents the results of chi-squared tests for the joint significance of the before-after difference variables considered in Tables 9, 10 and 11. The test statistics and p-values in this table confirm that respondents' self-reported evaluations depend (in a statistical sense) on these before-after differences. Indeed, the joint tests for the employment change variable look stronger than the individual t-tests, suggesting that our omitted group lies in the middle of the categories in terms of its effect on the self-evaluation measure. Overall, the findings in this section lend support to the view that respondents implicitly or explicitly use natural and cognitively simple (but nonetheless quite biased) before-after comparisons in constructing their self-reported evaluations.

7.0 Results with performance measures

In this section we present results on the relationship between participant self-reported evaluations and performance measures based on program outcomes commonly used in employment and training programs both in the U.S. and elsewhere. Performance standards systems attempt to provide information on the impacts of programs quickly and at low cost by relying on crude proxies. In this sense, as the reader will quickly discern, some of the performance measures are fairly closely related to the outcome proxies examined in Section 6.2. In the JTPA program, the performance measures had real effects on site budgets; sites that did well on them received budgetary rewards while sites that did exceptionally poorly could receive "technical assistance" and also experienced the threat of formal reorganization. See, e.g., Heckman, Heinirch and Smith (2002), Heckman and Heinrich (2005) and Barnow and Smith (2004) for more detailed descriptions of the performance standards systems in JTPA, WIA and other programs, for evidence that the usual performance measures have little, if any, relationship to the actual impacts of the program, for evidence of strategic behavior by program staff in response to the incentives provided by the JTPA performance standards system, and for additional pointers to the literature.

The performance measures we examine here are subsets of those included in the JTPA and WIA performance standards systems. From the JTPA system we consider employment status at termination from the program, wages at termination from the program (which is defined only for those employed at termination), employment at "follow-up", which is 13 weeks after termination, and weekly earnings (not including zeros) at follow-up. We use self-reported information to construct the JTPA performance measures, as was done in that program. From the WIA system we consider employment at termination, employment at six months after termination conditional on employment at termination (this measure aims to count "retention", although it does not require the individual to stay at the same job), and the difference in quarterly earnings between the two calendar quarters after termination and the two quarters prior to random assignment. Note that the earnings gain measure, which was an innovation in the WIA system relative to JTPA, exploits the pre-program dip in mean earnings discussed in, e.g. Heckman and Smith (1999), to obtain a measure that will invariably suggest positive earnings impacts whether the program works or not. We follow the WIA program in relying mainly on the UI earnings records in constructing the WIA performance measures for our sample.

The top panel of Table 13 presents results based on estimating logit models with self-reported evaluations as the dependent variable and one of the performance measures, again including all of the variables in Table 5 as covariates. The bottom panel reports results from similar models using the WIA performance measures.

Three patterns emerge from the findings in Table 13. First, among the JTPA measures, employment at termination and employment at follow-up are significantly related, in both senses, to self-reported evaluations. The estimated mean change in the probability of a positive self-evaluation due to employment at termination ranges from 0.08 for male youth to 0.13 for adult females. Employment at follow-up shows a similarly strong relationship. Second, the WIA measures, other than the earnings change measure, show little in the way of a consistent relationship with the self-reported evaluation measure. In particular, using the UI earnings data to measure employment at termination rather than survey data on employment spells adds enough measurement error to yield a much weaker relationship for all four groups and one that is statistically significant only for adult males. Third, the relationship between self-reported evaluations and the performance measures appears stronger for women than for men.

Overall, Table 13 yields a mixed picture. Some performance measures based on labor market outcomes have substantively and statistically significant relationships with self-reported evaluations but even these account for only a modest fraction of the variance. Thus, the self-reported evaluations capture something related to, but very different from, the performance measures.

8.0 Conclusions

Broadly speaking, and putting aside the material in Section 7.0 regarding the performance standards, we have two main findings. The first is that self-reported evaluations by treatment group members from the JTPA experimental evaluation have, in general, little if any relationship to either experimental impact estimates at the subgroup level or to what we regard as relatively plausible econometric impact estimates based on percentile differences. The second is that the self-reported evaluation measures do have consistent relationships with crude proxies for impacts, such as measures of service type (a proxy for the amount of money spent), labor market outcome levels (which measure impacts only if the counterfactual state consists of no employment or earnings, which it does not for the vast majority of our sample), and before-after comparisons.

Taken together, these two findings provide strong support for the view that respondents avoid the cognitive burden associated with trying to construct (implicitly or explicitly) the counterfactual outcome they would have experienced had they been in the control group and thus excluded from JTPA. Instead, they appear to use readily available proxies and simple heuristics to conclude, for example, that if they are employed at the time of the survey or if their earnings have risen relative to the period prior to random assignment, that the program probably helped them find a job or get a better job. At the same time, our evidence does not rule out the view that respondents consider factors in their answers not captured in our experimental and econometric impact estimates, such as expected impacts in later periods. The proxy variables still leave much variation in the self-evaluation measure to be explained by other factors.

Overall, we conclude that participant self-evaluation measures of the type analyzed here represent a very poor substitute for rigorous experimental or non-experimental estimates of program impact. At the same time, to our knowledge this paper represents the first attempt to seriously study what these questions actually measure. The literature on using surveys to measure expectations, as discussed in Manski (2004), provides some reason for thinking that more sophisticated survey questions might do a better job of measuring the underlying objects of interest. Without additional research, we hesitate at this point to make any claims about the broadest of the questions that motivated this study, namely whether or not individuals can effectively construct the counterfactual outcomes required for them to make economically rational decisions.

References

Abbring, Japp, James Heckman and Edward Vytlacil. 2005. "Econometric Evaluation of Social Programs" in James Heckman and Edward Leamer (eds.) *Handbook of Econometrics, Volume 6*. Amsterdam: North-Holland: forthcoming.

Abadie, Alberto, Joshua Angrist and Guido Imbens. 2002. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings." *Econometrica*. 70(1): 91-117.

Angrist, Joshua. 2004. "Treatment Effect Heterogeneity in Theory and Practice." *Economic Journal* 114(494): C52-C83.

Barnow, Burt and Jeffrey Smith. 2004. "Performance Management of U.S. Job Training Programs: Lessons from the Job Training Partnership Act." *Public Finance and Management* 4(3): 247-287.

Bitler, Marianne, Jonah Gelbach and Hilary Hoynes. 2004. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." Unpublished manuscript, University of Maryland.

Bloom, Howard, Larry Orr, George Cave, Stephen Bell and Fred Doolittle. 1993. *The National JTPA Study: Title II-A Impacts on Earnings and Employment at 18 Months*. Bethesda, MD: Abt Associates.

Bloom, Howard, Larry Orr, Stephen Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Findings from the National Job Training Partnership Act Study." *Journal of Human Resources* 32: 549-576.

Carniero, Pedro, James Heckman and Edward Vytlacil. 2005. "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education." Unpublished manuscript, University of Chicago.

Devine, Theresa, and Heckman, James. 1996. "The Structure and Consequences of Eligibility Rules for a Social Program" in Solomon Polachek (ed.) *Research in Labor Economics Volume 15*. Greenwich, CT: JAI Press. 111-170.

Dominitz, Jeff and Charles Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources*. 31(1): 1-26.

Doolittle, Fred, and Linda Traeger. 1990. *Implementing the National JTPA Study*. New York, NY: Manpower Demonstration Research Corporation.

Dunnett, Charles and Ajit Tamhane. 1992. “A Step-Up Multiple Test Procedure.” *Journal of the American Statistical Association* 87(417): 162-170.

Greenberg, David. 1997. “The Leisure Bias in Cost-Benefit Analyses of Employment and Training Programs.” *Journal of Human Resources* 32(2): 413-439.

Greenberg, David and Daniel Shroder. 2004. *Digest of the Social Experiments, Third Edition*. Washington, DC: Urban Institute Press.

Heckman, James. 1997a. “Instrumental Variables: A Study of the Implicit Behavioral Assumptions Used in Making Program Evaluations.” *Journal of Human Resources* 32(3): 441-452.

Heckman, James. 1997b. “Randomization as an Instrumental Variable.” *Review of Economics and Statistics* 78(2): 336-341.

Heckman, James and Carolyn Heinrich, eds. 2005. *Performance Standards in a Government Bureaucracy*. W.E. Upjohn Institute for Employment Research.

Heckman, James, Carolyn Heinrich, and Jeffrey Smith. 2002. “The Performance of Performance Standards.” *Journal of Human Resources* 36(4): 778-811.

Heckman, James, Neil Hohmann, Jeffrey Smith, with the assistance of Michael Khoo. 2000. “Substitution and Drop Out Bias in Social Experiments: A Study of an Influential Social Experiment.” *Quarterly Journal of Economics* 115(2): 651-694.

Heckman, James, Robert LaLonde, and Jeffrey Smith. 1999. “The Economics and Econometrics of Active Labor Market Programs” in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics, Volume 3A*. Amsterdam: North-Holland, 1865-2097.

