Does the Quality of Public-Sponsored Training Programs Matter? Evidence from Bidding Processes Data

Jose Galdo*
McMaster University and IZA
galdojo@mcmaster.ca

Alberto Chong
Research Department
Inter-American Development Bank
albertoch@iadb.org

Abstract

This paper evaluates the effectiveness of market-based approaches on the provision of public-sponsored training programs. In particular, we study the link between training quality and labor earnings using a Peruvian program that targets disadvantaged youths. Multiple proxies for training quality are identified from bidding processes in which public and private training institutions compete for limited public funding. Using difference-in-differences kernel matching and standard regression-based approaches, we find that beneficiaries attending high-quality training courses have higher average and marginal treatment impacts. The program is highly cost-effective for women but not for men. External validity was assessed by using four different cohorts of individuals over an eight-year period.

JEL Classification Codes: I38, H43, C13, C14.
Key Words: Training, Quality, Earnings, Bidding, Matching Methods.

---

*Corresponding author: galdojo@mcmaster.ca. Previous version circulated under the title “Training Quality and Earnings: The Effects of Competition on the Provision of Public-Sponsored Training Programs”. Paul Glewwe, Ana Dammert, Jeffrey Kubik, Dan Black, Hugo Ñopo, Jeff Smith, and Oscar Mitnik provided valuable comments. We also thank seminar participants at the 2006 European Econometric Society Meeting in Vienna, Austria, the 2006 Northeast Universities Development Consortium Meeting at Cornell University, and the 2007 Society of Labor Economists in Chicago. Arturo Garcia provided excellent research assistance. We are grateful to the PROJOVEN team, particularly to Milagros Alvarado and Israel Gamarra. The standard disclaimer applies.

1. Introduction

Despite the fact that the empirical evidence on active labor policies suggests that training programs for youth and the displaced are not worth the cost, such programs keep being reinvented by policymakers. In fact, public-sponsored training programs appear to yield small and even negative returns in both developed and developing countries (Heckman, Lalonde, and Smith, 1999; World Bank 2004). In this context, it is by no means clear whether training programs are ineffective because they target relatively unskilled and less able individuals or simply because of the quality of the training itself. After all, the same government agencies that get low grades in training assessments are the ones that end up in charge of implementing training programs.[1]

While a number of authors have reported gains in earnings associated to increments in school or college quality (e.g., Black and Smith, 2003, 2005; Dale and Krueger, 2002, Card and Krueger 1992), corresponding evidence for public-sponsored training programs is non-existent. The predominant approach in the literature is to assume either that training programs have an equal impact on all participants or systematic heterogeneity in the impact of these programs on earnings arises from individual differences in observed and unobserved characteristics (e.g., Heckman, Smith, and Clements 1997, Heckman 2001). Yet training quality has not been incorporated formally in the evaluation of active labor market policies. Nor have the implications for public investment decisions of including training quality been explored.

In this paper, we study the link between the quality of training programs and beneficiaries' subsequent labor market earnings. To our knowledge, this is the first paper

[1]For instance, Campa (1997) shows the limited ability of training programs in Spain to reallocate workers to alternative industries, partly because training was focused on the update of previous skills rather than the acquisition of new ones.

that focuses on quality issues in public-sponsored training programs, with the added advantage that we are able to address the effects of market-based approaches on the provision of training services. In fact, the selection of training courses relies on formal bidding processes in which public and private training institutions compete for limited public funding. Similar program designs have been implemented since the mid-1990s in Chile, Argentina, Colombia, and Uruguay.

The detailed bidding questionnaires and instruments not only allow us to identify common proxies for quality such as expenditures per student, class size, and teacher characteristics, but they also provide information about curricular structure. Moreover, this bidding information allows us to use disaggregated data at the course level, rather than at the school or state level, which may improve the explanation for quality heterogeneity (Hanushek, Kain, O'Brien, and Rivkin, 2005). Furthermore, the availability of data for four different cohorts of individuals over an eight-year period (1996 to 2003) allows us to consider the robustness of our estimates over time with respect to the external validity assumption.

This paper takes advantage of a non-experimental program, the Peruvian Youth Training Program (PROJOVEN), which has provided training to around 40,000 disadvantaged young individuals aged 16 to 25 since 1996. The program has changed the government's intervention in the training market from unconditional funding of public institutions to conditional cash transfers to public and private institutions that offer the best quality courses at the best competing prices. The treatment consists of a mix of formal and on-the-job training organized into two sequential phases, at the training institution and at manufacturing or business firms for a period of six months. To guarantee a paid, on-the-job training experience for each trainee, the program follows a demand-driven approach in which competing institutions must offer training for those occupations with assured labor demand.

The comparison group individuals are selected from a random sample of "nearest-neighbor" households located in the same neighborhoods of those participants included in the evaluation sample. This costly evaluation design greatly ameliorates support problems in the data, which is one of the most important criteria needed for addressing bias due to selection on observables (Smith and Todd 2005). Furthermore, the evaluation framework allows us to identify and compare individuals in the treatment and comparison groups 6, 12, and 18 months after the program, which in turn allow us to test whether the effect of training-quality on labor earnings is constant over time.

The possibility that better students sort themselves into better training institutions is very limited in this program because eligible individuals enroll to the courses according on a first-come-first-served basis, and because there is a large variability in the quality of the training courses within training institutions. To control for potential bias arising from differences in unobserved characteristics, we implement difference-in-differences kernel matching methods, which allow for selection on time-invariant unobservables (Heckman, Ichimura, and Todd, 1997). We also implement an alternative semiparametric estimator that only requires data for the treatment group and thus can be applied when no comparison group data are available (Behrman, Chen, and Todd, 2004).

Our empirical findings can be summarized in three main conclusions. First, we find strong evidence about the effectiveness of market-based approaches in the provision of training services. The PROJOVEN program yield larger overall point estimates than those reported in the literature. This result is particularly robust for women, making the program highly cost-effective. Second, we find evidence of substantial heterogeneity in response to training quality. Our difference-in-differences models estimate a differential effect of 22 percentage points in the earnings of beneficiaries attending high- and low-quality courses 12 months after the program. Third, this paper also shows that individuals who complete both formal training and on-the-job training experience much higher earnings than individuals who complete only formal training. In fact, the returns to formal training are modest and consistent with previous findings in the literature on training programs. The earning differentials between individuals attending high- and low-quality courses are, however, larger within the subsample of individuals who complete only the formal training. This result suggests that the second stage of the program, the on-the-job training experience, smooths productivity gains between people attending training courses of varying quality.

The remainder of the paper is organized as follows. In section 2 we discuss relevant features related to the economics of training quality. Section 3 provides an overview of the PROJOVEN program. We then discuss the measurement of training quality in section 4. Section 5 presents the evaluation data. In section 6 we discuss the empirical strategy along with the main results. In section 7 we report some robustness tests. Section 8 presents the cost-benefit estimates, and section 9 concludes.

2. The Economics of Training Quality

Since the seminal work of Becker (1962) and Mincer (1962), economists have acknowledged the role of training as a potential determinant of labor earnings. This association may be due to human capital accumulation as trained individuals are more productive and, as a result, obtain higher earnings.[2] Several papers using a variety of data sources and econometric approaches have confirmed the main prediction that on-the-job training and earnings are positively correlated (see Parsons 1986 for a review of the training literature).

To gauge the impact of training on earnings, the conventional "quantitative only" approach is to specify an earnings equation, augmented with training measures. The theoretical foundation for this approach assumes that the labor earnings of trainees would not equal their marginal product but would be less for the total cost of training and the post training work span is fixed at *t* and is independent of training. Thus, the equilibrium condition of equating the present discounted value of two income streams associated with training (Y_1) and no training (Y_0) can be written as

$$\ln Y_1 = \ln Y_0 + X^{'}\beta + \delta T + \varepsilon \qquad (1)$$

where X is a set of observed covariates such as schooling, experience, and age, T is a training indicator, and ε is the stochastic term. Figure 1A illustrates the basic effect of training. Trained persons would received lower earnings during the investment training period because the training is paid for at the time, and higher earning are collected after training because of the returns to the investment. This earnings profile (*TT*) is concave as long as the effect of training on earnings is higher in the short-term than in the medium-term. On the contrary, we assume that untrained individuals receive the same earnings before and after the training (*NN*), with the difference between *TT* and *NN* greater the greater the cost of investment and the return from investment (Becker 1993).

[2] Alternatively, since the cost of acquiring training is lower for high-ability individuals, even if training is unproductive, firms may make inferences about productive differences from training choices and workers respond by selecting longer training to signal higher quality. For our purposes, both models yield similar empirical predictions.

To incorporate training quality on this conventional framework, we need two additional assumptions: 1) training quality varies across courses or programs; and 2) individuals do not sort into courses or programs in response to training quality. The first assumption is necessary to obtain empirical estimates because the effect of training quality is only identified if quality is not homogenous. The second assumption is also needed to guarantee that individuals with a lower private cost of learning do not select into high-quality courses or programs. In the context of the PROJOVEN program, there is no question about the heterogeneity of the training services since we observe and measure large variability across training courses. The second assumption, which we address in detail in next section, is more challenging since we cannot observe the level of ability for trainees. Because the enrollment into the courses is based on a first-come-first-served basis and because there is large variability in the quality of the training services within training institutions, we argue that this assumption is likely to be satisfied.

Figure 1B shows how the introduction of training quality alters the conventional framework. Holding fixed the quantity of training, the labor earnings are now conditional on the level of training quality ($q_3 > q_2 > q_1$). Thus, the returns to training investment are higher for higher quality and thus the difference between the earnings profile for untrained and trained individuals will be greater the greater the quality. Moreover, the concavity of the earnings profile may be more or less pronounced depending on whether the effect of the higher quality training on the future stream of earnings will depreciate faster or not.

To illustrate formally the relationship between earnings and training quality, we need to modify the earnings equation (1) to:

$$\ln Y_1 = \ln Y_0 + X^{'}\beta + \delta^{*}T^{*}(T,Q) + \mu \qquad (1')$$

where *T*(T,Q)* represents the effective level of training and depends positively on both the extent and the quality of training; and μ is the new stochastic term.

