Inter-American Development Bank
Banco Interamericano de Desarrollo (BID)
Research Department
Departamento de Investigación
Working Paper #555

Training Quality and Earnings: The Effects of Competition on the Provision of Public-Sponsored Training Programs

by

Alberto Chong*
José Galdo**

*Inter-American Development Bank, Washington, D.C.
**McMaster University/IZA

Original June 2006
Revised August 2006

Abstract[*]

This paper evaluates the effectiveness of market-based approaches in 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 that operate for profit 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. These earnings differentials are larger for females rather than males, and are larger in the medium term rather than in the short run. External validity was assessed by using five different cohorts of individuals over an eight-year period.

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

[*] Chong: Research Department, Inter.-American Development Bank; Galdo: McMaster University and IZA; Corresponding author: Galdo E-mail: galdojo@mcmaster.ca. Dan Black, Paul Glewwe, Jeffrey Kubik, Oscar Mitnik, Hugo Ñopo, and Enrico Rettore provided valuable comments. We also thank seminar participants at the 2006 European Econometric Society Meeting in Vienna, Austria and the 2006 Northeast Universities Development Consortium Meeting at Cornell University. Arturo Garcia provided excellent research assistance. The views and opinions in this paper should not be attributed to the Inter-American Development Bank or its Executive Directors. 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. This has been particularly true in recent years when massive privatization processes and dramatic personnel reductions in overstaffed public sectors have driven a large fraction of workers into the unemployment ranks or the underground economy. 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. In fact, 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., Bitler, Gelbach, and Hoynes, 2004). 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 public-sponsored training programs and beneficiaries' subsequent labor market earnings. To our knowledge, this is the first paper that addresses quality issues in training programs, with the added advantage that we are able to address the role of market-based approaches in the provision of public-sponsored training services. In fact, the selection of training courses relies on formal bidding processes in which public and private training institutions

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

compete for limited public funding.[2] The detailed bidding questionnaires and instruments not only allow us to identify common proxies for quality such as expenditures per student, class size, infrastructure, equipment, and teacher characteristics, but they also provide information about curricular structure, such as the consistency of goals, contents, and activities. Moreover, this bidding information allows us to use disaggregated data at the course level, rather than at the school o state level, which may improve the explanation for quality heterogeneity (Hanushek, Kain, O'Brien, and Rivkin, 2005). Furthermore, the availability of data for five different cohorts of individuals over an eight-year period (1996 to 2003) allows us to consider the robustness of our estimates 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 35,000 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 relative 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. Hence, this unique data design allows us to examine the effectiveness of market-based approaches in the selection of pertinent training services.

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 unobservables. Indeed, both the standardized quality scores based on bidding data and the unique evaluation framework allow us to minimize two crucial problems frequently encountered in the literature, data limitations and econometric

[2]Similar programs have been implemented since the mid-1990s in Chile (Chile Joven), Argentina (Proyecto Joven), Colombia (Youth Training Program), and Uruguay (Youth Training Program).

problems, and to provide alternative measures to typical point estimates which have been highly criticized (Glewwe and Kremer, 2005; Glewwe, 2002). 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 the extent that socioeconomic variables and family background raise or lower earnings for all levels of training attainment, we control our estimates from differences in observables arising from demographic, socioeconomic, and parental variables. 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 marginal matching estimator that assumes that sorting into different quality training courses arises from both observables and unobservables. An advantage of this estimator is that it 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 four conclusions. First, we find strong evidence about the effectiveness of market-based approaches in the provision of training services. In fact, the combination of bidding processes with demand-driven approaches, which ensures both quality and pertinence of the training courses yield larger overall point estimates than those reported in the literature. This result is particularly robust for females who show much higher treatment impacts than male participants.