Heckman, James and Jeffrey Smith. 1998. “Evaluating the Welfare State” in Steiner Strom (ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*. Cambridge University Press for Econometric Society Monograph Series, 241-318.

Heckman, James, and Jeffrey Smith. 1999. “The Pre-Program Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies.” *Economic Journal* 109(457): 313-348.

Heckman, James, and Jeffrey Smith. 2000. “The Sensitivity of Experimental Impact Estimates: Evidence from the National JTPA Study” in David Blanchflower and Richard Freeman (eds.), *Youth Employment and Joblessness in Advanced Countries*, Chicago: University of Chicago Press for NBER, 331-356.

Heckman, James, and Jeffrey Smith. 2004. "The Determinants of Participation in a Social Program: Evidence from the Job Training Partnership Act." *Journal of Labor Economics* 22(4): 243-298.

Heckman, James, Jeffrey Smith, with the assistance of Nancy Clements. 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies* 64(4): 487-535.

Heckman, James and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects and Econometric Policy Evaluation. *Econometrica*, forthcoming.

Heinrich, Carolyn, Gerald Marschke and Annie Zhang. 1998. "Using Administrative Data to Estimate the Cost-Effectiveness of Social Program Services." Technical report. The University of Chicago.

Jacob, Brian and Lars Lofgren. 2005. "Principals as Agents: Subjective Performance Measurement in Education." NBER Working Paper No. 11463.

Kornfeld, Robert and Howard Bloom. 1999. "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports From Employers Agree With Surveys of Individuals?" *Journal of Labor Economics* 17(1): 168-197.

Kemple, James, Fred Doolittle, and John Wallace.1993. *The National JTPA Study: Site Characteristics and Participation Patterns*. New York, NY: Manpower Demonstration Research Corporation.

Liu, Wei. 1997. "Stepwise Tests When the Test Statistics Are Independent." *Australian Journal of Statistics* 39(2): 169-177.

Manski, Charles. 1990. "The Use of Intentions Data to Predict Behavior: A Best-Case Analysis." *Journal of the American Statistical Association* 85(412): 934-940.

Manski, Charles. 2004. "Measuring Expectations." *Econometrica* 72(5): 1329-1376.

Orr, Larry, Howard Bloom, Stephen Bell, Winston Lin, George Cave and Fred Doolittle. 1994. *The National JTPA Study: Impacts, Benefits and Costs of Title II-A*. Bethesda, MD: Abt Associates.

Orr, Larry, Howard Bloom, Stephen Bell, Fred Doolittle, Winston Lin, and George Cave. 1996. *Does Training Work for the Disadvantaged? Evidence from the National JTPA Study*. Washington, DC: Urban Institute Press.

Philipson, Tomas and Larry Hedges. 1998. "Subject Evaluation in Social Experiments." *Econometrica* 66(2): 381-408.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3: 135-146.

Smith, Jeffrey. 1997a. "Measuring Earnings Dynamics Among the Poor: Evidence from Two Samples of JTPA Eligibles." Unpublished manuscript, University of Western Ontario.

Smith, Jeffrey. 1997b. "Measuring Earnings Levels Among the Poor: Evidence from Two Samples of JTPA Eligibles." Unpublished manuscript, University of Western Ontario.

Smith, Jeffrey and Alexander Whalley. 2005. "How Well Do We Measure Public Job Training?" Unpublished manuscript, University of Maryland.

Wood, Michelle. 2005. "National JTPA Study – SDA Unit Costs." Abt Associates Memo to Jerry Marsky [sic] and Larry Orr.

Appendix 1: Data Appendix

1. Sample Selection Criteria for the Samples Used

Our data set combines self-reported information from the Background Information Form, completed at or near the time of random assignment and the First Follow-Up Survey, collected around 18 months after random assignment with administrative data on quarterly earnings from matched UI wage records.

The full experimental sample contains 6639 observations in the control group and 13972 observations in the treatment group. If we restrict our sample to only those with valid self-reported earnings for the 18 months after random assignment we lose 2080 observations from the control group and 4329 observations from the treatment group. If we instead restrict the sample to only those with valid UI earnings over the six quarters after random assignment we lose 122 observations from the control group and 232 observations from the treatment group. We only require sample members to have valid values for earnings for the analysis in question; that is, we use all available observations for a given dependent variable. The analyses presented in Tables 5 to 13 require only the data from the experimental treatment group.

Our self-reported earnings data consists of the self-reported data used in Bloom et al. (1993), the official 18-month impact report. The data we use include the recoded values for outliers (which were examined individually and by hand by staff of Abt Associates) but do not include the imputed values based on the matched UI earnings records that they employed in some of their analyses. This earnings variable is not available on the public use CD but is available from the authors by request.

The matched administrative data from UI records consists of earnings in each calendar quarter. As a result, for some sample members, the six calendar quarters after the calendar quarter of random assignment (the period used in some of our dependent variables from the UI data) will cover a slightly different set of months than the 18 months after the month after random assignment (the period covered in some of our dependent variables from the self-reported data).

We do not drop observations with missing values of covariates from the sample for any of our analyses; instead we include dummy variables for those with missing values of the covariates used in each analysis. If we had instead listwise deleted observations from the sample having any missing value for the covariates we would lose 18327 observations out of the 20601 observations in the full experimental sample.

2. Variable Definitions

Predicted impact: This consists of the experimentally estimated predicted impact of the program for an individual based on either the individual's measured characteristics or the individual's quantile in the untreated outcome distribution.

Percent positive self-evaluation: This is the mean of a binary indicator for a positive participant self-evaluation. It is defined only for individuals in the treatment group.

Earnings one: This is total earnings over the 18 months after random assignment based on the self-reported earnings data.

Employment one: This is a binary variable indicating any employment over the 18 months after random assignment using self-reported earnings data. The variable equals one if self-reported earnings over the 18 months after random assignment are positive and zero otherwise.

Earnings two: This is total earnings in the 18th month after random assignment based on the self-reported earnings data.

Employment two: This is a binary variable indicating employment in month 18 after random assignment based on the self-reported earnings data. The variable equals one if self-reported earnings in the 18th month after random assignment are positive and zero otherwise.

Earnings three: This is total earnings in the six calendar quarters after the calendar quarter of random assignment based on the matched UI administrative earnings data.

Employment three: This is a binary variable indicating any employment over the six calendar quarters after the calendar quarter of random assignment based on the matched UI administrative earnings data. This variable equals one if UI earnings over the six calendar quarters after the calendar quarter of random assignment are positive and zero otherwise.

Earnings four: This is total earnings in month 18 after random assignment based on the matched UI administrative earnings data.

Employment four: This is a binary variable indicating any employment in the sixth calendar quarter after the calendar quarter of random assignment based on the matched UI administrative earnings data. This variable equals one if UI earnings in the sixth calendar quarter after random assignment are positive and zero otherwise.

EXHIBIT 1: JTPA Self-Evaluation Survey Questions

(D7)

According to (LOCAL JTPA PROGAM NAME) records, you applied to enter (LOCAL JTPA PROGRAM NAME) in (MONTH/YEAR OF RANDOM ASSIGNMENT). Did you participate in the program after you applied?

YES (SKIP TO D9)
NO (GO TO D8)

(D9)

Do you think that the training or other assistance that you got from the program helped you get a job or perform better on the job?

YES
NO

Source: JTPA First Follow-Up Study Survey Instrument

TABLE 1: Bivariate Results for the relationship between Experimental Impacts and Positive Self-Evaluation By Demographic Group

	Percentage Positive Self-Evaluation	Earnings One	Employ One	Earnings Two	Employ Two	Earnings Three	Employ Three	Earnings Four	Employ Four
Adult Males	0.63	538.20	0.03	23.58	0.02	-36.42	0.00	-24.10	-0.03
	(0.01)	(379.22)	(0.01)	(28.55)	(0.02)	(293.50)	(0.01)	(65.69)	(0.02)
Adult Females	0.65	750.87	0.03	56.79	0.04	594.08	0.04	131.24	0.03
	(0.01)	(236.17)	(0.01)	(18.34)	(0.14)	(195.48)	(0.01)	(44.18)	(0.01)
Male Youth	0.67	-777.33	0.01	-82.93	-0.03	-381.03	-0.02	-128.07	-0.03
	(0.02)	(463.33)	(0.01)	(37.00)	(0.02)	(328.19)	(0.02)	(73.54)	(0.02)
Female Youth	0.72	-44.89	0.04	8.38	-0.00	-233.74	0.01	-13.84	0.00
	(0.01)	(295.12)	(0.02)	(29.87)	(0.02)	(227.97)	(0.02)	(50.74)	(0.02)
Correlation with Positive Self-Evaluation	--	-0.4620	0.5510	-0.2239	-0.4553	-0.4381	-0.1486	-0.1858	0.1426
		[0.538]	[0.449]	[0.776]	[0.545]	[0.562]	[0.851]	[0.814]	[0.857]

Notes: Source: Authors' calculations using the NJS data. Values in the table are means for Positive Self-Evaluation, and experimental impacts for the eight outcomes. The values in parentheses are standard errors and the values in square brackets are p-values. Percentage Positive Self-Assessment is calculated as the mean of the binary indicator positive self-assessment variable for those who self-report participating and are in the treatment group. Earnings one and employment one are earnings and any employment over the 18-months after random assignment using self-reported earnings data. Earnings two and employment two are earnings and employment in month 18 after random assignment using self-reported earnings data. Earnings three and employment three are earnings and any employment over the 18-months after random assignment using UI-reported earnings data. Earnings four and employment four are earnings and employment in month 18 after random assignment using UI-reported earnings data. Those with missing outcomes are dropped from the estimate for that outcome only.