By looking at equations (1) and (1') it is obvious that if the true relationship is (1') but one estimate the equation (1) the resulting least squared estimates of the returns to training will be biased.[3] In the context of public-sponsored training programs, the identification of the quality effects gets even more complicated since the programs target individuals who self-select into training. Thus, the correlation between the effective level of training and the stochastic errors in the regressions will yield biased estimates. To eliminate bias on observables and time-invariant unobservables one can relax the linear assumption in the earnings equation and implement more data-hungry econometric estimators such as difference-in-differences matching (Heckman et al. 1997).[4]

3. The PROJOVEN Program

To smooth the short-run negative effects of structural reforms on the welfare of poor households in Latin America during the mid-1990s, several countries launched active labor-market policies. In particular, the disproportionately large unemployment rates for young individuals galvanized governments to implement training programs across the region (ILO 2003). The most distinctive element differentiating this generation of training programs from previous public-sponsored experiences was the decentralization

---

[3] In the simplest case if one assume that $T^* = \delta_0 + \delta_1 T + \delta_2 Q$, it is straightforward to show that the OLS estimate of the returns to training will be biased towards zero (see Behrman and Birdsall 1983 and Card and Krueger 1996 for applications in the context of school quality).

[4] For instance, Heckman, Layne-Ferrar, and Todd (1995) find that estimated earnings-quality relationships for schooling quality are sensitive to specification of the earnings function. When false linearity assumptions are relaxed, the evidence for a strong effect of schooling quality on earnings is greatly weakened.

of the training services through market mechanisms in which public and private training institutions compete for public funding (World Bank 2004). The competition was intended to reverse a long period of neglect of the quality of training in public institutions and, thus, to increase the returns to training. The assignment of public funds to any training institution, private or public, is similar in spirit to the school vouchers approach, which is motivated by the idea that competition will be translated into expanded access and enhanced service quality, and thereby improved labor market outcomes (see Carnoy 2001 for a review of the school vouchers literature).

The Youth Training Program PROJOVEN was implemented in 1995 with the goal of increasing the employability and productivity of disadvantaged young individuals aged 16 to 25 through vocational training in blue-collar occupations.[5] The treatment consists of a mix of formal and on-the-job training organized into two sequential phases. The first stage consists of 300 hours of classes at the training center locations; or roughly five hours per day for three months. The program provided a stipend to trainees during these 3 months of the program to cover for transportation and lunch of US$2 per day for men and women without children and of US$3 per day for women with children under 6 years of age to cover childcare expenses. In the second phase, training institutions must place trainees into paid, on-the-job training experiences in private manufacturing firms for an additional period of three months.

To ensure the relevance of the training courses, the program relies on a demand-driven mechanism that stipulates that all training centers must present, as part of their bids, formal agreements with private manufacturing firms that guarantee a paid, on-the-job training slot for each beneficiary. This program design requires a strong match

[5]These occupations are, for example, maintenance mechanic, electricians, receptionist clerks, construction laborers, plumbers, pipefitters, sewing machine operators, textile operators and tenders, etc.

between the content of the training courses and the firm's labor skill requirements and thus a strict coordination between the training institutions and the manufacturing firms in designing and implementing the training courses. As a result, the coverage of this training program is limited because of its costly design and relatively intense package of services.

If the firms receive unproductive workers, they are entitled by law to drop their labor contracts at any time. Responsibility for the completion of both phases of training falls solely on the training institutions. A system of conditional payments based on the training centers' effectiveness at getting trainees to successfully complete the six-month course provides the incentives to train only for those occupations with assured labor demand.[6]

3.1 The Beneficiary Selection Process

PROJOVEN's selection process consists of several stages governed by different actors: target individuals, bureaucrats, and training centers. Figure 2 shows the dynamics of this process. The program awareness strategy (position A) constitutes the first formal effort to reach out to the target population and aims to inform potential participants about the program's benefits and rules. This first filter focuses only on poor neighborhoods. Those prospective participants attracted by the expected benefits and perceived opportunity costs of participation voluntarily show up in the registration centers (position B) where qualified personnel determine their eligibility status. A standardized targeting system based on some key observable variables (poverty status, age, schooling, labor market

[6] Payments are structured in per capita terms according to the following scheme: 100, 80, 60, and 30 percent if completing both phases of the program, at least one month of on-the-job training, only formal training, and at least a month of formal training, respectively.

status) determines who is eligible and who is not. This process concludes when there are nearly twice as many eligible individuals as training slots.[7]

Eligibility status does not guarantee participation in the program. Program enrollment depends on the applicants' willingness to pursue the application process to its conclusion. Eligible individuals are invited to an orientation process (position C), where they choose the courses they want to attend following a first-come-first-served criterion. Finally, training institutions select beneficiaries from the pool of eligible applicants sent by the program operator (position D). This is the only step where training institutions intervene in the selection process and does not follow standardized criteria since each institution applies its own rules. Institutional evidence suggest, however, that training centers have very limited role in selecting beneficiaries because the number of eligible individuals exceeds by about 25 percent the number of available training slots in each course.

4. Measuring Training Quality

The selection of training services follows a two-step standardized process. The first step targets the selection of training institutions. The program operator consults a training directory called the RECAP, which lists all the training institutions eligible to participate in the program. To be included in the RECAP, the training centers must pass a minimum quality threshold after the program administrator verifies their legal status (formality) and

[7] A two-tiered monitoring and supervision process guarantees the reliability of the information given by the prospective applicants to determine their eligibility status. In addition to focusing only on targeted poor districts, the program administrator makes house visits to those applicants who provided dubious or inconsistent information. Finally, a random sample of eligible and non-eligible individuals is subject to an *ex-post* visit, which allows the program administrator to detect misreported cases and improve the eligibility survey and instruments.

the existence of some acceptable level of human resources and infrastructure. In this first step, institutions do not compete with each other, there are no restrictions as to the number of institutions that can be listed in the RECAP, and the quality threshold is loosely determined.

In the second step, the program administrator invites institutions included in the RECAP to participate in public bidding processes in which the selection of training courses rather than training institutions takes place. The PROJOVEN design competition relies in a model of two-dimensional auctions in which each training institution bids on both quality and price. The PROJOVEN's Terms and Conditions regulate these processes and follow a set of international standards that were previously approved by the Inter-American Development Bank and the United Nations as part of their role as guarantors in this program.[8]

This formal document also includes standardized forms and instruments that must be presented at the bidding. They are constructed by education specialists with the express purpose of capturing both quantitative (e.g., number of computers, number of instructors, etc), and qualitative (e.g., curricula, activities, etc) information about each competing course. The training institutions must clearly respond to all points set up in these technical forms. Once the deadline is reached, two sealed envelopes containing the technical specifications and price offers are opened in a public act, where the price offers are made public. The documents containing the technical specifications, on the other

[8] The PROJOVEN's Terms and Conditions can be found at *http://www.projoven.gob.pe.* This document stipulates, for instance, the starting and closing dates for the bidding, the schedule of the payments, potential conflict of interest, penalties, etc; and the technical specifications for the courses including the selection of the trainees, the minimum and maximum number of students per class, the duration of the courses, the core activities, etc. Both the Inter-American Development Bank and the United Nations have played an important role in the transparency of these processes because of their involvement in the funding and administration of the funds, respectively.

hand, are subject to blind evaluation during a two-month period or so. In this process, a small team of education specialists assigns standardized scores to multiple proxies for quality following a battery of standardized instruments. The score system is confidential and, therefore, unknown to the competing institutions. The program operator maximizes the (summed) ratio scores/prices subject to both budget and training slots restrictions to determine the list of winning courses. Appendix A1 describes in detail how course quality and prices get taken account of in the bidding processes.

Three distinctive features characterize the quality measurement. First, all proxies for quality are disaggregated at the course level rather than at the school level, which allows us to measure the quality of the training services in great detail. Thus, variations can be found within training centers depending on the relative distribution of school supplies or differential teacher experience across courses.

Second, detailed questionnaires and instruments not only collect information on common proxies for quality such as expenditures per student, class size, infrastructure, and equipment, but also put emphasis on the curricular structure (i.e., consistency among goals, contents, and activities) and teacher "skills" (i.e., experience in dealing with disadvantaged young individuals).

Third, the measurement of quality proxies follows a standardized system of scores rather than the classical approach of computing raw quantities (e.g., number of computers). In this way, the evaluators are able to evaluate both the number of items in each subcategory and their intrinsic quality. For example, in evaluating a course on computing software, the total score in the equipment variable will depend on both the quantity of computers per student and the model and age of the machines. The use of standardized scores also allows for the evaluation of variables such as curricular structure that do not per se have a corresponding quantitative content. Only two proxies for quality are measured in raw form: expenditures per student and class size.

This paper focuses on 6 different categories of proxies for quality: class size, expenditures per trainee, 8 teacher variables, 6 variable measuring infrastructure and equipment physical characteristics, 19 curricular structure variables, and 9 variables characterizing the link between the content of the training courses and the institution's knowledge about workers and occupational analysis of labor demand. We have information only for the aggregate scores in each quality category. As a whole, these variables largely exceed the number of school and teacher characteristics considered to be core variables in the literature (Fuller 1987; Harbison and Hanushek, 1992).

Table 1 displays summary statistics of these quality measures using re-scaled indices for all categories. We use data from 1996 to 2003, which allows us to identify four different bidding processes corresponding to the first, second, fourth, and sixth programs.[9] We observe strong variation in the scores assigned to each category within and across programs. Moreover, as one might expect, there is an increasing trend in the average scores for some categories over time, particularly for variable measuring infrastructure and equipment characteristics. This is explained by a natural learning curve on the part of continuously participating institutions and by the relatively small number (9) of new entering training institutions.

Because we think that each individual quality proxy represents an error-ridden measure of underlying quality, we combine the information for all quality categories using factor analytic methods to produce a one-dimensional quality index. In doing so, we use the first principal component, which is a linear combination of the quality proxies

[9] A previous version of this study considered five rounds of the program. We do not consider the eighth round because the lack of data for the third follow-up survey

that accounts for the highest proportion of their variance. This proportion ranges from 42 percent in the first program to 20 percent in the last one, which indicates a larger quality variation in the first rounds of the program. The resulting scoring factors (weights) on individual indicators are relatively constant over time for all categories but infrastructure and equipment variables, which also receive the lowest weights. Curricular structure and market knowledge, on the contrary, are especially important for the index.

The lower panel of Table 1 reveals that the average number of training institutions is 33 per program, ranging from 30 to 35. These institutions offer on average 183 courses per program. We also observe that the supply of training courses and the number of selected courses follow parallel paths. The ratio of funded courses to competing courses reaches 0.59, which indicates a relatively high probability of success for those training institutions included in the RECAP.