Second, we find evidence of substantial heterogeneity in response to training quality. Our main result is anticipated in Figure 1, which depicts the beneficiaries' earnings along a quality-index grid. In general, individuals attending high-quality training courses show much higher labor earnings than individuals attending low-quality courses. These earning differentials are larger in the medium-term than in the short-run, which is explained by a sharp decrease in the medium-term earnings of individuals attending low-quality courses. Our difference-in-differences models estimate that the effect suggested by Figure 1 corresponds to a differential of 32 percent in the earnings of beneficiaries attending high- and low-quality courses 18 months after the program.

Third, this paper also shows that individuals who complete both formal training and on-the-job training experience have 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. Fourth, there is evidence of Ashenfelter's dip in this program. Parametric estimators that are consistent when the model of program participation stipulates pre-program earnings dip show that our results are robust to alternative identifying assumptions.

The remainder of the paper is organized as follows. In section 2 we develop the economics of training quality. In section 3 we provide an overview of the PROJOVEN program. We then discuss the measurement of training quality in section 4. In section 5 we present the evaluation data. In section 6 we discuss the empirical strategy along with the identification assumptions. Our main results appear in section 7. In section 8 we show some robustness tests, and we conclude in section 9

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.[3] Numerous 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 extensive review of 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

[3] 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.

theoretical foundation for this approach assumes that the labor earnings of trainees would not equal their opportunity marginal product but would be less for the total cost of training, there is no risk aversion, and the post training work span is fixed at *N* 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 measure, and ε is the stochastic term. Figure 2A 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-run than in the medium-term. On the contrary, we assume that untrained individuals receive the same earnings before and after the training (*NN*), being the difference between *TT* and *NN* greater the greater the cost of investment and the return from investment (Becker 1993).

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 the training quality. These assumptions are similar to the statements done in the school quality literature (e.g.; Behrman and Birdsall 1983). The first assumption is necessary to obtain empirical estimates since the effect of training quality is only identified if quality is not homogenous. The second assumption is also needed to guarantee that individuals with lower private cost of learning do not select into high-quality courses or programs.[4] 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 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

[4] If individuals with lower cost of training tend to have higher level of earnings regardless of their training choice, then the estimated marginal productivity will be upward-biased as an estimate of the true marginal productivity.

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 2B 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 grater the greater the quality. Put differently, higher training quality increases the slope of the observed relationship between earnings and training. Moreover, the concavity of the earnings profile may be more or less pronounced depending on whether the effect of the level of training quality 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 extend and the quality of training; and μ is the new stochastic term. Because the underlying relationship between earnings and the effective level of training is unknown, one can assume a variety of functional forms for the smooth continuous function *T*(.),* including splines and polynomials on *Q*.[5]

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. In the context of public-sponsored training programs, the identification of the quality effects gets even more complicated since the programs target unskilled 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 we can relax the linear assumption in the earnings equation and implement more complex econometric estimators such as difference-in-differences

[5] 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 only effect of measured schooling quality is the earnings returns for college graduates.

matching (Heckman et al 1997), which is based on time-invariant identifying assumptions.

3. The PROJOVEN Program

To smooth the short-run negative effects of structural reforms on the welfare of poor households in Latin American 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 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 review of 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 job-specific training in blue-collar occupations.[6] 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 roughly five hours per day for three months. In the second phase, training institutions must place trainees into a paid, on-the-job training experience in private manufacturing firms for an additional period of three months.

The selection of the training courses relies on bidding processes that targets the relatively best training courses at the best competing prices. Thus, private and public training institutions that operate for profit compete for limited public funding following standard processes and strict timetables. To ensure the relevance of the training courses,

[6]These occupations are, for example, maintenance mechanic, electricians, janitors and building cleaners, cashiers, receptionist clerks, construction laborers, plumbers, pipefitters, maintenance and repair workers, sewing machine operators, textile operators and tenders, and computer operators.

the program relies on a demand-driven mechanism that stipulates that all training centers must present, as part of their offers, formal agreements with private manufacturing firms that guarantee a paid, on-the-job training for each beneficiary. This program design requires a strong match 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 in successfully completing the six-month course provides the incentives to train only for those occupations with assured labor demand.[7]

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 3 shows the dynamic 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 those neighborhoods with a high concentration of households below the poverty line. 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 five key observable variables (poverty status, age, schooling, labor market status, and pre-treatment earnings) determines who is eligible and who is not. This process concludes when there are nearly twice as many eligible individuals as training slots.[8]

A two-tiered monitoring and supervision process guarantees the reliability of the information given by the prospective applicants to determine their eligibility status. In

[7] 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.