TABLE 2A: Bivariate results for the Correlation between Experimental Impacts and Self-Evaluation for Eight Outcomes, Adult Males

	Earnings One	Employment One	Earnings Two	Employment Two	Earnings Three	Employment Three	Earnings Four	Employment Four
Race	0.1742 [0.826]	-0.3556 [0.644]	-0.2184 [0.782]	-0.0346 [0.965]	0.7508 [0.249]	0.6956 [0.304]	0.7023 [0.298]	0.6447 [0.355]
Age Category	0.9974 [0.046]	-0.9989 [0.030]	0.9870 [0.103]	-0.9137 [0.266]	0.8198 [0.388]	-0.0455 [0.971]	0.7573 [0.453]	0.0169 [0.989]
Education Category	0.4984 [0.393]	-0.7003 [0.188]	0.7613 [0.135]	-0.3996 [0.505]	0.9077 [0.033]	0.1717 [0.783]	0.9798 [0.003]	-0.8810 [0.048]
Marital Status	-0.5476 [0.631]	-0.9999 [0.007]	-0.9946 [0.066]	-0.9404 [0.229]	0.9939 [0.070]	-0.2063 [0.868]	0.7701 [0.440]	0.6103 [0.582]
Employ Category	-0.1606 [0.897]	-0.8638 [0.336]	-0.1177 [0.925]	-0.5880 [0.600]	-0.6909 [0.514]	0.3855 [0.748]	-0.3794 [0.752]	-0.9451 [0.212]
Site	0.3380 [0.200]	0.1495 [0.581]	0.1682 [0.533]	0.4170 [0.108]	-0.2132 [0.428]	-0.1497 [0.580]	-0.1015 [0.709]	0.0265 [0.923]

Positive Correlations
Overall: 24 of 48 (50 %); significant at 0.10: 4 of 48 (8 %); significant at 0.05: 3 of 48 (6 %)
Negative Correlations
Overall: 24 of 48 (50 %); significant at 0.10: 4 of 48 (8 %); significant at 0.05: 3 of 48 (6 %)

Notes: Source: Authors' calculations using the NJS data. Values in the table are the correlation between the mean of Positive Self-Evaluation, and the experimental impacts by subgroup. The values in square brackets are p-values. Percentage Positive Self-Evaluation is calculated as the mean of the binary indicator positive self-evaluation variable for those who self-report participating and are in the treatment group. Earnings one and employment one are earnings and any employment over the 18-months after random assignment using self-reported earnings data. Earnings two and employment two are earnings and employment in month 18 after random assignment using self-reported earnings data. Earnings three and employment three are earnings and any employment over the 18-months after random assignment using UI-reported earnings data. Earnings four and employment four are earnings and employment in month 18 after random assignment using UI-reported earnings data. Those with missing outcomes are dropped from the estimate for that outcome only. The categories are defined as the following. Race: White, Black, Hispanic and Other. Age: less than 19 years, 19-21 years, 22-25 years, 26-34 years and 35+ years. Education: under 10 years, 10-11 years, 12 years, 13-15 years and 16+ years. Marital Status: single, married, and divorced/widowed/separated. Employment Status: out of labor force, unemployed, and employed. Site: sixteen site categories.

TABLE 2B: Bivariate results for the Correlation between Experimental Impacts and Self-Evaluation for Eight Outcomes, Adult Females

	Earnings One	Employment One	Earnings Two	Employment Two	Earnings Three	Employment Three	Earnings Four	Employment Four
Race	0.3681 [0.632]	0.4127 [0.587]	-0.2490 [0.751]	-0.2490 [0.751]	0.3618 [0.638]	0.2877 [0.712]	0.4188 [0.581]	0.6203 [0.380]
Age Category	-0.8947 [0.295]	0.1282 [0.918]	-0.9136 [0.267]	-0.9681 [0.161]	0.5726 [0.612]	-0.4926 [0.672]	-0.9490 [0.204]	-0.9597 [0.181]
Education Category	-0.6549 [0.230]	-0.6872 [0.200]	-0.5260 [0.363]	0.1167 [0.852]	-0.1673 [0.788]	0.6492 [0.236]	-0.6820 [0.205]	-0.2442 [0.692]
Marital Status	0.0802 [0.949]	0.8084 [0.401]	0.4329 [0.715]	0.5232 [0.650]	0.4284 [0.718]	0.3944 [0.742]	-0.7971 [0.413]	0.5074 [0.661]
Employ Category	0.9296 [0.240]	0.7264 [0.482]	0.6294 [0.567]	0.2549 [0.836]	0.9978 [0.042]	0.8663 [0.333]	0.2949 [0.809]	0.8199 [0.388]
Site	-0.0745 [0.784]	-0.2296 [0.392]	-0.0628 [0.817]	0.1812 [0.502]	-0.0753 [0.782]	0.1143 [0.674]	-0.0250 [0.927]	0.1923 [0.476]

Positive Correlations
Overall: 28 of 48 (58 %); significant at 0.10: 0 of 48 (0 %); significant at 0.05: 0 of 48 (0 %)
Negative Correlations
Overall: 20 of 48 (42 %); significant at 0.10: 0 of 48 (0 %); significant at 0.05: 0 of 48 (0 %)

Notes: Source: Authors' calculations using the NJS data. Values in the table are the correlation between the mean of Positive Self-Evaluation, and the experimental impacts by subgroup. The values in square brackets are p-values. Percentage Positive Self-Evaluation is calculated as the mean of the binary indicator positive self-evaluation variable for those who self-report participating and are in the treatment group. Earnings one and employment one are earnings and any employment over the 18-months after random assignment using self-reported earnings data. Earnings two and employment two are earnings and employment in month 18 after random assignment using self-reported earnings data. Earnings three and employment three are earnings and any employment over the 18-months after random assignment using UI-reported earnings data. Earnings four and employment four are earnings and employment in month 18 after random assignment using UI-reported earnings data. Those with missing outcomes are dropped from the estimate for that outcome only. The categories are defined as the following. Race: White, Black, Hispanic and Other. Age: less than 19 years, 19-21 years, 22-25 years, 26-34 years and 35+ years. Education: under 10 years, 10-11 years, 12 years, 13-15 years and 16+ years. Marital Status: single, married, and divorced/widowed/separated. Employment Status: out of labor force, unemployed, and employed. Site: sixteen site categories.

TABLE 2C: Bivariate results for the Correlation between Experimental Impacts and Self-Evaluation for Eight Outcomes, Male Youths

	Earnings One	Employment One	Earnings Two	Employment Two	Earnings Three	Employment Three	Earnings Four	Employment Four
Race	0.471	0.2314	-0.0844	-0.3902	0.2621	0.1283	0.2866	0.2097
	[0.528]	[0.769]	[0.916]	[0.610]	[0.738]	[0.872]	[0.713]	[0.790]
Education Category	-0.4412	0.2749	-0.1031	0.8667	0.4519	-0.8642	0.6511	0.9301
	[0.559]	[0.725]	[0.897]	[0.133]	[0.548]	[0.136]	[0.349]	[0.070]
Marital Status	-0.8985	-0.8039	-0.9816	-0.2344	0.9262	-0.9926	0.9954	0.9671
	[0.289]	[0.406]	[0.122]	[0.849]	[0.246]	[0.077]	[0.061]	[0.164]
Employ Category	0.8770	0.9606	0.7889	0.8352	0.9985	0.8704	0.9865	0.7949
	[0.319]	[0.179]	[0.421]	[0.371]	[0.035]	[0.328]	[0.105]	[0.415]
Site	0.3258	-0.1278	0.1120	0.6023	0.2063	0.1062	-0.1386	-0.4438
	[0.236]	[0.637]	[0.680]	[0.014]	[0.443]	[0.696]	[0.609]	[0.085]

Positive Correlations
Overall: 27 of 40 (68 %); significant at 0.10: 4 of 40 (10 %); significant at 0.05: 2 of 40 (5 %)
Negative Correlations
Overall: 13 of 40 (32 %); significant at 0.10: 2 of 40 (5 %); significant at 0.05: 0 of 40 (0 %)

Notes: Source: Authors' calculations using the NJS data. Values in the table are the correlation between the mean of Positive Self-Evaluation, and the experimental impacts by subgroup. The values in square brackets are p-values. Percentage Positive Self-Evaluation is calculated as the mean of the binary indicator positive self-evaluation variable for those who self-report participating and are in the treatment group. Earnings one and employment one are earnings and any employment over the 18-months after random assignment using self-reported earnings data. Earnings two and employment two are earnings and employment in month 18 after random assignment using self-reported earnings data. Earnings three and employment three are earnings and any employment over the 18-months after random assignment using UI-reported earnings data. Earnings four and employment four are earnings and employment in month 18 after random assignment using UI-reported earnings data. Those with missing outcomes are dropped from the estimate for that outcome only. The categories are defined as the following. Race: White, Black, Hispanic and Other. Education: under 10 years, 10-11 years, 12 years, 13-15 years and 16+ years. Marital Status: single, married, and divorced/widowed/separated. Employment Status: out of labor force, unemployed, and employed. Site: sixteen site categories.