Two potential factors that may affect the accuracy with which the quality proxies are measured are evaluation bias and misreporting. In the first case, evaluators may introduce bias when assigning scores due to subjective evaluation. The program administrator, however, minimizes this risk by hiring a small team of education specialists who are trained to follow a standardized score system and are under strict supervision. The competition for limited public resources may also encourage training centers to misreport their public offers. To minimize this problem, the program administrator has implemented a monitoring system that uses inspections before and during the training to ensure the validity of all technical specifications contained in the offers. The bidding data are then merged with the evaluation data, which implies that all treated individuals attending the same training course receive the same quality scores.

5. The Evaluation Data

From 1996 to 2003, the period for which we currently have data, the PROJOVEN evaluation datasets consist of 8 different sub-samples associated with 4 different cohorts of beneficiaries receiving treatment in Lima, and 4 corresponding comparison group samples. The beneficiary subsamples are selected from a stratified random sample of the population of participants corresponding to the first, second, fourth, and sixth rounds of the programs.[10] Individuals in the corresponding comparison subsamples are selected from a random sample of "nearest-neighbor" households located in the same neighborhood as those participants included in the evaluation sample. The neighborhood dimension may have the ability to control somewhat for unobservables, including geographic segregation, transportation costs, and firms' location, which may affect the propensity to work and the potential outcomes. The program operator builds the comparison samples by using the same eligibility instruments applied to the treatment sample and by pairing each beneficiary to a random neighbor who has the same sex, age, schooling, labor market status, and poverty status.

For each treated and untreated cohort combination, we have panel data collected in 4 rounds including a baseline and 3 follow-up surveys taken 6, 12, and 18 months after ending the program. Comparison group individuals are not allowed to participate in the PROJOVEN program until the end of the third follow-up survey. The baseline survey provides rich information for all variables that define the eligibility status. It also contains demographics and labor-market information. There is information, for example, on

[10] Individuals that satisfy the same eligibility criteria in terms of age, education, poverty status, and labor outcomes compose all four cohorts. The only difference among these groups is the period when they receive treatment. These periods extend from November 1996 to April 1997; February 1998 to July 1998; March 1999 to August 1999; and June 2000 to December 2000, respectively. The total number of participants in these program rounds is 1507, 1812, 2274, and 2583, respectively. The corresponding number of treated individuals in the random sample is 299, 321, 343, and 405.

educational attainment, marital status, number of children, and parents' schooling. The labor-market module includes information about labor force participation, experience, monthly earnings, working hours, occupation, firm's size, and participation in previous training courses. In addition, the datasets provide detailed information on dwelling characteristics including source of drinking water, toilet facilities, and house infrastructure, which proxy for household's long-run economic status (Filmer and Pritchett 2001).

Moreover, the follow-up surveys provide detailed labor-market information for both treated and comparison groups, using the same definitions and variables as the baseline instruments, which minimize potential biases due to misalignment in the measurement of variables. The response rate to the initial survey was 100 percent and the attrition rates are small ranging between 4 percent (12 months after the program) and 7 percent (18 months after the program).

5.1 Data Description

Columns 2 to 7 in Table 2 compare the pre-treatment means of several covariates for the treatment and comparison samples for both men and women. Panels A and B show the effectiveness of the "neighborhood" strategy in balancing the distribution of covariates that determine the eligibility status. Treatment and comparison groups have the same average age, sex composition, schooling attainment, and labor market participation. On the other hand, the data show statistically significant differences in marital status and monthly earnings. We also observe that women are more educated, have higher married rates, and smaller labor earnings than men.

Panel C compares household and dwelling characteristics. The data shows that a higher proportion of treated individuals live in houses with somewhat better infrastructure and access to a flush toilet and piped water. These statistical significant differences hold for both male and female samples. Finally, Panel D shows parental schooling attainment. In general, the schooling distribution is similar for the treatment and comparison groups, with mothers having fewer years of formal education than their spouses.

Columns 8 to 13 in Table 2 repeat the same analysis for treated individuals above and below the 50 percentile of the quality index distribution. We do not reject the null hypothesis of equality of means for all observed covariates. We observe, for both men and women, that all demographic, labor, and household characteristics are similar across individuals attending high- and low-quality courses. This result suggests that the first-come-first-served strategy had indeed the ability to randomize individuals in courses of varying quality.

6. The Empirical Strategy

We focus on both average and marginal treatment impacts conditional on the quality of the training courses. Let $Y_1(q)$ be the potential outcome in the treatment state $(T=1)$ for an individual who participated in a training course of quality q and let $Y_0(q)$ be the potential outcome in the untreated state $(T=0)$. In our application, the untreated state refers to either no participation in the program, in which case $q=0$, or participation in a training course of lower quality. Our parameter of interest is the impact of treatment on the treated, which estimates the mean effect of attending a high-quality training course rather than not participating (or attending a low-quality course):

$$\Delta_{TT} = E(Y_1(q) - Y_0(q) \mid T = 1) = E(Y_1(q) \mid T = 1) - E(Y_0(q) \mid T = 1)\,. \qquad (2)$$

While $E(Y_1(q) \mid T = 1)$ may be estimated from the observed treatment sample, the right-hand side of the equation (2) contains the missing data $E(Y_0(q) \mid T = 1)$. Using the outcomes of untreated individuals to approximate the missing counterfactual yield the well-known selection bias because of differences in the distribution of observed and unobserved characteristics between *T=1* and *T=0.* To eliminate bias due to selection on unobservables, we relax any linear assumption and proceed under the assumption that the distribution of unobservables varies across *T=1* and *T=0* but not over time within groups, which is the standard assumption of difference-in-differences matching models.

6.1 Identifying Mean Impacts when the Counterfactual is Non-Participation ($q = 0$)

In general, standard matching methods eliminate selection on observables by balancing the distribution of relevant covariates of the untreated group with that of the treated group. However, even after conditioning on a rich set of observables, selection bias may not be eliminated because of differences in unobservables between the treated and comparison group. Such differences may arise in the PROJOVEN program from three different sources. First, it is impossible to control differences in innate ability or motivation using our data. Second, we do not observe all the factors that govern the transition from eligible status to beneficiary status. Third, we may not observe and measure certain aspects of teacher and school quality that are not correlated with the quality index.

To eliminate bias arising from (time-invariant) unobservables, we use difference-in-differences matching methods (Heckman et al. 1997) that are conditional semiparametric versions of the widely used parametric approach. This method solves the evaluation problem by subtracting the before-after change in untreated outcomes from the before-after change for treatment outcomes. The identifying assumption justifying this matching estimator is that there exists a rich set of conditioning variables *X* such that

$$E(Y_t(q) - Y_{t'}(q) \mid X, T = 1) = E(Y_t(q) - Y_{t'}(q) \mid X, T = 0). \qquad (3)$$

where *t'* and *t* refer to before and after the start of the program. This assumption ensures that after conditioning on a rich set of observable variables, the outcomes for treated and untreated individuals follow a parallel path. The availability of an unusual rich baseline data makes plausible this assumption. Several relevant factors affecting both the propensity to participate in the program and labor market outcomes are available. Moreover, the program's strategy to construct the comparison group based on a random sample of "nearest-neighbor" households has the potential to control for unobservables affecting the outcomes.

Matching methods force us to compare comparable individuals by relying on the common support assumption

$$\Pr(T = 1 \mid X) < 1 \text{ for all } X. \qquad (4)$$

The support condition ensures that for each *X* satisfying assumption (3) there is a positive probability of finding a match for each treatment individual. Otherwise, if there are *X* for which everyone received treatment, then it is not possible for matching to construct the counterfactual outcomes for these individuals.

Under conditions (3) and (4), we estimate the treatment impacts by computing first the counterfactual outcome for each treatment unit using a weighted average of the comparison units' outcomes over the common support region, and then averaging these results over the treatment group sample

$$\Delta^{DID} = \frac{1}{n_1} \sum_{i \in n_1 \cap S_p} \left\{ [Y_t(q) - Y_{t'}(q)] - \left\{ \sum_{j \in n_0 \cap S_p} W(i,j)[Y_t(q) - Y_{t'}(q)] \right\} \right\}. \quad (5)$$

where n_1 and n_0 are the sample of treatment and comparison individuals, S_p is an indicator function that takes the value 1 for individuals in the common support region (0 otherwise) and $W(i, j)$ is the kernel weighting function that depends on the Euclidian distance between each comparison group individual and the treatment group individual for which the counterfactual is being constructed. We implement local constant regression methods (Pagan and Ullah 1999) rather than local linear methods (Fan 1992) because of their finite sample performance (Frölich 2004, Galdo, Smith, and Black 2007).[11]

6.2 Identifying Marginal Program Impacts

We are also interested in the marginal treatment impacts of increasing quality in the program using data on program participants who have received different qualities of treatment. An important advantage of using only treatment individuals is that we do not require assumptions about the process governing selection into the program. On the other hand, this approach may introduce a potential source of nonrandom selection because of

[11] Local linear estimators were developed in the early 1990s by Fan (1992) and have more recently been considered in the evaluation literature by Heckman, Ichimura, Smith, and Todd (1998). These methods converge faster near boundary points and thus have lower boundary bias in regions of sparse data. At the same time, these methods demand more of the data because they estimate one additional parameter in every local regression. This suggests the possibility that the local constant estimator might have lower mean squared error in finite samples. Indeed, Frölich (2004) and Galdo et al. (2007) show smaller mean squared error for the local constant estimator via Monte Carlo experiments.

potential sorting. Indeed, this is the main econometric problem in studies addressing the link between college quality and labor earnings (e.g., Black and Smith, 2005).

Three features limit the chances that high-ability individuals select into high-quality courses in the PROJOVEN program. First, as shown in Table 2, the first-come-first-served strategy has indeed randomized eligible individuals in courses of varying quality. This is not the case in studies addressing college education. Second, there is large variability in the quality of the training courses within training institutions. Thus, even if more able individuals manage to get in line ahead of less able individuals and select the training centers where they would like to go, they may end up enrolled in low-quality courses. Third, there is no evidence that training institutions use any sort of IQ tests to select the program's beneficiaries among the eligible population.[12]

Even though this favorable assessment for the PROJOVEN program, we cannot ignore some unobservables that may cause sorting in our data. Thus, we again implement difference-in-differences matching methods that assume selection on time-invariant unobservables to eliminate selection bias. Formally, we assume that the mean outcomes for individuals participating in high-quality courses follow a parallel path with respect to individuals attending low-quality courses. We estimate the marginal treatment impact using the same matching estimator (equation 5), although implemented using data only for treatment observations.