[8] The number of selected courses depends on the available training slots, which are determined *ex-ante*.

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 applicant's eligibility status does not guarantee participation in the program. Program enrollment depends on both training centers' and 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. This process concludes when the number of eligible individuals exceeds by 75 percent the number of available slots in each course.

Finally, the training institutions select beneficiaries from the pool of eligible applicants selected by the program administrator (position D). This final step does not follow standardized criteria since each institution applies its own rules. Because the eligible-beneficiary ratio is around 1.75:1, the training centers have a limited role in selecting beneficiaries.

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 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.[9]

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's Terms and Conditions regulate these processes and follow a set of international standards that were

[9]For those training centers that participate in two or more consecutive programs, the previous performance is also considered as an additional evaluation factor. It explains almost half of the total score.

previously approved by the Inter-American Development Bank and the United Nations as part of their role of guarantors in this program. This document stipulate, for instance, the starting and closing dates for the bidding, the schedule of the payments, and potential conflict of interest; 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.[10]

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 extracting both quantitative (e.g., number of computers, number of instructors, etc), and qualitative information (e.g., curricula, activities, etc) about each competing course. Once the deadline is reached, two sealed envelopes containing the technical specifications and price offers are open in a public act, where the price offers are made public. The documents containing the technical specifications, on the other 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.[11]

Three distinctive features characterize the quality measurement in this second step. 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 target 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). The inclusion of this new set of "soft" variables that defy an objective description may dramatically improve the explanation for differences in quality

[10]The PROJOVEN's Terms and Conditions can be found at *http://www.projoven.gob.pe*. Both the Inter-American Development Bank and the United Nations have played an important role in the transparency of these processes because their involvement in the funding and administration of the funds, respectively.

[11]For instance, the variable that measures the quality of the equipment considers four different scores: 4 for very good quality, 3 for good quality, 2 for low quality, and 0 for very low quality.

(Hanushek, 1986). 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 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. 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 five different bidding processes corresponding to the first, second, fourth, sixth, and eighth programs. Two features emerge. First, there is variation in the scores assigned to each category within and across programs. In particular, expenditures per trainee and curricular structure are the variables that vary most between and within programs. On the other hand, infrastructure and equipment are the variables that show the smallest variation. Second, as one might expect, there is an increasing trend in the average quality for some proxies over time. This is explained by a natural learning curve on the part of continuously participating institutions, and by the relatively small number of new entering training institutions.[12]

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

[12]The average number of new training centers entering in successive programs is 9.

the first principal component, which is a linear combination of the quality proxies that accounts for the highest proportion of their variance. The lower panel of Table 1 shows large variability in the resulting index within and across programs. We also include the number of competing training institutions, courses offered, and courses accepted for these five rounds of the program. The average number of training institutions is 33 per program, ranging from 30 to 48. These institutions offer an average of 200 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.[13] 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 10 different sub-samples associated with 5 different cohorts of beneficiaries receiving treatment in Lima, and 5 corresponding comparison group samples.[14] The beneficiary subsamples are selected from a stratified random sample of

---

[13] For instance, the program administrator cannot assign any resulting contract without prior written approval from the United Nations or the Inter-American Development Bank.

[14] Individuals that satisfy the same eligibility criteria in terms of age, education, poverty status, and labor outcomes compose all five cohorts. The only difference among these groups is the period when they receive the treatment. These periods extend from November 1996 to April 1997; February 1998 to July 1998; March 1999 to August 1999; June 2000 to December 2000; and August 2001 to January 2002, respectively.

the population of participants corresponding to the first, second, fourth, sixth, and eighth rounds of the programs.[15] 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 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. The neighborhood dimension may have the ability to control some unobservables, including geographic segregation, transportation costs, and firms' location, which may affect the propensity to work and the potential outcomes.