TABLE 2D: Bivariate results for the Correlation between Experimental Impacts and Self-Evaluation for Eight Outcomes, Female Youths

	Earnings One	Employment One	Earnings Two	Employment Two	Earnings Three	Employment Three	Earnings Four	Employment Four
Race	-0.7560 [0.244]	-0.2111 [0.789]	-0.7309 [0.269]	-0.5835 [0.417]	-0.5849 [0.415]	0.4595 [0.541]	-0.6132 [0.387]	0.7253 [0.274]
Education Category	-0.9988 [0.000]	-0.9995 [0.000]	-0.9914 [0.001]	-0.9896 [0.001]	-0.9459 [0.015]	-0.9950 [0.000]	-0.9616 [0.009]	-0.9702 [0.006]
Marital Status	-0.4687 [0.690]	-0.1860 [0.881]	0.6774 [0.526]	0.3977 [0.740]	0.8910 [0.300]	-0.1495 [0.905]	0.9998 [0.012]	0.3252 [0.789]
Employ Category	0.8395 [0.366]	-0.5065 [0.662]	-0.8400 [0.365]	-0.9703 [0.156]	0.2660 [0.829]	0.9998 [0.012]	-0.3951 [0.741]	0.1339 [0.915]
Site	0.2620 [0.346]	0.3277 [0.233]	0.2164 [0.439]	0.3086 [0.263]	0.2643 [0.341]	0.1690 [0.547]	0.2554 [0.358]	0.2112 [0.450]

Positive Correlations
Overall: 19 of 40 (47 %); significant at 0.10: 2 of 40 (5 %); significant at 0.05: 2 of 40 (5 %)
Negative Correlations
Overall: 21 of 40 (53 %); significant at 0.10: 8 of 40 (20 %); significant at 0.05: 8 of 40 (20 %)

Notes: Source: Authors' calculations using the NJS data. Values in the table are the correlation between the mean of Positive Self-Evaluation, and the experimental impacts by subgroup. The values in square brackets are p-values. Percentage Positive Self-Evaluation is calculated as the mean of the binary indicator positive self-evaluation variable for those who self-report participating and are in the treatment group. Earnings one and employment one are earnings and any employment over the 18-months after random assignment using self-reported earnings data. Earnings two and employment two are earnings and employment in month 18 after random assignment using self-reported earnings data. Earnings three and employment three are earnings and any employment over the 18-months after random assignment using UI-reported earnings data. Earnings four and employment four are earnings and employment in month 18 after random assignment using UI-reported earnings data. Those with missing outcomes are dropped from the estimate for that outcome only. The categories are defined as the following. Race: White, Black, Hispanic and Other. Education: under 10 years, 10-11 years, 12 years, 13-15 years and 16+ years. Marital Status: single, married, and divorced/widowed/separated. Employment Status: out of labor force, unemployed, and employed. Site: sixteen site categories.

TABLE 3: Regression results for the relationship between Predicted Impacts and Positive Self-Evaluation for Eight Outcomes, By Demographic Group

	Adult Males		Adult Females		Male Youths		Female Youths	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Earnings over 18 Months	-121.04	48.66	45.71	-16.86	-21.95	273.06	-208.51	24.82
	(134.85)	(83.41)	(85.10)	(57.75)	(244.01)	(214.97)	(89.36)	(97.87)
Any Employment During	-0.009	-0.002	0.001	0.005	-0.003	0.010	-0.005	-0.003
18 Months	(0.003)	(0.04)	(0.023)	(0.004)	(0.003)	(0.006)	(0.002)	(0.008)
Earnings in Month 18	-21.20	-6.05	2.37	0.97	35.53	-6.67	-2.32	1.54
	(20.65)	(7.73)	(4.95)	(3.80)	(31.26)	(16.58)	(9.49)	(10.37)
Employment in Month 18	-0.005	-0.002	-0.004	0.002	-0.003	-0.001	-0.002	-0.003
	(0.003)	(0.004)	(0.004)	(0.003)	(0.014)	(0.011)	(0.001)	(0.011)
Earnings (UI) over 6 Quarters	-63.67	-61.17	-85.43	14.51	-71.24	271.47	-103.53	78.86
	(94.48)	(85.90)	(50.92)	(35.79)	(133.03)	(134.34)	(68.76)	(68.61)
Any Employment (UI)	-0.003	-0.002	0.001	0.002	-0.007	0.000	0.003	0.007
During 6 Quarters	(0.002)	(0.003)	(0.003)	(0.003)	(0.007)	(0.006)	(0.013)	(0.006)
Earnings (UI) in Quarter 6	-22.56	-14.52	0.73	-10.17	-0.16	80.59	2.60	-13.14
	(23.25)	(17.18)	(10.96)	(7.65)	(18.90)	(31.78)	(13.53)	(12.60)
Employment (UI) in Quarter 6	-0.004	0.000	0.000	-0.004	0.008	0.019	-0.001	0.007
	(0.003)	(0.004)	(0.003)	(0.003)	(0.005)	(0.010)	(0.002)	(0.009)
Positive (overall / 0.10 / 0.05)	0/0/0	1/0/0	5/0/0	5/0/0	2/0/0	5/3/2	2/0/0	5/0/0
Negative (overall / 0.10 / 0.05)	8/1/1	6/0/0	2/1/0	3/0/0	6/0/0	2/0/0	5/2/2	3/0/0

Notes: Source: Authors' calculations using the NJS data. Each cell in the table is a coefficient estimate from the regression of the estimated impacts for an individual (based on their X) as the dependent variable and self-evaluation as the independent variable. The population used is the treatment sample. The values in parentheses are the heteroskedastic-consistent standard errors. The values in the bottom two rows are the counts of the number of cells in the column above which are positive or negative, and counts of those that are significantly different from zero at the 10% and 5% levels. Specification (1) selects the set of X's used to predict the impacts for each individual by a stepwise procedure. Specification (2) uses the specification of X' s used in Heckman, Heinrich and Smith (2003) (HHS) to predict the impacts for each individual. The HHS set of X's contains: are race, age, education, marital status, employment status, AFDC receipt, receipt of food stamps and site.

TABLE 4: Relationship between Quantile Treatment Effects for 18-Month Earnings and the Percent with Positive Self-Evaluation, By Demographic Group

	Adult Males		Adult Females		Male Youths		Female Youths	
	Quantile Treatment Effects	Percentage Positive Self-Evaluation	Quantile Treatment Effects	Percentage Positive Self-Evaluation	Quantile Treatment Effects	Percentage Positive Self-Evaluation	Quantile Treatment Effects	Percentage Positive Self-Evaluation
5th	0	0.51	0	0.57	0	0.56	0	0.68
	(0.90)	(0.03)	(0.38)	(0.02)	(1.15)	(0.07)	(1.07)	(0.04)
25th	1233	0.66	501	0.63	-516	0.62	402	0.72
	(452)	(0.05)	(193)	(0.04)	(515)	(0.08)	(193)	(0.06)
50th	825	0.56	747	0.63	-1161	0.83	-39	0.71
	(608)	(0.05)	(416)	(0.04)	(681)	(0.06)	(371)	(0.07)
75th	8	0.72	938	0.78	-1261	0.68	-479	0.83
	(590)	(0.05)	(383)	(0.04)	(701)	(0.08)	(566)	(0.06)
95th	1589	0.65	1910	0.70	-887	0.81	-53	0.64
	(1323)	(0.05)	(740)	(0.04)	(1959)	(0.07)	(1012)	(0.07)
Correlation with Percentage Positive Self-Evaluation	0.0760	--	0.7652	--	-0.4527	--	-0.4209	--
	[0.750]		[0.000]		[0.045]		[0.065]	
Coefficient on Percentage Positive Self-Evaluation	511	--	5489	--	-2232	--	-1576	--
	(1686)		(1204)		(909)		(931)	

Notes: Source: Authors' calculations using the NJS data. The values in the left column of the upper panel for each demographic group are quantile treatment effects estimates with standard errors in parentheses for five quantiles. The values in the right column of the upper panel for each demographic group are the means of the binary positive self-evaluation indicator variable for each quantile of the outcome distribution for those in the treatment group. The first row of the lower panel contains the correlation between the treatment effect estimates and the percentage positive self-evaluation by quantile (where one observation is one of the 20 quantiles) and the p-value for the correlation is in square brackets. The second row of the lower panel contains the coefficient of the regression with percentage positive self-evaluation as the independent variable and the treatment effect estimate as the dependant variable (where one observation is one of the 20 quantiles). The hetero-skedastic consistent standard errors for these estimates appear in parentheses.