Because the "curse of dimensionality" arises when X is high dimensional, we follow the celebrated result of Rosenbaum and Rubin (1983), who show that if the information set contained on X justifies matching, then matching on the propensity score

[12] From interviews with both the program administrator and training institution personnel, it seems that the selection of beneficiaries among the eligible individuals is driven by variables such as sex and specific physical requirements arising from the courses (e.g., body mass for handling weights).

$P(X)$ is also justified. In the empirical work, we estimate the propensity score using a logit approach and implement the balancing test suggested by Dehejia and Wahba (1999).[13] Appendix A2 shows the logit results. Importantly, the distributions of the estimated propensity scores indicate no support problems in our data. Less than 5 percent of the observations are out of the empirical overlapping region, which illustrates the relative efficiency of constructing comparison groups among eligible "neighbor" individuals. In this respect, our data satisfy one of the most important criteria needed for solving the evaluation problem (Smith and Todd 2005).[14]

To implement the local constant kernel matching (equation 5) we also need to compute kernel functions along with their optimal bandwidths. We adopt the unbounded Epanechnikov kernel and choose bandwidth values by weighted least squares cross-validation (Galdo et al. 2007), which selects the value that minimizes the mean square error of the local constant regression estimator over a bandwidth search grid. The weights account for the location of the treated units because precise estimation of counterfactuals is more important in regions containing much of the probability mass of the treatment group individuals than in regions where few treated individuals are located. The bandwidth grid is defined over values 0.1 through 5, with a step size of 0.1.[15]

---

[13]This test considers valid any parametric models that balance the distribution of pre-treatment covariates between matched individuals conditional on the propensity score. It is important to indicate, however, that multiple versions of the balancing test exist in the literature, and little is know about their statistical properties or the relative efficiency among them.

[14]We follow the "trimming" method (Heckman et al. 1998), which seems to be more stringent than alternative approaches suggested in the literature. Hence, we estimate the propensity score density distributions for $T=1$ and $T=0$ using Epanechnikov kernel functions. Then, the estimated densities are evaluated at all observed data points and, all points with zero density and points corresponding to the lowest 5 percent of estimated density values are trimmed.

[15]Relative to their frequency in a random population, the treatment group individuals are oversampled. Thus, we apply matching methods to choice-based sampled data and thus we use the log of the odd ratio $\hat{P}(X)/1-\hat{P}(X)$ as the matching variable (Heckman and Todd 1995).

6.3 Difference-in-Differences Matching Estimates

Table 3 presents difference-in-differences matching estimates for each one of four different cohorts and for the pooled male and female samples. The upper panel (A) depicts short-term treatment impacts whereas the lower panels (B and C) present medium-term impacts. Within each panel, three different parameters of interest are presented: the average treatment effect on the treated, the average treatment effect on those attending a high-quality course, and the average treatment effect on those attending a low-quality course. In all three cases, we estimate the counterfactuals using the comparison group sample. The point estimates along with their corresponding percentage gains (in brackets) and bootstrap standard errors (in parentheses) are reported.[16]

By looking at the first row of each panel, one can observe that the beneficiaries attending PROJOVEN show large earnings impacts. The overall treatment impacts on the treated are 48 percent 6 months after the program, 30 percent 12 months after the program, and 54 percent 18 months after the program. This positive assessment of the PROJOVEN program are in line with partial evaluations of the first (Galdo 1998) and sixth (Ñopo, Saavedra, and Robles 2001) programs, which were based on conventional parametric estimators.

These earnings gains are driven by female beneficiaries that show much higher treatment impacts than male beneficiaries (74 versus 5 percent 12 months after the program). The large number of individuals that relocated from low-quality jobs toward productive ones in private firms protected by international labor-standard laws may explain these large gains. For example, the number of beneficiaries working as either

[16] The percentage gains are calculated by using the comparison group mean earnings in the baseline period: 127 soles for the whole sample, 171 soles for men, and 90 soles for women. All figures in real values of December 2001. The exchange rate (dollar/sol) was 3.4 in December 2001.

unpaid family workers or housekeepers decreases by 33 and 20 percentage points for women and men 12 months after the program.

The second and third rows within each panel show the average treatment impacts for those attending high- and low-quality training courses. In general, the matching estimates indicate that trainees attending high-quality courses have higher labor-market earnings than those trainees attending low-quality courses after controlling for systematic differences in observed and time-invariant unobserved covariates. By looking at the whole sample, we observe that 6 months after the program the differential effect between high and low-quality courses reaches 16 percentage points, and increases to 22 and 27 percentage points 12 and 18 months after the program. These earnings-quality relationships differ when splitting the sample for gender. Whereas men attending high-quality courses report higher earnings than those attending low-quality courses, we observe large treatment impacts for women attending both low- and high-quality training courses.

Table 4 presents the marginal matching estimates for the pooled sample.[17] We show short- and medium-term treatment impacts in the upper and lower panels. Within each panel, we present four marginal treatment impacts: the effect of increasing the quality of the training services from q_1 (lowest quartile) to q_4 (top quartile), from q_2 (second quartile) to q_4, and from q_3 (third quartile) to q_4. Two main patterns emerge. First, the marginal impacts for the whole sample indicate positive treatment impacts for those attending high-quality courses rather than low-quality courses, although we lose statistical significance due to sample size issues. Second, for men, the treatment impacts

[17]We match on the predicted probability of attending a high-quality training course using the same set of regressors as before plus a set of dummies for the type of institution attended. These propensity score models are not reported but they are available upon request.

are monotonically increasing as one moves to higher quartiles of training quality. In particular, the impacts for attending a training course in the fourth quartile rather than the first quartile are very large and statistically significant. For women, however, the quality-earnings relationship are only positive when one moves from the first to second quartile and with no clear results when one moves to the third and fourth quartiles.

When the estimates from Tables 3 and 4 are taken together, several lessons emerge. First, market-based approaches that put great emphasis in the quality and pertinence of the training courses yield larger overall point estimates than those reported in the literature. This intensive and costly program has the ability to relocate workers from low-quality jobs toward productive ones in firms protected by international laws that guarantee minimum work conditions.[18] These earnings gains are heightened by the fact that the per-capita expenditures on participants are not small relative to the deficits that this program is being asked to address. Whereas the Peruvian public school system spent about $350 dollars per-capita in 2001, the PROJOVEN program spends $515 dollars per-capita. The cost-benefit section will address in more detail this point.

Second, female beneficiaries demonstrate much larger treatment effects than male participants. Three factors may explain this finding. First, a larger proportion of women (64 versus 49 percent) complete at least one month out of three of the on-the-job training experience. Second, women are more educated than men (86 versus 81 percent have completed high school), which may affect the likelihood of getting or keeping a job

[18]We also illustrate this fact by estimating conditional probabilistic models that use firms' size as the dependent variable (1 if working in firms with more than 20 workers, 0 otherwise) and treatment status as the key independent variable for the pooled data. The estimates show that treatment group individuals are 52 percent more likely to work in medium- and large-size firms than comparison group individuals 12 months after the program. These estimates are statistically significant at 5 the percent level. It is important to note that the distribution of treated individuals across firm size is symmetric for individuals in the top and bottom quartiles of quality index.

conditional on participating in the PROJOVEN program. Finally, in the absence of the program, men have much higher earnings and labor participation rates than women. Thus, the program's effect on the earnings growth for men who primarily work in low-pay occupations is very limited.

Third, reporting simple average treatment impacts hide important distributional gains due to heterogeneity in the quality of the training services even within a selected group of institutions that pass a quality threshold. This result suggests that the earnings gap between high- and low-quality courses would be higher if the program administrator allowed the participation of training institutions located below the cut-off point.

Four, the results also indicate a downward trend in the evolution of the gains over time, which is consistent with theoretical predictions emerging from human capital models (Becker 1993). Our estimates suggest that the stock of training depreciates between 10 and 25 percent in the second year after the program, which are in line with international evidence on on-the-job training that suggest a depreciation rate of about 20 percent (Lillard and Tan 1986, Almeida and Carneiro 2005).

Five, this study highlights the importance of having multiple cohorts of participants for the same program design when assessing the effectiveness of active labor market programs. The sensitivity of some estimates to the sample used is illustrative. For instance, the combined effect for the first two program are statistically significant different from the combined effect for the last two ones.

7. Robustness Checks

7.1 Ashenfelter's Dip and Sensitivity to the Econometric Estimators

The difference-in-differences approach may be sensitive to the specific period over which the 'before" period is defined if we observe a drop in the mean earnings of participants prior to program entry (Ashenfelter 1978, Heckman and Smith 1999). Figures 3A and 3B compare the earnings trajectory for the treatment and comparison groups. There is evidence about the existence of Ashenfelter's Dip in the PROJOVEN program for both men and women that may bias the estimates. Because of data limitations, we cannot argue whether the pre-program drop in earnings is permanent or transitory. However, evidence from employment patterns in the months prior to the program is more consistent with the hypothesis of transitory drops in earnings, which implies that our estimates may be upwardly biased. It is noteworthy that the full post-program earnings trajectory is consistent with the point estimates emerging 6, 12, and 18 months after the program for both men and women.

Alternative econometric estimators that are consistent when the model of program participation stipulates pre-program earnings dip can address the issue about the robustness of our estimates. We use a standard regression-based estimator of the difference between the post-treatment earnings of treatment and comparison group members, holding constant the level of pre-treatment earnings and a set of control variables that includes the propensity score. This estimator identifies consistently the parameters of the regression model in the context of Ashenfelter's dip (Lalonde 1986). In addition to the conditional covariates, we include dummy variables for having attended a course in the fourth, third, second, and first quartile of the quality distribution. The control group indicator is the omitted group and, therefore, the implicit counterfactual.

The OLS analysis estimates the effect of a treatment under the assumptions of selection on observables and that simply conditioning linearly on *X* suffices to eliminate selection bias.

The results in Table 5 indicate that the difference-in-differences matching estimates are somewhat bigger than the OLS estimates, which is consistent with the pre-treatment earnings dip observed in the data. The OLS treatment impacts for the whole sample are 42, 24, and 33 percent 6, 12, and 18 months after the program. For the same reference periods, the treatment impacts for those attending high-quality courses (top quartile) are 39, 18, and 33 percent, while for those attending low-quality courses (bottom quartile) are 29, 6, and 31 percent. All these OLS treatment impacts are statistically significant at the 5 percent level. Furthermore, these parametric estimates corroborate the evidence of large treatment impacts for women and modest impacts for men. The small average gains for men are driven by the negative treatment impacts experienced by those attending low-quality training courses. For women, on the other hand, the treatment impacts are large across all quartiles of training quality.