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 the program. The baseline survey provides rich information for all variables that define the eligibility status. It also contains demographics and labor-market information. In fact, relevant factors affecting both the propensity to participate in the program and labor market outcomes are available. There is information, for example, on education attainment, marital status, number of children, parents' schooling, and participation in welfare programs. 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. At the household level, we have information about family size, family income, and household's density rate. In addition, the datasets provide detailed information on dwelling characteristics including source of drinking water, toilet facilities, and house infrastructure (type of materials used in the floor, ceiling, and walls), which is used to measure the poverty status. 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.

[15] The total number of participants in these program rounds is 1507, 1812, 2274, 2583, and 3114, respectively. The corresponding number of treated individuals in the random sample is 299, 321, 343, 405, and 421.

5.1 Comparison of Pre-Treatment Sample Means

Table 3 compares the means of several covariates for the treatment and comparison samples for each one of five different cohorts. Column 2 shows the means using the pooled sample and columns 3 to 7 show the p-values for the test of difference of means for each cohort. In terms of demographic and socioeconomic characteristics, Panel A shows the effectiveness of the "neighborhood" strategy to balance the distribution of covariates that determine the eligibility status. Both groups have the same average age (19), sex ratio (42 percent are males), and schooling attainment (85 percent have completed high school). The p-values for all cohorts under analysis do not reject the null hypothesis of equality of means. The data show, however, that both marital status and children variables have different distributions. About 90 percent of the participants are single and only 14 percent have children, which differs from the comparison sample, which has a lower proportion of single people (77 percent) and a higher proportion of individuals with offspring (25 percent). The p-values show that this is a robust result for all cohorts.

Panel B compares labor market characteristics for treatment and comparison samples. Both groups have the same proportion of individuals in and out of the labor force. Approximately 52, 25, and 22 percent of individuals were employed, unemployed, and out of the labor force, respectively. These non-significant differences are consistent across all cohorts as is shown by the p-values. The type of work depicts a somewhat different pattern. A higher proportion of comparison individuals were working in the formal private sector (63 versus 54 percent) whereas a higher proportion of treated individuals were non-paid family workers (17 versus 10 percent). A comparison of monthly earnings also shows that treated units receive on average smaller earnings than their counterpart comparison sample, which is a steady result across all cohorts.

Panel C compares households and dwelling characteristics. On average family income is somewhat smaller for treated individuals, although the p-values show mixed results across different cohorts. In addition, the analysis of dwelling characteristics shows that a higher proportion of treated individuals live in houses with somewhat better infrastructure and access to flush toilet and piped water. These differences, however, are not significant for several cohorts. Finally, Panel D shows parental schooling attainment. In general, the schooling distribution in both samples is similar, with mothers having fewer years of formal education than their spouses. The p-values do not reject the null hypothesis of equality of means for most of the categories.

In summary, the baseline datasets show that we are dealing with a homogenous population in terms of several socioeconomic and labor-market characteristics, including sex, age, schooling, parents' education, type of work, previous training, and family size. On the other hand, the data also reveal some significant differences in variables such as marital status, presence of children, monthly earnings, family income, and some dwelling characteristics, which would play an important role in any econometric strategy intended to eliminate selection bias.

6. The Empirical Strategy

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 $Y_0(g)$ 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 $g=0$, or participation in a training course of quality g, where $g<q$. We observe the pairs $(Y_1(q),T_1)$ and $(Y_0(g),T_0)$ but never $(Y_1(q),T_0)$ or $(Y_0(g),T_1)$. Because of this missing data problem, we cannot identify for any particular individual the treatment gains $\Delta_i=(Y_1(q)-Y_0(g))$. We focus, instead, on both average and marginal treatment impacts conditional on the quality of the training courses.