TABLE 5: Logit Estimates of the Determinants of Positive Self-Evaluation, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
Age: 19-21 Years	--	--	0.026 (0.039) [0.505]	-0.036 (0.031) [0.255]
Age: 26-34 Years	-0.020 (0.029) [0.500]	-0.023 (0.026) [0.383]	--	--
Age: 35+ years	-0.071 (0.034) [0.036]	-0.122 (0.032) [0.000]	--	--
Marital Status: Married	-0.035 (0.032) [0.283]	0.039 (0.030) [0.188]	-0.009 (0.065) [0.889]	-0.013 (0.048) [0.794]
Marital Status: Divorced/Widowed/ Separated	-0.032 (0.032) [0.327]	-0.009 (0.025) [0.706]	0.204 (0.119) [0.086]	-0.145 (0.053) [0.006]
Education: 10-11 Years	0.049 (0.035) [0.155]	-0.022 (0.032) [0.487]	0.040 (0.044) [0.364]	0.083 (0.035) [0.018]
Education: 12 Years	0.048 (0.032) [0.137]	-0.051 (0.029) [0.081]	-0.003 (0.049) [0.944]	0.082 (0.038) [0.030]
Education: 13-15 Years	-0.005 (0.041) [0.900]	-0.035 (0.037) [0.337]	-0.075 (0.090) [0.406]	0.102 (0.059) [0.066]
Education: 16+ Years	0.022 (0.057) [0.705]	0.065 (0.058) [0.266]	--	--
Race: Black	0.020 (0.034) [0.543]	-0.016 (0.030) [0.598]	0.122 (0.045) [0.007]	-0.035 (0.043) [0.425]
Race: Hispanic	0.048 (0.047) [0.313]	-0.001 (0.042) [0.979]	0.159 (0.052) [0.002]	0.103 (0.048) [0.031]
Race: Other	0.035 (0.075) [0.646]	-0.054 (0.072) [0.458]	0.039 (0.130) [0.764]	0.017 (0.106) [0.872]
English Language	0.051	0.106	0.135	-0.058

	(0.070)	(0.059)	(0.107)	(0.158)
	[0.471]	[0.073]	[0.210]	[0.715]
AFDC Receipt	0.039	-0.015	0.085	-0.007
	(0.042)	(0.024)	(0.052)	(0.037)
	[0.352]	[0.521]	[0.097]	[0.850]
Work for Pay	0.037	0.039	-0.059	0.015
	(0.045)	(0.028)	(0.053)	(0.036)
	[0.415]	[0.164]	[0.264]	[0.697]
Child less than Six	-0.019	0.003	-0.146	0.003
	(0.033)	(0.023)	(0.073)	(0.035)
	[0.571]	[0.900]	[0.044]	[0.940]
Self-Report	0.135	0.122	0.084	0.010
Training: CT-OS	(0.028)	(0.023)	(0.040)	(0.034)
	[0.000]	[0.000]	[0.038]	[0.764]
Self-Report	0.153	0.115	0.109	0.081
Training: OJT/WE	(0.040)	(0.034)	(0.062)	(0.055)
	[0.000]	[0.001]	[0.076]	[0.138]
Self-Report	0.085	0.022	0.000	0.029
Training: JSA	(0.042)	(0.038)	(0.088)	(0.066)
	[0.044]	[0.570]	[0.998]	[0.661]
Self-Report	0.046	0.019	0.075	0.095
Training: ABE	(0.049)	(0.038)	(0.049)	(0.040)
	[0.347]	[0.614]	[0.129]	[0.018]
Self-Report	0.153	0.048	0.079	0.090
Training: Other	(0.053)	(0.048)	(0.082)	(0.061)
	[0.004]	[0.313]	[0.334]	[0.140]
Administrative-	0.039	-0.005	-0.023	0.154
Report Training:	(0.058)	(0.046)	(0.104)	(0.052)
CT-OS	[0.493]	[0.916]	[0.824]	[0.003]
Administrative-	0.055	0.088	-0.084	0.067
Report Training:	(0.058)	(0.048)	(0.113)	(0.067)
OJT/WE	[0.340]	[0.066]	[0.459]	[0.313]
Administrative-	0.036	-0.045	0.077	0.062
Report Training:	(0.058)	(0.052)	(0.099)	(0.063)
JSA	[0.532]	[0.387]	[0.433]	[0.326]
Administrative-	0.108	-0.122	-0.184	-0.064
Report Training:	(0.066)	(0.068)	(0.108)	(0.078)
ABE	[0.102]	[0.073]	[0.089]	[0.409]
Administrative-	0.050	-0.121	-0.148	0.123
Report Training:	(0.063)	(0.057)	(0.110)	(0.060)
Other	[0.425]	[0.033]	[0.178]	[0.040]

Notes: Source: Authors' calculations using the NJS data. Columns two through five of the table report the results from a logit model where the binary positive self-evaluation variable is the dependant variable and the categorical variables listed in column one are the independent variables. The values in the table are mean numerical derivatives, with the standard errors in parentheses and p-values in square brackets. The population for these regressions is the treatment sample. Indicator variables for missing values for the independent variables are also included in the regression. The omitted age category for adults is age 22-25 years and is age less than 19 for youths. The omitted marital status is single, the omitted education category is less than 10 years, the omitted racial group is white, and the omitted training type for both self-report and administrative report is no training for all demographic groups.

TABLE 6: Test Statistics from Logit Models of the Determinants of Positive Self-Evaluation, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
Site	65.30	73.06	29.66	61.12
	[0.000]	[0.000]	[0.009]	[0.000]
Age Category	5.09	21.71	0.45	1.26
	[0.078]	[0.000]	[0.504]	[0.262]
Marital Status	1.44	3.68	1.44	8.55
	[0.487]	[0.159]	[0.488]	[0.014]
Education Category	4.50	6.47	2.46	6.37
	[0.343]	[0.167]	[0.482]	[0.095]
Race	1.20	0.80	10.49	6.36
	[0.753]	[0.849]	[0.015]	[0.095]
English Language	0.91	3.16	1.12	0.15
	[0.633]	[0.206]	[0.290]	[0.703]
Other Individual Characteristics	6.13	3.16	6.72	1.23
	[0.294]	[0.675]	[0.242]	[0.942]
Self-Reported Training Type	30.21	30.66	6.23	7.05
	[0.000]	[0.000]	[0.284]	[0.217]
Administrative Reported Training Type	30.67	53.40	21.55	25.61
	[0.000]	[0.000]	[0.002]	[0.000]

Notes: Source: Authors' calculations using the NJS data. Columns two through five of the table report the results from a logit model where the binary positive self-evaluation variable is the dependent variable and the categorical variables summarized in column one are the independent variables. The values in the table are χ^2-statistics for joint tests that all of the coefficients equal zero for a given group of variables, with the p-values in square brackets. The population for these regressions is the treatment sample. The variables in 'Other Individual Characteristics' are AFDC receipt, child less than six indicator, and worked for pay indicator. Indicator variables for missing values for the independent variables are also included in the regressions.

TABLE 7: Logit Estimates of the Relationship between Outcomes and Positive Self-Evaluation: Four Outcomes, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
Earnings over 18 Months = 0	-0.171	-0.184	-0.151	-0.065
	(0.042)	(0.038)	(0.084)	(0.064)
	[0.000]	[0.000]	[0.073]	[0.312]
Earnings over 18 Months Bottom Quartile	-0.096	-0.094	-0.158	-0.028
	(0.040)	(0.035)	(0.062)	(0.053)
	[0.015]	[0.007]	[0.011]	[0.593]
Earnings over 18 Months Lower Middle Quartile	-0.038	-0.133	-0.001	0.068
	(0.035)	(0.034)	(0.056)	(0.049)
	[0.269]	[0.000]	[0.992]	[0.172]
Earnings over 18 Months Upper Middle Quartile	-0.083	-0.033	-0.052	0.053
	(0.033)	(0.034)	(0.056)	(0.053)
	[0.013]	[0.337]	[0.358]	[0.319]
Earnings over 18 Months = 0 (UI)	-0.076	-0.027	-0.063	-0.068
	(0.041)	(0.034)	(0.073)	(0.064)
	[0.062]	[0.423]	[0.387]	[0.290]
Earnings over 18 Months Bottom Quartile (UI)	-0.080	-0.112	-0.120	-0.081
	(0.039)	(0.034)	(0.061)	(0.057)
	[0.038]	[0.001]	[0.049]	[0.153]
Earnings over 18 Months Lower Middle Quartile (UI)	-0.068	-0.067	-0.050	-0.089
	(0.035)	(0.032)	(0.056)	(0.057)
	[0.049]	[0.035]	[0.368]	[0.113]
Earnings over 18 Months Upper Middle Quartile (UI)	-0.026	-0.035	-0.030	0.014
	(0.032)	(0.030)	(0.059)	(0.054)
	[0.420]	[0.249]	[0.605]	[0.795]
Any Employment Over 18 Months	0.122	0.105	0.084	0.082
	(0.038)	(0.027)	(0.071)	(0.044)
	[0.001]	[0.000]	[0.237]	[0.060]
Any Employment over 18 Months (UI)	0.038	-0.029	0.000	0.006
	(0.035)	(0.025)	(0.054)	(0.039)
	[0.289]	[0.253]	[1.000]	[0.886]