7.2 Quality Dose versus Treatment Dose

The estimates for the returns to training quality may also be interpreted as returns to treatment dose rather than quality dose, because of differences in the duration of the on-the-job training experience among trainees. To address this potentially confounding factor, we use two different approaches. First, we check whether individuals enrolled in high-quality courses (quartile 4) have larger treatment doses than individuals enrolled in low-quality courses (quartile 1). For both men and women, we find a slight difference in favor of individuals attending low-quality courses. Over 98 percent of trainees enrolled in both low- and high-quality courses complete at least the first stage of the program, whereas 58 and 53 percent of women (45 and 41 percent of men) attending low- and high-quality courses complete the three months of on-the-job training experience.

Table 6 presents a second, more stringent test. We estimate average treatment impacts on the treated by using both difference-in-difference matching and OLS methods applied to the subset of individuals who complete the training course at the training center location but do not participate in the paid, on-the-job training experience. In this way, we hold fixed the treatment dose and, at the same time, we eliminate any potential effects arising from differences among manufacturing firms that may mask the causal effect of the training quality.

Two patterns emerge. First, the treatment impacts for the formal training are positive although much smaller with respect to the overall mean program impacts. This result holds for both men and women. In fact, the returns to formal training, particularly those emerging from the OLS estimation, are modest and consistent with previous findings reported in the literature on training programs (Heckman et al. 1999). Second, the treatment impacts are much larger for those beneficiaries attending high-quality courses within the group that complete only formal training. For instance, one year after the program, the difference-in-differences matching estimates for males attending high- and low-quality courses are 31 and -51 percent, whereas the corresponding OLS estimates are 9 and -58 percent. For women, these earnings-quality relationships are present in the short-term but not in the medium-term. The point estimates, however, are not statistically significant because the small sample sizes involved in the estimation.

Taken together, the estimates in Tables 3 and 6 impart two related lessons. First, the on-the-job training experience matters in terms of both magnitude and statistical significance of the point estimates. When comparing the impacts for those who completed only the first stage of the program (Table 6) with the overall impacts (Table 3), we observe large differences that suggest that formal training alone may not worth the cost. Second, the earnings differentials between people attending high- and low-quality courses are higher for the subsample of individuals who participate only in the first stage as compared to those for the whole sample. This suggests that the on-the-job training experience has the ability to smooth the strong training quality effects on labor earnings across individuals attending low- and high-quality courses. This in turn may explain why the earnings-quality relationships are stronger for men than for women given the larger proportion of women that complete the program.

8. Cost-Benefit Analysis

The estimates of the program's effect on monthly earnings can be used to compare the total costs of implementing the training courses to the additional earnings gains they generated. As our measure of the return to training we use the standard concept of internal rate of return of an investment. Let B_{t+s} the flow of earnings gains in each period and let C_t be the total cost of investment. Assuming the cost is incurred in one period and that the investment generates benefits for T years, the internal rate of return of the investment is given by the rate r that equalizes the present discounted value of net benefits (NPV) to zero:

$$NPV = -C_t + \sum_{s=1}^{T} \frac{B_{t+s}}{(1+r)^s} = 0 \qquad (6)$$

Let $C_t = C_t^d + F_t$ be the total cost of training, where C_t^d is the direct cost of training and F_t the forgone cost of training (Mincer 1989). The rich administrative data allows us to account for all direct costs incurred in the program including payments to the training centers, stipends-subsidies and medical insurance received by trainees, and administrative costs incurred by the training operator. Because the PROJOVEN program is a three-month full-time activity for beneficiaries, they forgone the monthly wage they would have earned had they entered or remain in the labor market without training. Thus, the forgone costs of training are computed as $F_t = 3\gamma W_{t-1}$, where W_{t-1} is the pre-treatment average monthly earnings conditional on working and γ is the probability of being employed.

The administrative data show that the direct cost of training is $515 dollars per trainee whereas the estimated forgone earnings reach $100 dollars for men and $50 dollars for women.[19] These numbers reveal that the direct costs of training are substantial components of the total costs (see Mincer 1989, Almeida and Carneiro 2005), and they vary monotonically across quartiles of quality, ranging from $475 dollars per trainee (quartile 1) to $572 dollars (quartile 4).

When computing the benefits of the investment in training we acknowledge that is the stock and not the flow of training which determines current earnings. It implies that training increases earnings contemporaneously but also in future periods. Yet, human capital stock acquired while participating in training depreciates as knowledge and skills becomes obsolete and workers forget past learning (Becker 1993). Therefore, we assume

[19] All figures in real values of December 2001. The exchange rate (dollar/sol) was 3.4.

that the stock of training depreciates at a rate ρ per year. Our estimates suggest that ρ ranges between 10 and 25 percent in the second year after the program.

Further assumptions are needed to compute the yearly benefits of the program. First, we assume that tenure is 11 months for men and 10 months for women. Second, the stipends-subsidies received by trainees are also computed as benefits in the first year. Finally, we assume that the benefits are spread over a 40-year period. Therefore, the annual benefits of the program in year *s* is estimated using the formula $B_{t+s} = (\Delta W_t / (1+\rho)^s)\delta$, where ΔW_t is the post-treatment earnings gains and δ represents tenure. We take the OLS treatment impacts (Table 5), which are the lower bound estimates, as a benchmark. Because the treatment impacts are computed using censored real monthly earnings variable that assigns zero earnings to non-working individuals, this effect gives the combined gain in earnings and employment.

Four results emerge from this analysis. First, the internal rate of return for the whole sample ranges from 14 percent (ρ=0.25) to 25 percent (ρ=0.10), which is above the average interest rate on government debt in Peru during the period of analysis. These returns may underestimate the true internal rate of return since we do not consider potential gains on other outcomes (e.g., education attainment).

Second, as expected, there is considerable heterogeneity of the net gains by gender. Women show quite high returns to training (32 to 43 percent) whereas men show low returns (0 to 6 percent). These estimates are in line with Attanasio, Kugler, and Meghir (2007) that find net gains for women but not for men participating in a similar training program in Colombia.

Third, when considering the quality of the training services, we find large and nearly similar returns for women attending low- and high-quality training courses (24 versus 19 percent for ρ=0.25 and 35 versus 30 percent for ρ =0.10). For men, on the other hand, we find positive returns (below10 percent) only for those attending high-quality courses. Finally, for the whole sample, it takes a relatively short time (5 years) to generate positive returns. This time span is shorter for women (3 years) than for men (8 years).

9. Conclusions

The adoption of market-based approaches that ensure both quality and pertinence in the provision of training services has been shown to effectively increase the earnings of disadvantaged individuals, who frequently emerge from public schools operating far from any efficient frontier (Glewwe 2002). The large treatment effects are mainly explained by the program's demand-driven approach that has the ability to relocate individuals from low-quality jobs to productive ones in firms protected by international laws that guarantee minimum work conditions.

The entire set of positive training impacts is largely determined by the performance of female beneficiaries, who demonstrate much larger treatment effects than male participants. Consequently, the program generates large net gains for women, making the program highly cost-effective. For men, on the other hand, the program shows nearly zero returns. These findings suggest that the PROJOVEN dollars are misallocated to the extent that the program operator must deny services to eligible males to reserve the funds for additional female beneficiaries.

We also find strong heterogeneity of the treatment impacts by considering the quality of the training services. Overall, individuals attending high-quality training courses show higher impacts than those attending low-quality courses. These earnings-quality relationships are stronger for males rather than females. For men, we find net gains only for those attending high-quality courses, whereas women attending low- and high-quality courses show large and somewhat similar net gains.

This paper also shows that individuals who complete both the formal training and the on-the-job training experience have much higher earnings than individuals who complete only the formal training. In fact, the returns to the formal training are modest and consistent with the findings in the literature on training programs. In addition, the earnings differentials between people attending high- and low-quality courses are higher for the subsample of individuals who participate only in the first stage as compared to those for the whole sample. This result indicates that the second stage of the program, the on-the-job training experience, smooths productivity gains between people attending courses of varying quality. This in turn may explain why the earnings-quality relationships are stronger for men than for women given the larger proportion of women that complete the program. Therefore, a policy implication that follows from this result is that on-the-job training experience may mitigate the outcomes of low-quality training services and constitute the most effective way of helping disadvantaged individuals.

This favorable assessment of the PROJOVEN program should be tempered by the existence of a trade-off between the costs and the potential coverage of this program. In fact, the large costs associated to this program prevent a large-scale expansion of its operations and thus the aggregate impact on the youth unemployment problem is very limited. Finally, the reader should bear in mind that the quality premiums observed in this paper are based on a sample of training institutions that pass a minimum quality threshold. It is important to consider what the magnitude of these earnings differentials would be if training institutions located below the cut-off point were included.

References

Almeida, R., P. Carneiro. 2005. "The Internal Rate of Return to On-the-Job Training", Manuscript.

Ashenfelter, O. 1978. "Estimating the Effect of Training on Earnings." *Review of Economics and Statistics* 60: 47-57.

Attanasio O., A. Kugler, and C. Meghir. 2007. "Effects of Youth Training in Developing Countries: Evidence from a Randomized Training Program in Colombia". Manuscript.

Becker, G. 1962. "Investment in Human Beings". NBER, Special Conference 15, *Journal of Political Economy,* 70 (S9:S49).

Becker, G. 1993. Human Capital. Columbia University Press, Third Edition.

Behrman, J., Y. Cheng, and P. Todd. 2004. "Evaluating Pre-School Programs when Length of Exposure to the Program Varies: A Nonparametric Approach." *Review of Economics and Statistics* 86(1): 108-132.

Behrman, J., N. Birdsall. 1983. "The Quality of Schooling: Quantity Alone is Misleading", *American Economic Review*, Vol 73, No 5 (928-946)

Black, D. and J. Smith. 2003. "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics* 121(1): 99-124.

-----. 2005. "Estimating the Returns to College Quality with Multiple Proxies for Quality." Journal of Labor Economics, forthcoming.

Campa, J. 1997. "Public Sector Retrenchment: Spain in the 1980s", Department of Economics, New York University.

Card, D. and A. Krueger. 1996. "Labor market Effects of School Quality: Theory and Evidence", NBER, WP 5450.

Card, D. and A. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100(1): 1-40.

Carnoy, M. 2001. "School Vouchers: Examining the Evidence", Economic Policy Institute, Washington D.C.