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) on the individuals who attend a high-quality course:

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

While $E(Y_1(q)\,|\,T=1)$ may be estimated from the observed treatment sample, the right-hand side of the equation (2) contains the missing data $E(Y_0(g)\,|\,T=1)$. 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 implement matching methods to estimate the counterfactual outcome for program participants by taking weighted averages over the outcomes of observationally similar untreated individuals. Thus, we relax any linear assumption that may mask the earnings-quality relationship. We 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 models.

6.1 Identifying Mean Impacts when the Counterfactual is not Participation ($g = 0$)

In general, standard matching methods eliminate selection bias by balancing the distribution of observables of the untreated group with that of the treated group. However, there may be systematic differences in *T=1* and *T=0* outcomes even after conditioning on a rich set of observables. Such differences may arise in the PROJOVEN program from three different sources. First, it is impossible to control differences in innate ability or motivation. 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 correlated with the quality index.

To eliminate bias arising from unobservables, we can 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 set of conditioning variables *X* such that

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

where *t'* and *t* refer to before and after the start of the program and $g = 0$. This assumption ensures that after conditioning on a rich set of observable variables, the outcomes for treated and untreated individuals follow a parallel path.

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. In this sense, matching forces us to compare comparable individuals in a way that standard regression methods do not.

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

$$\Delta^{DID} = \frac{1}{n_1} \sum_{i \in n_1 \cap S_p} \left\{ [Y_t(q) - Y_{t'}(0)] - \left\{ \sum_{j \in n_0 \cap S_p} W(\|i - j\|)[Y_t(0) - Y_{t'}(0)] \right\} \right\}. \quad (5)$$

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 key 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 estimate the counterfactual outcome $\sum W(\|i, j\|)[Y_t(0) - Y_{t'}(0)]$ using local linear regression methods that 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). This nonparametric approach relies on standard kernel weighting functions that assign greater weight to individuals who are similar, and is more efficient than local constant regression methods because of its lower boundary bias in regions of sparse data.[16]

6.2 Identifying Marginal Program Impacts $(g < q)$

We are also interested in the marginal treatment impacts of increasing quality in the program from g to q, where g>0, 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 potential sorting. Indeed, this is the main econometric

[16]The local constant regression presents bias inversely proportional to the density distribution of the untreated sample. This feature may be problematic because typically the mass of the treated individuals is located in regions where the number of untreated ones is sparse.

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, eligible individuals choose the course they want to attend on a first-come-first- served basis, which is not the case in studies addressing college education. Second, there is large variability in the quality of the training courses within training institutions.[17] 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. From interviews with both the program operator and training institution personnel, it seems that the selection of beneficiaries among the eligible individuals is driven by variables such as marital status, children, and specific physical requirements arising from the courses (e.g., body mass for handling weights).[18]

Because we cannot ignore sorting in our data, we again implement difference-in-differences matching methods that assume selection in observables and unobservables to eliminate selection bias. Formally, the identifying condition (3) changes to

$$E(Y_t(q_1) - Y_{t'}(0) \mid X, T = 1) = E(Y_t(q_2) - Y_{t'}(0) \mid X, T = 1). \qquad (6)$$

which states 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 this is adjusted for the changes implied in assumption (6). This new estimator gives the impact of increasing the quality of the program from q_1 to q_2 for the group of individuals who enrolled in the training course of quality q_1, where $q_1 > q_2$.

6.3 Empirical Issues

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 balancing score

[17]The pooled data show that the standard deviation for the within-institution quality-index ranges from 0.10 to 0.27 whereas the overall standard deviation is 0.18.
[18] It is against the program's rules to select individuals based on age, race, sex, or schooling.

$b(X)$ is also justified. The balancing scores is a function of attributes at least as "fine" as the valued index that predicts the probability of participation: the propensity score.