Notes: Source: Authors' calculations using the NJS data. Columns two through five of this table report the results from logit regressions where the binary positive self-evaluation variable is the dependant variable and the categorical variables listed in column one of Table 5 are the independent variables, in addition an outcome variable is included in each regression. The values in the table are mean numerical derivatives, with the standard errors in parentheses and p-values in square brackets. For earnings outcomes the continuous variables are entered as four categorical variables: zero earnings, an indicator for being in the lowest quartile of the non-zero earnings distribution, lower middle quartile of the non-zero earnings distribution, upper middle quartile of the non-zero earnings distribution. The omitted category is for those with earnings in the highest quartile of the non-zero earnings distribution. For the employment outcomes a binary variable is included indicating whether the respondent was employed or not. Each set of cells in the table is the result for a different specification where the outcome to be included as an independent variable is different. The sets of cells are defined as two groups of four and two groups of two depending on how the outcome enters the regression. The population for these regressions is the treatment sample. Indicator variables for missing values for the independent variables are also included in the regression.

TABLE 8: Test Statistics from Logit Models of the Relationship between Outcomes and Positive Self-Evaluation, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
Earnings over 18 Months	20.37	34.33	15.96	12.18
	[0.001]	[0.000]	[0.007]	[0.032]
Any Employment during	11.38	15.60	5.58	4.51
18 Months	[0.003]	[0.000]	[0.062]	[0.105]
Earnings in Month 18	15.75	30.55	8.02	1.32
	[0.008]	[0.000]	[0.155]	[0.933]
Employment in Month 18	6.69	12.36	5.67	0.99
	[0.035]	[0.002]	[0.059]	[0.609]
Earnings over 6 Quarters (UI)	7.37	13.92	4.98	10.06
	[0.195]	[0.016]	[0.418]	[0.074]
Any Employment During 6 Quarters (UI)	1.17	1.27	0.39	2.41
	[0.556]	[0.529]	[0.825]	[0.300]
Earnings in Quarter 6 (UI)	12.23	12.49	3.53	7.14
	[0.032]	[0.029]	[0.473]	[0.211]
Employment in Quarter 6 (UI)	0.98	0.51	0.28	0.68
	[0.612]	[0.776]	[0.595]	[0.710]
Earnings in the Month of the Survey	9.85	21.78	15.57	4.65
	[0.080]	[0.001]	[0.008]	[0.460]
Employment in the Month of the Survey	8.78	5.59	8.33	1.50
	[0.012]	[0.061]	[0.016]	[0.473]
Earnings in the Quarter of the Survey (UI)	10.46	9.00	11.74	3.53
	[0.063]	[0.109]	[0.039]	[0.618]
Employment in the Quarter of the Survey	3.45	0.24	10.59	0.82

(UI)	[0.178]	[0.889]	[0.005]	[0.664]

Notes: Source: Authors' calculations using the NJS data. Columns two through five of this table report the results from logit models where the binary positive self-evaluation variable is the dependant variable and the categorical variables listed in column one of Table 5 are the independent variables, in addition an outcome variable is included in each regression. Each cell in the table is the result for a different specification where the outcome to be included as an independent variable is different. The values in the table are χ^2-Statistics for joint tests that all of the coefficients are zero for a given outcome, with the p-values in square brackets. For earnings outcomes the continuous variables are entered as four categorical variables: zero earnings, an indicator for being in the lowest quartile of the non-zero earnings distribution, lower middle quartile of the non-zero earnings distribution, upper middle quartile of the non-zero earnings distribution. The omitted category is for those with earnings in the highest quartile of the non-zero earnings distribution. For the employment outcomes a binary variable is included indicating whether the respondent was employed or not. The population for these regressions is the treatment sample. Indicator variables for missing values for the independent variables are also included in the regression.

TABLE 9: Logit Estimates of the Relationship between Before-After Self-Reported Earnings Changes and Positive Self-Evaluation, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
Before-After Self Reported Earnings 2nd Quintile	-0.025 (0.041) [0.540]	-0.008 (0.035) [0.825]	0.030 (0.057) [0.600]	-0.077 (0.059) [0.190]
Before-After Self Reported Earnings 3rd Quintile	0.033 (0.037) [0.375]	0.055 (0.031) [0.077]	0.107 (0.050) [0.031]	0.015 (0.047) [0.752]
Before-After Self Reported Earnings 4th Quintile	0.044 (0.038) [0.250]	0.067 (0.031) [0.029]	0.115 (0.049) [0.019]	0.053 (0.045) [0.239]
Before-After Self Reported Earnings 5th Quintile	0.020 (0.034) [0.547]	0.109 (0.027) [0.000]	0.132 (0.046) [0.004]	0.035 (0.040) [0.379]

Notes: Source: Authors' calculations using the NJS data. "Before-After Self-Reported Earnings" consists of monthly self-reported earnings over the 18 months after random assignment minus monthly self-reported earnings in the 12 months prior to random assignment. The estimates come from logit models with an indicator for a positive self-evaluation as the dependent variable and the before-after earnings change variable and the categorical variables listed in column one of Table 5 as independent variables. The values in the table are mean numerical derivatives, with standard errors in parentheses and p-values in square brackets. The before-after earnings changes enter in the form of indicator variables for being in the 2nd, 3rd, 4th, and 5th quintiles of the before-after earnings change distribution. The omitted category is the 1st quintile of the distribution. The population for these regressions is the treatment group. Indicator variables for missing values for the independent variables are also included in the regression.

TABLE 10: Logit Estimates of the Relationship between Before-After UI Earnings Changes and Positive Self-Evaluation, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
Before-After UI	0.008	0.006	0.164	0.029
Reported Earnings	(0.036)	(0.031)	(0.044)	(0.044)
2nd Quintile	[0.825]	[0.853]	[0.000]	[0.494]
Before-After UI	0.053	0.032	0.127	-0.021
Reported Earnings	(0.035)	(0.030)	(0.044)	(0.043)
3rd Quintile	[0.124]	[0.275]	[0.004]	[0.623]
Before-After UI	0.071	0.083	0.135	0.076
Reported Earnings	(0.034)	(0.028)	(0.044)	(0.039)
4th Quintile	[0.037]	[0.003]	[0.002]	[0.050]
Before-After UI	0.093	0.107	0.143	0.135
Reported Earnings	(0.033)	(0.027)	(0.044)	(0.035)
5th Quintile	[0.005]	[0.000]	[0.001]	[0.000]
Before-After UI (2)	0.015	-0.075	0.015	0.156
Reported Earnings	(0.051)	(0.051)	(0.059)	(0.112)
2nd Quintile	[0.765]	[0.143]	[0.792]	[0.163]
Before-After UI (2)	-0.032	-0.060	0.021	0.024
Reported Earnings	(0.038)	(0.035)	(0.054)	(0.044)
3rd Quintile	[0.407]	[0.083]	[0.697]	[0.579]
Before-After UI (2)	0.022	-0.001	0.045	0.012
Reported Earnings	(0.030)	(0.026)	(0.045)	(0.035)
4th Quintile	[0.473]	[0.960]	[0.326]	[0.723]
Before-After UI (2)	0.065	0.085	0.073	0.092
Reported Earnings	(0.030)	(0.025)	(0.045)	(0.034)
5th Quintile	[0.027]	[0.001]	[0.102]	[0.006]

Notes: Source: Authors' Calculations using the NJS data. "Before–After UI Reported Earnings" consist of monthly UI earnings in the six quarters after random assignment minus monthly UI earnings in the 18 months before random assignment. "Before-After UI (2) Reported Earnings" consist of monthly UI earnings in the 6th quarter after random assignment minus monthly UI earnings in the 6th quarter before random assignment. The estimates come from logit models with an indicator for a positive self-evaluation as the dependent variable and the before-after earnings change variable and the categorical variables listed in column one of Table 5 as independent variables. The values in the table are mean numerical derivatives, with standard errors in parentheses and p-values in square brackets. The before-after earnings changes enter in the form of indicator variables for being in the 2nd, 3rd, 4th, and 5th quintiles of the before-after earnings change distribution. The omitted category is the 1st quintile of the distribution. The population for these regressions is the treatment group. Indicator variables for missing values for the independent variables are also included in the regression.