Dehejia R. and S. Wahba. 1999. "Causal effects in Non-Experimental Studies: Re-evaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94: 1053-1062.

Dale, S. and A. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of the Selection on Observables and Unobservables." *Quarterly Journal of Economics* 117 (4): 1491-1528.

Fan, J. 1992. "Design-Adaptive Nonparametric Regression." *Journal of the American Statistical Association* 87: 998-1004.

Filmer D. and L. Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data-or Tears: An application to educational Enrollment in States of India". *Demography*, 38 (1): 115-132.

Frölich, M. 2004. "Finite Sample Properties of Propensity-Score Matching and Weighting Estimators" *The Review of Economics and Statistics*, 86(1): 77-90.

Fuller, B. 1987. "What School Factors Raise Achievement in the Third World." *Review of Education Research* 57(3): 255-292.

Galdo, J. 1998. "La Evaluación de Proyectos de Inversión Social: Impacto del Programa de Capacitación Laboral Juvenil PROJOVEN." *Boletín de Economía Laboral* 9, Ministerio de Trabajo y Promoción Social.

Galdo, J., D. Black, and J. Smith. 2007. "Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data." Manuscript.

Glewwe, P. 2002. "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes." *Journal of Economic Literature* XL: 436-482.

Hanushek, E. 1986. “The Economics of Schooling: Production Function and Efficiency in Public Schools.” *Journal of Economic Literature* 24(3): 1141-1177.

Hanushek, E., Kain, J., O’Brien, D., and Rivkin, S. 2005. “The Market for Teacher Quality”, Manuscript.

Harbison, R. and E. Hanushek. 1992. *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. New York: Oxford University Press/World Bank.

Heckman, J. 2001. “Micro data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture.” *Journal of Political Economy* 109: 673-748.

Heckman, J., H. Ichimura, and P. Todd. 1997. “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme.” *Review of Economics Studies* 64(4): 605-654.

Heckman, J., J. Smith, and N. Clement. 1997. “Making the Most Out of Program Evaluations and Social Experiments: Accounting for Heterogeneity in Program Impacts.” *Review of Economic Studies* 64: 421-471.

Heckman, J., H.R. LaLonde, and J. Smith. 1999. “The Economics and Econometrics of Active Labor Programs.” In: O. Ashenfelter and D. Card, eds. *Handbook of Labor Economics, Volume 3A*. Amsterdam: North Holland, pp. 1865-2097.

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

Heckman, J., A. Layne-Ferrar, , and P. Todd. 1995. “Does Measured School Quality Really Matter? An Examination of the Earnings Quality Relationship, NBER, WP 5274

Heckman, J., and P. Todd, 1995. “Adapting Propensity Score Matching and Selection Models to Choice-Based Samples”. Manuscript.

International Labor Organization. 2003. “Youth Training and Employment”. CINTERFOR/ILO. Geneva.

LaLonde, R. 1986. “Evaluating the Econometric Evaluation of Training Programs with Experimental Data”, *The American Economics Review*, 76(4) , 604-620

Lillard, L., and Tan H. 1986. “Training: Who Gets It and What are Its Effects on Employment and Earnings? RAND Corporation

Mincer, J. 1989. “Job Training: Costs, Returns, and Wage Profiles”, NBER WP, 3208.

Mincer, J. 1962. “On-the-Job Training: Costs, Returns, and Some Implications”, *Journal of Political Economy*, 70 (S50-S79).

Ñopo, H., J. Galdo. 1997. “Un Algoritmo de Adjudicacion de los Cursos de PROJOVEN”, Manuscript, Ministry of Labor, Lima, Peru.

Ñopo, H., J. Saavedra, and M. Robles. 2001. “Una Medición del Impacto del Programa de Capacitación Laboral Juvenil PROJOVEN.” Lima, Peru: Group for the Analysis of Development (GRADE).

Pagan, A., A. Ullah. 1999. *Nonparametric Econometrics*. Cambridge University Press.

Parsons, D. 1986. “The Employment Relationship”, In Handbook of Labor Economics, Vol 2, Eds. Ashenfelter and Layard. Amsterdam: Elsevier Science Publishers.

Rosenbaum, P. and D. Rubin. 1983. “The Central Role of the Propensity Score In Observational Studies For Causal Effects.” *Biometrika* 70(1): 41-55.

Smith, J. and P. Todd. 2005. “Does Matching Overcome LaLonde’s Critique of Non-Experimental Estimators? *Journal of Econometrics* 125(1-2): 305-353.

World Bank. 2004. “Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries”, Social Protection Discussion Paper Series #0402.

Appendix 1A: The PROJOVEN's Procurement Competition

The PROJOVEN's design competition is based on a model of two-dimensional auctions in which each training firm bids on both quality and price and bids are selected following a classical optimization process (see Ñopo and Galdo 1997). Let *X* be the set of competing training courses. Let training course "y" be described by $y = (y_1, y_2)$, where y_1 represents prices and y_2 the promised quality and let training course "x" be represented by $x = (x_1, x_2)$. We define the superiority of "y" over "x" if and only if $y_1 < x_1$ and $y_2 > x_2$, which is denoted by $y \succ x$. We define the set of corresponding superiors for each training course by

$$\begin{aligned} &CSup : X \to \ell(X) \\ &\qquad x \mapsto CSup(x) = \{y \in X / y \succ x\} \end{aligned}$$

The design competition relies on the premise that if a training course "x" is selected to receive a contract then its set of corresponding superiors must be also selected, provide two restrictions. First, the program operator targets at least *Q* training slots that is determined *ex-ante*, which defines a quantity restriction $Q : X \to R, x \mapsto Q(x) \geq Q$

In addition, the program operator has a limited public funding (*F*), which lead us to define a budget restriction $F : X \to R, x \mapsto F(x) = \sum_{y \in CSup(x)} y_2 \leq F$

Thus, the selection of the training courses follows a maximization problem:

$$\begin{aligned} &\underset{x \in X}{Max} \quad \sum_{y \in CSup(x)} y_1 \Big/ \sum_{y \in CSup(x)} y_2 \\ &s.t. \\ : \quad &\quad \sum_{y \in CSup(x)} Q(y) \geq Q \\ &\quad \sum_{y \in CSup(x)} y_2 \leq F \end{aligned}$$

Appendix A2: Coefficient Estimates from Balanced Logit Models for Program Participation
PROJOVEN, Lima 1996-2003

	Men		Women	
	coeff.	std. error	coeff.	std. error
age	0.052	0.031	0.131	0.029
schooling				
incomplete primary	0.583	0.621	1.441	0.760
incomplete high school	0.360	0.319	0.557	0.306
complete high school	0.776	0.558	-0.756	0.330
marital status				
single	-15.524	0.707	-0.174	0.691
married and/or cohabitating	-16.410	0.851	-0.603	0.627
number of children	-0.458	0.251	-0.353	0.123
work status				
have a job	-0.972	0.474	-0.416	0.330
unemployed	-0.224	0.215	-0.238	0.152
kind of work				
self-employed	1.748	0.510	0.937	0.366
worker in private sector	1.828	0.490	0.795	0.360
unpaid family worker/housekeeper	1.777	0.417	0.975	0.297
monthly earnings	-0.003	0.001	-0.004	0.001
participation in training courses	-0.236	0.204	-0.051	0.159
hours of training	0.002	0.001	0.000	0.000
family size	-0.074	0.078	0.007	0.024
number of rooms/ number of persons	0.005	0.131	0.113	0.037
floor: high-quality materials	0.275	0.139	0.365	0.125
ceiling: high-quality materials	0.516	0.155	0.396	0.137
walls: high-quality materials	-0.125	0.158	-0.096	0.134
flush toilet in the house	0.157	0.137	0.263	0.117
#	1111		1494	

Figure 1A: Training and Earnings Over Time

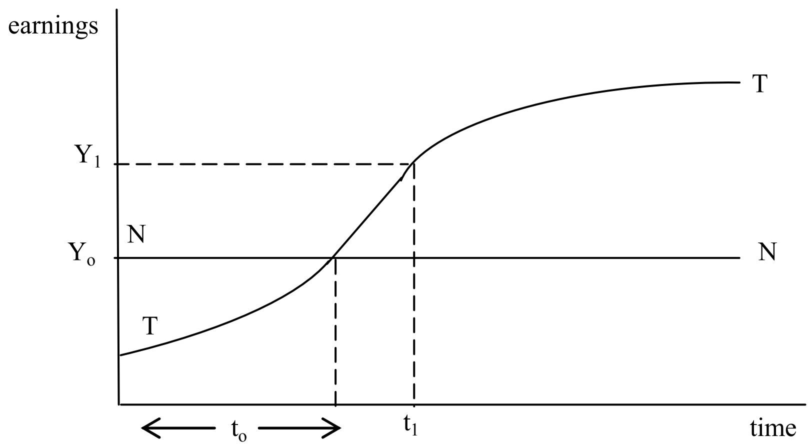


Figure 1B: Training and Earnings Over Time

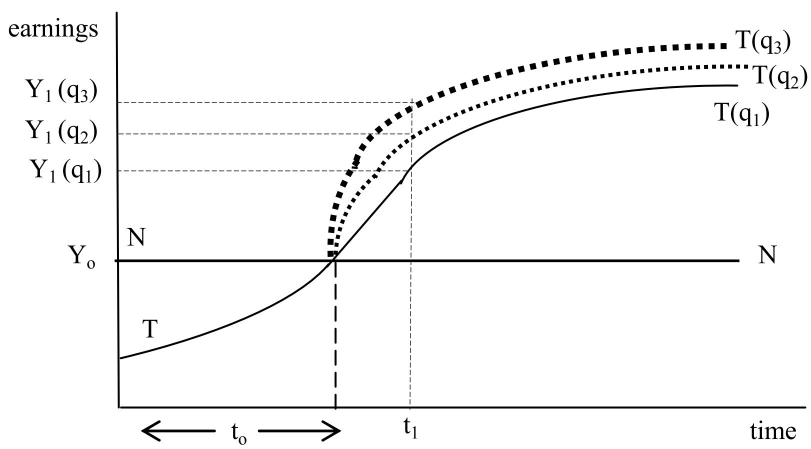


Figure 2. Beneficiary Selection Process
PROJOVEN, Lima 1996 to 2003.