TABLE 11: Logit Estimates of the Relationship between Before-After Employment Status Changes and Positive Self-Evaluation, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
Employed Before & Not Employed After	-0.059	-0.077	-0.502	-0.282
	(0.102)	(0.093)	(0.230)	(0.139)
	[0.567]	[0.408]	[0.029]	[0.043]
Not Employed Before & Employed After	0.040	0.014	-0.112	-0.013
	(0.028)	(0.024)	(0.045)	(0.034)
	[0.161]	[0.570]	[0.013]	[0.696]
Always Not Employed	-0.123	-0.042	0.071	-0.073
	(0.060)	(0.042)	(0.142)	(0.064)
	[0.038]	[0.319]	[0.616]	[0.249]

Notes: Source: Authors' calculations using the NJS data. Employment status changes are based on changes in self-reported employment status measured at the date of random assignment and 18 months after random assignment. The omitted category is always employed. The estimates come from logit models with an indicator for a positive self-evaluation as the dependent variable and the before-after employment change variable and the categorical variables listed in column one of Table 5 as independent variables. The values in the table are mean numerical derivatives, with standard errors in parentheses and p-values in square brackets. The population for these regressions is the treatment group. Indicator variables for missing values for the independent variables are also included in the regression.

TABLE 12: Test Statistics from Logit Models of the Relationship between Before-After Estimates and Positive Self-Evaluation, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
Before-After Self Reported Earnings	3.95 [0.4130]	25.18 [0.0000]	12.18 [0.0160]	7.54 [0.1100]
Before-After UI Reported Earnings	11.25 [0.0239]	23.20 [0.0001]	18.66 [0.0009]	23.06 [0.0001]
Before-After UI (2) Reported Earnings	7.14 [0.1286]	23.04 [0.0001]	2.95 [0.5659]	9.74 [0.0451]
Before-After Employment Status Changes	10.46 [0.0150]	2.90 [0.4070]	11.15 [0.0109]	5.23 [0.1555]

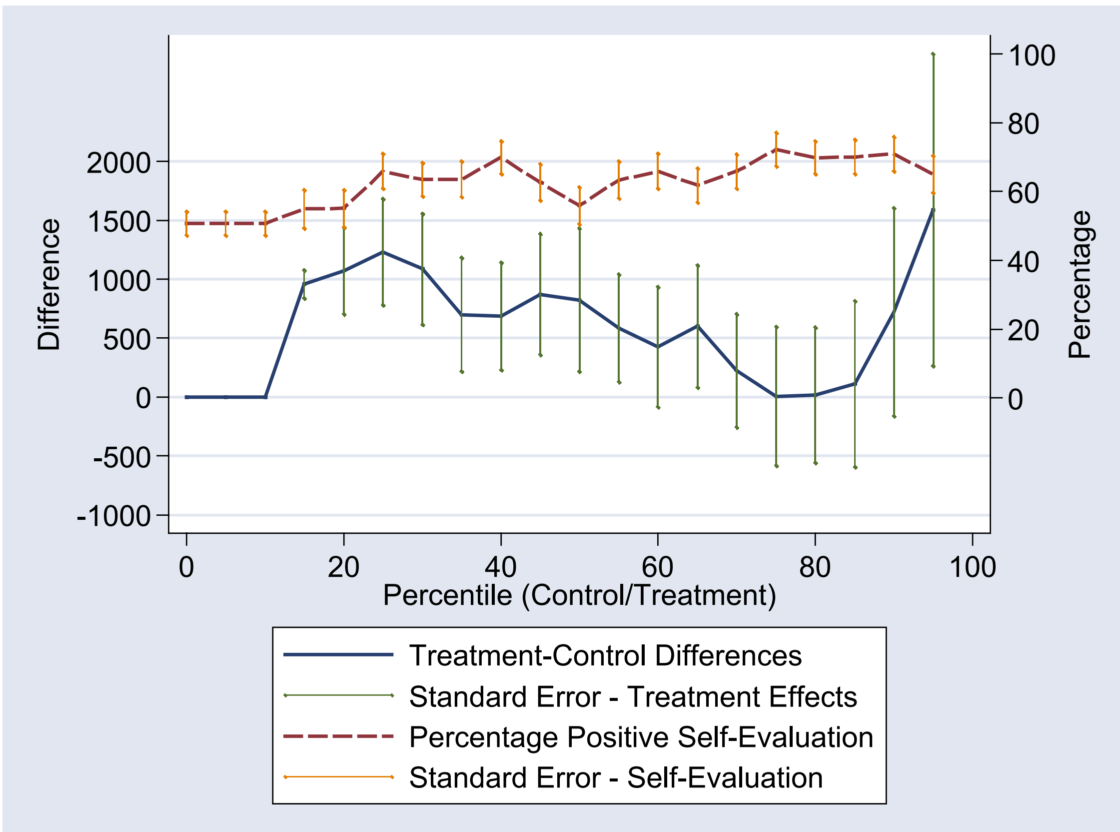
Notes: Source: Authors' Calculations using the NJS data. The values in the table are χ^2-Statistics for joint tests of the null hypothesis that all of the coefficients equal zero for a given outcome, with the p-values in square brackets. The tests correspond to the estimates presented in Tables 9, 10 and 11. See the notes for those tables for further details.

TABLE 13: Logit Estimates of the Relationship between Positive Self-Evaluation and Performance Standards, By Demographic Group

	Adult Males	Adult Females	Male Youths	Female Youths
A. JTPA:				
Employment at Termination	0.115 (0.034) [0.001] n=1507	0.126 (0.029) [0.000] n=1882	0.083 (0.047) [0.078] n=699	0.091 (0.039) [0.020] n= 890
Wages at Termination	0.012 (0.009) [0.173] n=617	0.028 (0.010) [0.003] n=873	0.026 (0.022) [0.232] n=280	0.021 (0.021) [0.316] n=319
Employment at Follow-up	0.097 (0.035) [0.005] n=1507	0.137 (0.029) [0.000] n=1882	0.043 (0.049) [0.380] n=699	0.076 (0.040) [0.056] n=890
Weekly Earnings at Follow-up	0.007 (0.015) [0.623] n=617	0.040 (0.015) [0.007] n=883	0.051 (0.033) [0.120] n=302	0.065 (0.030) [0.028] n=336
B. WIA:				
Employment at Termination	0.052 (0.036) [0.151] n=1155	0.048 (0.032) [0.122] n=1536	-0.073 (0.054) [0.179] n=528	0.035 (0.051) [0.484] n=705
Employment at 6-Months	0.030 (0.046) [0.511] n=659	-0.069 (0.044) [0.123] n=778	0.131 (0.065) [0.045] n=291	0.143 (0.060) [0.018] n=293
Earnings Gain at 6-Months	0.000 (0.000) [0.340] n=566	0.001 (0.000) [0.022] n=634	0.002 (0.001) [0.079] n=240	0.004 (0.001) [0.003] n=204

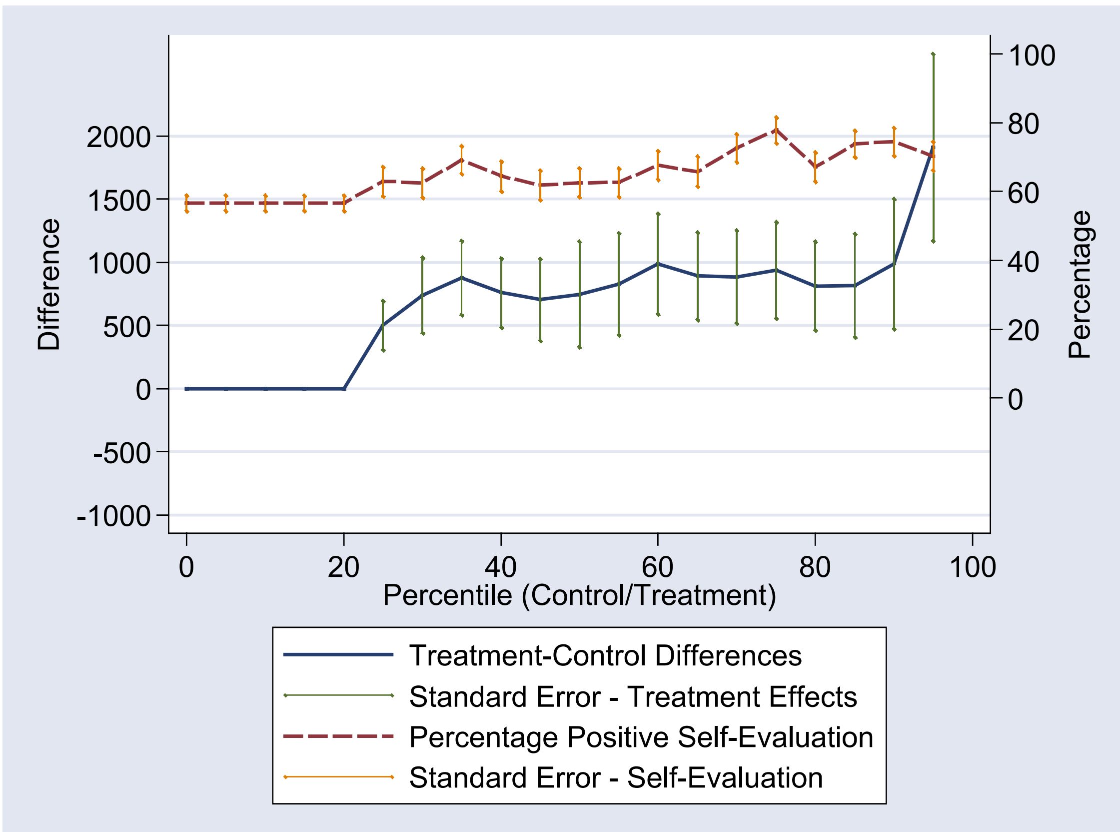
Notes: Source: Authors' calculations using the NJS data. The JTPA performance measures consist of (1) employment at JTPA termination date; (2) employment at follow-up, which is 13 weeks after termination in JTPA; (3) wage per hour at termination date (conditional on employment) in dollars; and (4) the average total weekly earnings at follow up (conditional on employment). Our construction of all of the JTPA performance measures relies on self-reported data. The WIA performance measures consist of (1) employment at exit, which we calculate as non-zero UI earnings in the calendar quarter of termination (conditional on non-employment at the date of random assignment based on self-reported labor force status); (2) employment at six months after termination, which we calculate as non-zero UI earnings in the third calendar quarter after termination (conditional on employment in the first quarter after termination); (3) earnings differences (conditional on employment in the first quarter after termination), which we calculate as the sum of UI earnings in the second and third calendar quarters after program termination minus the sum of earnings in the two calendar quarters prior to random assignment. The estimates in the table correspond to logit models with an indicator for a positive self-evaluation as the dependent variable and one of the performance measures as the only independent variable. The models also include all of the variables listed in Table 5 as additional covariates. The values in the table are mean numerical derivatives, with standard errors in parentheses and p-values in square brackets. We multiply the values for the earnings-based performance measures by 100 for ease of presentation. The final row in each cell gives the sample size for the sample used to produce each estimate. Before deleting observations with missing values of the performance measures, the treatment group samples contain 3067 adult males, 3922 adult females, 1308 male youths, and 1711 female youths. The population for these regressions is the treatment group.

FIGURE 1A: Quantile Treatment Effects and Percentage Reporting Positive Self-Evaluation, Adult Males



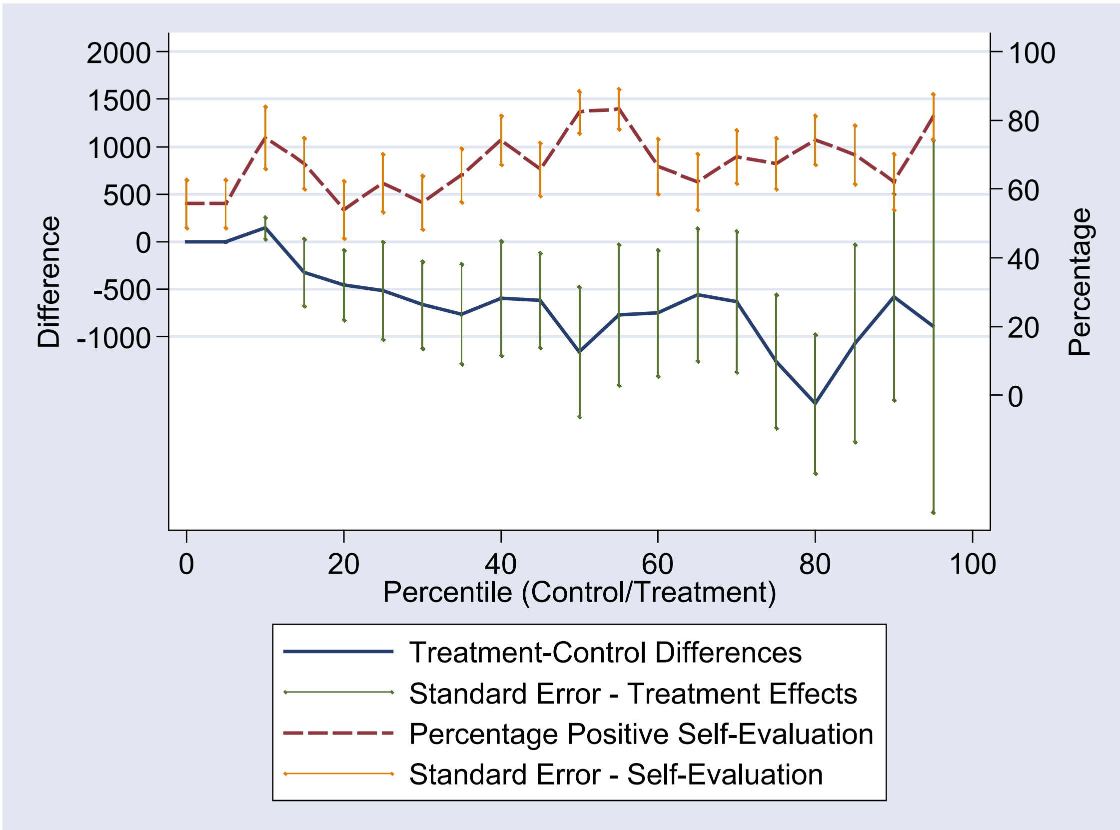
Notes: Source: Authors' Calculations using the JTPA data. The outcome used here is self-reported earnings over the 18months after random assignment.

FIGURE 1B: Quantile Treatment Effects and Percentage Reporting Positive Self-Evaluation, Adult Females



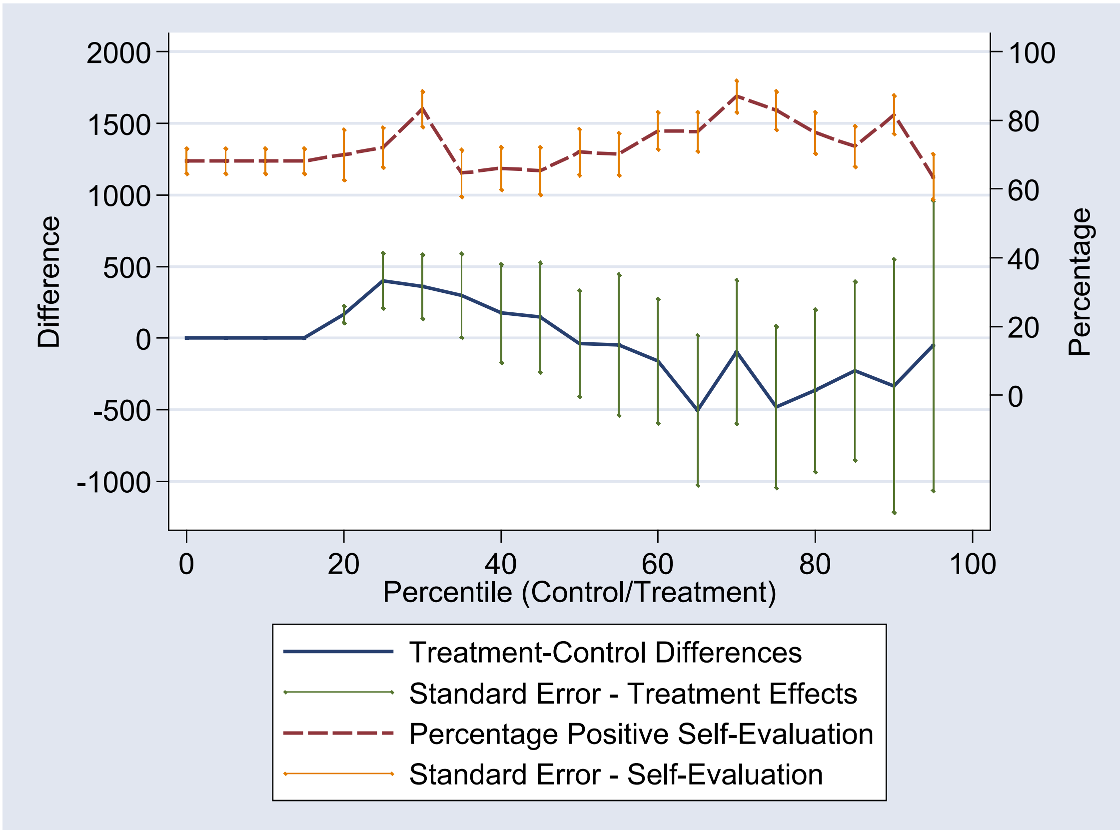
Notes: Source: Authors' Calculations using the JTPA data. The outcome used here is self-reported earnings over the 18months after random assignment.

FIGURE 1C: Quantile Treatment Effects and Percentage Reporting Positive Self-Evaluation, Male Youth



Notes: Source: Authors' Calculations using the JTPA data. The outcome used here is self-reported earnings over the 18months after random assignment.

FIGURE 1D: Quantile Treatment Effects and Percentage Reporting Positive Self-Evaluation, Female Youth



Notes: Source: Authors' Calculations using the JTPA data. The outcome used here is self-reported earnings over the 18months after random assignment.