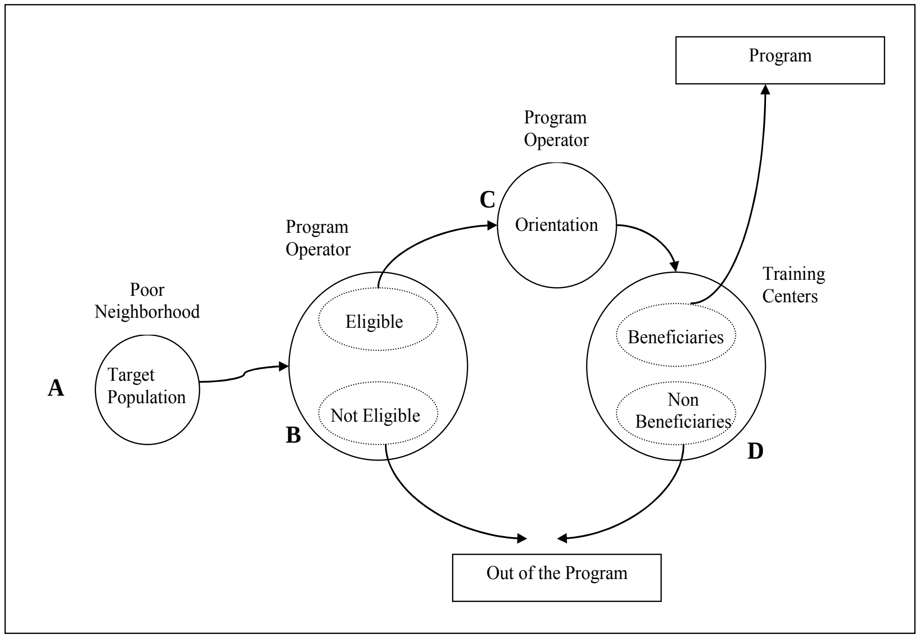
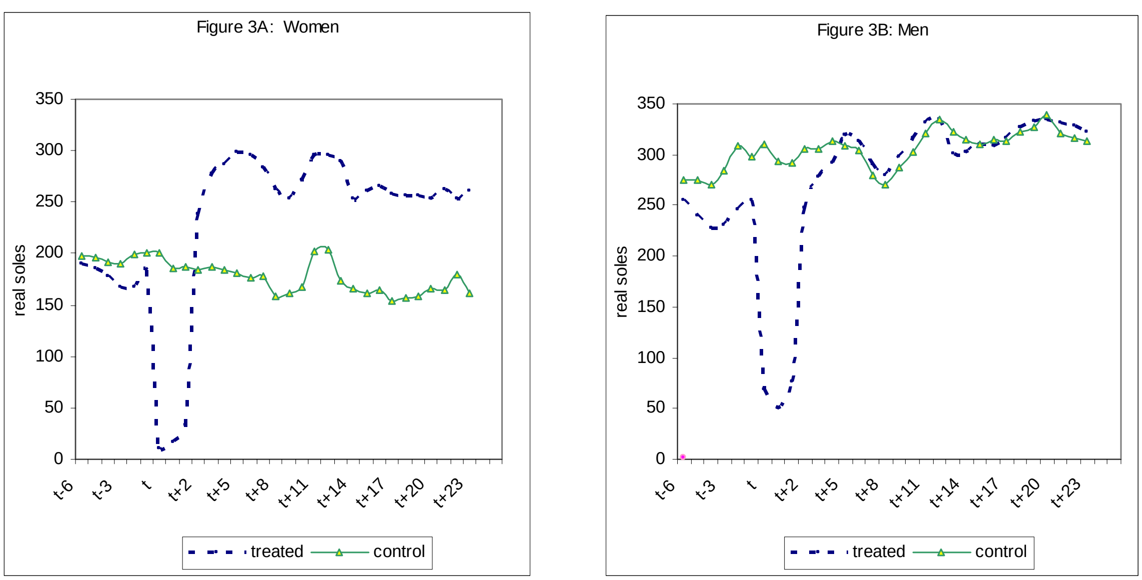


Figure 3. Unconditional Monthly Earnings over Time
PROJOVEN, Lima 1996 to 2003.



Note: Pooled means are unweighted and based on unbalanced panel data.

Table 1. Standardized Scores for Multiple Quality Proxies
PROJOVEN, Lima 1996 to 2003

	Cohort 1		Cohort 2		Cohort 3		Cohort 4	
	mean	std. dev.	Mean	std. dev.	mean	std. dev.	mean	std. dev.
Class size	0.28	0.20	0.35	0.26	0.41	0.23	0.43	0.29
Expenditures/trainees	0.39	0.23	0.39	0.26	0.55	0.20	0.48	0.23
Human Resources	0.72	0.25	0.57	0.24	0.64	0.22	0.65	0.18
Infrastructure	0.85	0.27	0.94	0.19	0.97	0.15	0.95	0.16
Equipment	0.54	0.26	0.67	0.30	0.65	0.21	0.81	0.16
Curricular Structure	0.56	0.28	0.85	0.26	0.78	0.24	0.69	0.29
Market Knowledge	0.74	0.22	0.68	0.21	0.83	0.15	0.71	0.19
% explained by PCA	42		30		24		20	
# institutions	30		33		35		33	
# courses	154		158		215		204	
# funded courses	75		98		118		148	

Notes: The available bidding data have aggregate scores for each category. The scores are normalized as the ratio of the difference between the raw indicator value and the minimum value divided by the range. All normalized proxies are between 0 and 1. The quality index is constructed by principal component analysis based on the first factor.

Table 2: Summary Statistics
PROJOVEN, Lima 1996-2003

							Training Quality Distribution					
	Males			Females			Males			Females		
	treated	comparison	p-value	treated	comparison	p-value	treated >50th	treated <50th	p-value	treated >50th	treated <50th	p-value
A. Socio-Demographic												
age	19.86	19.87	0.920	19.95	20.07	0.355	19.94	19.76	0.363	20.04	19.87	0.321
schooling (%)												
incomplete primary	1.58	1.22	0.993	0.91	0.38	0.196	0.32	3.20	0.006	0.59	1.18	0.396
complete primary	5.43	6.94	0.602	4.55	6.79	0.056	4.43	6.80	0.219	4.42	4.72	0.848
incomplete high school	11.56	10.24	0.287	8.18	7.30	0.515	12.34	10.80	0.571	6.78	9.20	0.225
complete high school	81.09	81.25	0.475	86.11	85.02	0.543	82.59	78.80	0.254	88.20	84.43	0.135
marital status (%)												
single	94.22	88.75	0.000	87.56	65.73	0.000	94.94	93.60	0.619	87.94	87.06	0.714
married and/or cohabitating	3.34	8.82	0.000	8.94	25.96	0.000	0.64	4.00	0.502	10.00	8.13	0.821
number of children	1.09	1.21	0.152	1.28	1.33	0.376	1.17	1.34	0.209	1.22	1.33	0.291
B. Labor information												
work status (%)												
have a job	62.35	62.63	0.921	44.95	44.76	0.939	61.71	63.20	0.716	45.59	44.00	0.661
unemployed	21.89	21.97	0.973	25.65	27.11	0.513	22.15	21.20	0.785	22.65	28.00	0.092
out of labor force	15.76	15.40	0.865	29.40	28.13	0.580	16.14	15.60	0.862	31.76	28.00	0.257
kind of work (%) (current job)												
self-employed	11.56	12.28	0.704	11.40	12.02	0.703	10.13	13.20	0.255	12.06	10.82	0.593
worker in private sector	39.58	42.74	0.277	16.58	20.59	0.042	40.19	38.80	0.737	17.06	15.53	0.569
worker in public sector	0.18	0.35	0.570	0.52	0.51	0.895	0.32	0.00	0.374	0.59	0.48	0.822
unpaid family worker	15.59	7.79	0.000	16.19	8.70	0.000	15.19	16.40	0.695	14.71	17.65	0.274
housekeeper	0.53	0.52	0.988	5.44	4.60	0.450	0.32	0.40	0.510	5.88	5.18	0.670
monthly earnings (current job)	132.37	177.90	0.000	62.21	90.02	0.000	134.08	130.03	0.762	62.21	61.01	0.883
training courses before Projoven	23.81	19.72	0.092	26.62	25.38	0.578	26.58	20.80	0.110	28.02	25.70	0.473
hours of training	71.29	36.27	0.000	64.96	61.74	0.734	82.22	58.90	0.202	67.66	63.75	0.790
C. Household characteristics												
family size	6.01	6.11	0.469	6.36	5.88	0.000	6.05	5.99	0.805	6.25	6.49	0.231
household members/rooms	2.92	2.82	0.283	3.22	2.98	0.004	2.94	2.90	0.869	3.14	3.29	0.255
floor: high quality materials	43.38	35.19	0.005	40.77	31.45	0.002	40.20	46.86	0.121	42.19	39.85	0.526
ceiling: high-quality materials	36.40	25.57	0.000	37.07	24.81	0.000	35.55	37.24	0.685	38.44	35.89	0.481
walls: high-quality materials	68.01	65.63	0.396	68.36	61.87	0.008	68.77	66.53	0.580	68.13	68.49	0.917
drinking water piped into house	76.71	78.61	0.445	75.10	74.33	0.728	75.08	79.25	0.250	75.90	74.51	0.663
flush toilet in the house	65.44	60.17	0.068	61.42	52.69	0.000	66.11	65.69	0.918	62.19	61.14	0.773
D. Parent's schooling												
father (%)												
primary	31.52	35.19	0.246	41.29	33.82	0.007	32.79	30.21	0.566	42.16	41.46	0.863
incomplete high school	26.08	19.82	0.026	17.08	20.59	0.118	27.46	25.00	0.564	14.55	19.21	0.133
complete high school	23.81	28.06	0.148	27.20	30.07	0.269	23.36	25.00	0.691	29.10	25.61	0.340
higher education	9.75	5.79	0.027	7.46	5.72	0.220	9.84	9.90	0.980	6.72	8.23	0.486
mother (%)												
primary	43.31	40.09	0.330	50.08	41.99	0.000	48.36	38.02	0.030	49.25	51.22	0.633
incomplete high school	21.09	20.94	0.955	18.08	18.46	0.861	20.08	22.92	0.474	19.03	17.68	0.672
complete high school	19.27	18.71	0.829	15.09	21.08	0.006	18.85	19.79	0.805	13.43	16.46	0.304
higher education	4.54	2.67	0.136	4.15	4.08	0.957	3.69	5.73	0.313	4.85	3.66	0.471
#	571	578		772	782		316	250		340	425	

Notes: Pooled means are unweighted. p-values refers to the test for differences in means for two different samples.

Table 3. Average Treatment Impacts on Monthly Earnings
Difference-in-Differences Matching Estimator
PROJOVEN, Lima 1996 to 2003

					Pooled Data		
	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Men	Women	All
A. 6 months after program							
$\Delta = Y_1(q) - Y_0(0)$	115 (39)	39 (26)	32 (25)	60 (22)	35 (20)	84 (15)	61 (13)
	[90]	[31]	[25]	[47]	[20]	[93]	[48]
$\Delta = Y_1(q_4) - Y_0(0)$	195 (43)	37 (47)	38 (53)	33 (42)	45 (32)	72 (34)	69 (25)
	[154]	[29]	[30]	[26]	[26]	[80]	[54]
$\Delta = Y_1(q_1) - Y_0(0)$	92 (51)	10 (44)	38 (56)	45 (29)	12 (34)	76 (24)	48 (21)
	[72]	[8]	[30]	[35]	[7]	[84]	[38]
B. 12 months after program							
$\Delta = Y_1(q) - Y_0(0)$	21 (27)	89 (31)	30 (30)	21 (24)	8 (22)	67 (16)	38 (15)
	[17]	[70]	[24]	[17]	[5]	[74]	[30]
$\Delta = Y_1(q_4) - Y_0(0)$	72 (44)	78 (53)	36 (56)	13 (43)	27 (35)	49 (31)	44 (25)
	[57]	[61]	[28]	[12]	[16]	[54]	[35]
$\Delta = Y_1(q_1) - Y_0(0)$	-56 (40)	74 (47)	50 (54)	-4 (42)	-8 (38)	57 (25)	17 (22)
	[-44]	[58]	[39]	[3]	[-5]	[63]	[13]
C. 18 months after program							
$\Delta = Y_1(q) - Y_0(0)$	42 (45)	88 (43)	47 (28)	54 (33)	12 (25)	106 (27)	69 (19)
	[33]	[69]	[37]	[43]	[7]	[117]	[54]
$\Delta = Y_1(q_4) - Y_0(0)$	115 (61)	12 (60)	60 (51)	49 (50)	37 (37)	83 (59)	78 (33)
	[90]	[22]	[47]	[39]	[22]	[92]	[61]
$\Delta = Y_1(q_1) - Y_0(0)$	8 (62)	55 (61)	58 (47)	36 (42)	21 (42)	110 (28)	43 (26)
	[6]	[43]	[45]	[28]	[12]	[122]	[34]

Notes: Point estimates are in real soles. Bootstrapped standard errors based on 500 replications are in parentheses. Percentage gains with respect to earnings in the baseline period are in brackets. q_4 and q_1 are the top and bottom quartiles of the quality index distribution. The propensity scores are estimated using a logit model. The matching variable is the log of the odd-ratio. We use Epanechnikov kernel function with the bandwidths determined by weighted cross-validation. The optimal bandwidths for the whole sample are 0.4, 0.5, and 0.4 when estimating the treatment impacts 6, 12, and 18 months after the program; 0.3, 0.2, and 0.3 for men; and 0.6, 0.5, and 0.6 for women. These data include 989, 931, and 918 males and 1342, 1279, and 1255 females 6, 12, and 18 months after the program.

Table 4. Marginal Treatment Impacts on Monthly Earnings
Difference-in-Differences Matching Estimator
PROJOVEN, Lima 1996 to 2003

	Men	Women	All
A. 6 months after program			
$\Delta = Y_1(q_4) - Y_1(q_1)$	130 (43) [76]	5 (32) [6]	62 (26) [49]
$\Delta = Y_1(q_3) - Y_0(q_1)$	43 (41) [25]	-4 (28) [-4]	8 (21) [6]
$\Delta = Y_1(q_2) - Y_0(q_1)$	19 (51) [11]	18 (24) [20]	19 (21) [15]
B. 12 months after program			
$\Delta = Y_1(q_4) - Y_1(q_1)$	61 (49) [36]	-16 (21) [-18]	29 (28) [23]
$\Delta = Y_1(q_3) - Y_0(q_1)$	22 (45) [13]	22 (25) [24]	22 (23) [17]
$\Delta = Y_1(q_2) - Y_0(q_1)$	-15 (63) [-8]	21 (24) [23]	16 (24) [13]
C. 18 months after program			
$\Delta = Y_1(q_4) - Y_1(q_1)$	115 (43) [67]	-38 (27) [-42]	-2 (33) [-2]
$\Delta = Y_1(q_3) - Y_0(q_1)$	62 (53) [36]	-2 (32) [-2]	2 (25) [2]
$\Delta = Y_1(q_2) - Y_0(q_1)$	19 (51) [11]	3 (28) [3]	-8 (23) [-6]

Notes: Point estimates are in real soles. Bootstrapped standard errors based on 500 replications are in parentheses. Percentage gains with respect to earnings in the baseline period are in brackets. q_4, q_3, q_2, and $q_1$are the fourth, third, and first quartiles of the quality index distribution. The propensity scores are estimated using a logit model. The matching variable is the log of the odd-ratio. We use Epanechnikov kernel function with the bandwidths determined by weighted cross-validation. The optimal bandwidths for the whole sample are (from above to below): 0.5, 0.7, 0.8, 3.2, 1.8, 1.4, 0.5, 0.3, and 0.5.

Table 5. Treatment Impacts on Monthly Earnings
Parametric Least Square Estimator
PROJOVEN, Lima 1996 to 2003

	Men	Women	All
A. 6 months after program			
$\Delta = Y_1(q) - Y_0(0)$	23 (15) [13]	80 (11) [89]	53 (10) [42]
$\Delta = Y_1(q_4) - Y_0(0)$	32 (22) [19]	67 (19) [74]	50 (15) [39]
$\Delta = Y_1(q_1) - Y_0(0)$	8 (24) [5]	62 (16) [69]	37 (14) [29]
B. 12 months after program			
$\Delta = Y_1(q) - Y_0(0)$	-5 (16) [-3]	70 (12) [78]	30 (10) [24]
$\Delta = Y_1(q_4) - Y_0(0)$	5 (23) [3]	50 (20) [56]	23 (16) [18]
$\Delta = Y_1(q_1) - Y_0(0)$	-25 (25) [-15]	47 (17) [52]	7 (15) [6]
C. 18 months after program			
$\Delta = Y_1(q) - Y_0(0)$	-2 (16) [-1]	86 (15) [95]	42 (12) [33]
$\Delta = Y_1(q_4) - Y_0(0)$	16 (25) [9]	63 (25) [70]	41 (18) [33]
$\Delta = Y_1(q_1) - Y_0(0)$	-1 (27) [-1]	81 (21) [90]	39 (17) [31]

Notes: Point estimates are in real soles. Percentage gains with respect to earnings in the baseline period are in brackets and standard errors in parentheses. The estimator is applied to the sample of individuals inside the overlapping support region. q_4 and q_1 are the fourth and bottom quartiles of the quality index. The parametric specification includes as regressors age, education, sex, marital status, pre-treatment earnings, whether has children, number of children, whether participate in previous training, house infrastructure (floor, ceiling, and walls), whether has access to flush toilet, and the estimated propensity score. Also, it considers dummy variables for having attended a course in the fourth, third, second, and first quartile of the quality distribution. The control group indicator is the omitted group. There are 989, 931, and 918 men and 1342, 1279, and 1255 women 6, 12, and 18 months after the program.

Table 6. Average Treatment Impacts for Formal Training
Difference-in-Differences Matching Estimator
PROJOVEN, Lima 1996 to 2003

	Men		Women		All	
	Matching	OLS	Matching	OLS	Matching	OLS
A. 6 months after program						
$\Delta = Y_1(q) - Y_0(0)$	4 (25)	-9 (18)	29 (20)	23 (13)	14 (17)	5 (11)
	[2]	[-5]	[32]	[26]	[11]	[4]
$\Delta = Y_1(q_4) - Y_0(0)$	25 (48)	3 (30)	77 (46)	54 (29)	54 (36)	26 (21)
	[15]	[2]	[86]	[60]	[43]	[20]
$\Delta = Y_1(q_1) - Y_0(0)$	-34 (47)	-38 (32)	15 (37)	-1 (23)	-10 (30)	-18 (19)
	[-20]	[22]	[17]	[-1]	[-8]	[-14]
B. 12 months after program						
$\Delta = Y_1(q) - Y_0(0)$	-14 (28)	-43 (20)	28 (23)	35 (16)	6 (19)	-7 (12)
	[-8]	[25]	[31]	[39]	[5]	[-6]
$\Delta = Y_1(q_4) - Y_0(0)$	53 (55)	16 (32)	19 (50)	3 (33)	47 (39)	13 (23)
	[31]	[9]	[21]	[3]	[37]	[10]
$\Delta = Y_1(q_1) - Y_0(0)$	-87 (51)	-100 (34)	26 (43)	18 (26)	-33 (33)	-43 (21)
	[-51]	[-58]	[29]	[20]	[-26]	[-34]
C. 18 months after program						
$\Delta = Y_1(q) - Y_0(0)$	16 (32)	-20 (21)	67 (39)	72 (21)	33 (22)	25 (15)
	[9]	[-12]	[70]	[80]	[26]	[20]
$\Delta = Y_1(q_4) - Y_0(0)$	59 (53)	29 (35)	40 (48)	42 (44)	70 (38)	36 (27)
	[35]	[17]	[44]	[47]	[55]	[28]
$\Delta = Y_1(q_1) - Y_0(0)$	-16 (60)	-40 (37)	106 (54)	85 (34)	33 (36)	28 (25)
	[-9]	[-23]	[117]	[94]	[26]	[22]

Notes: Point estimates are in real soles. Bootstrapped standard errors based on 500 replications are in parentheses. Percentage gains with respect to earnings in the baseline period are in brackets. Both difference-in-differences matching and OLS estimates are applied to the sample of individuals inside the overlapping support region. q_4 and q_1 are the top and bottom quartiles of the quality index distribution within the subsample of individuals that complete the formal training. The parametric specification includes as regressors age, education, sex, marital status, pre-treatment earnings, whether has children, number of children, whether participate in previous training, family members/rooms, house infrastructure (floor, ceiling, and walls), whether has access to flush toilet, and the estimated propensity score. Also, it considers dummy variables for having attended a course in the fourth, third, second, and first quartile of the quality distribution. The control group indicator is the omitted group. We use Epanechnikov kernel function for the matching estimates with the bandwidths determined by weighted cross-validation. The matching variable is the log of the odd ratio. There are 989, 931, and 918 men and 1342, 1279, and 1255 women 6, 12, and 18 months after the program.