The proof that assumptions (3) and (6) hold for $P(X)$ instead of *X*, is attained by a balancing property,

$$E(X \mid T=1, P(X)) = E(X \mid T=0, P(X)) = E(X \mid P(X))$$

This is a non-trivial property because in general $P(X_m) \approx P(X_n)$ does not imply $X_m \approx X_n$, and hence, $E(X_m) \neq E(X_n)$.[19] As a result, we assume that equations (3) and (6) hold when we replace X with $P(X) = \Pr(T=1 \mid X)$.

In the empirical work, we estimate the propensity score using a probit approach and implement the balancing test suggested by Dehejia and Wahba (1999).[20] Table 3 shows the probit results for all cohorts. As expected, the covariates used to construct the comparison samples (age, sex, schooling, and work status) are not significant predictors for program selection as they are balanced between treatment and comparison groups. In general, past earnings, experience, type of work, dwelling characteristics, mother's education, family income, and family density rate are the most important predictors of participation in the PROJOVEN program. The estimates also show that married individuals and people with offspring are less likely to participate, although the coefficients are not significant for some cohorts. Furthermore, 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.[21]

---

[19] This is a key difference from covariate matching where $X_m = X_n$ automatically implies $E(X_m) = E(X_n)$ for treatment and comparison samples.

[20] 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.

[21] 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 2 percent of estimated density values are trimmed.

To implement the local linear kernel matching (equation 6) 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, which selects the value that minimizes the mean square error of the local linear regression estimator over a bandwidth search grid.[22] 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 (Black, Galdo, and Smith, 2006).[23]

7. Matching Estimates

Table 4 presents matching estimates applied separately to each one of five different cohorts. Each column refers to one cohort, and the last column shows the pooled data estimates. The upper panel (A) depicts short-run 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 for the treatment impacts are presented along with their corresponding bootstrap standard errors (in parentheses) and percentage gains (in brackets), which are calculated using the mean earnings in the baseline period.

By looking at the first row of each panel, one can observe that the PROJOVEN program is an effective, active labor-market initiative, as was previously shown in partial evaluations of the program.[24] The overall treatment impacts on the treated are S./ 67 soles 6 months after the program, S./ 49 soles 12 months after the program, and S./ 44 soles 18 months after the program.[25] Compared with the mean pre-treatment earnings, these numbers show large increments (59, 43, and 39 percent, respectively), although they represent only about one-third to one-fourth of the Peruvian monthly minimum wage in

[22]The bandwidth grid is defined over values 0.8 through 8 for the logs odd ratio, with a step size of 0.1.
[23]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.
[24]Galdo (1998), Ñopo, Saavedra, and Robles (2001), Chacaltana and Sulmont (2002).
[25]The exchange rate between soles and U.S dollars is about 3:1 for the period under analysis.

the period of analysis. These statistical significant gains are mainly explained by the large number of individuals that relocated from unproductive jobs toward productive ones in private firms protected by international labor-standard laws. For example, the percentage of beneficiaries working as either unpaid family workers or housekeepers decreases from 31 to 6 percent 12 months after the program. 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).

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 unobserved covariates. By looking at the pooled sample estimates, we observe that 6 months after the program the differential effect between high and low-quality courses reaches 10 percentage points, and increases to 19 and 36 percentage points 12 and 18 months after the program. These results suggest that the earnings gap between individuals attending high- and low-quality courses increases over time, which is mainly explained by a sharp decrease in the medium-term earnings of those beneficiaries attending low-quality courses.[26]

Table 5 presents the marginal matching estimates in parallel format to Table 4.[27] Thus, we show short-run (upper panel) and medium-term (lower panels) treatment impacts for each cohort (columns 2 to 5) and the pooled data (column 6). Within each panel, we present two marginal treatment impacts: the effect of increasing the quality of the training services from q_1 (lowest quartile) to q_4 (top quartile), and the effect of increasing quality from q_3 (quartile 3) to q_4 (top quartile). Two main patterns emerge.

First, the marginal impacts indicate mostly positive treatment impacts for those attending high-quality courses, although we lose statistical significance due to sample

[26] The estimates emerging 6 and 12 months after the program are estimated without considering attrition in the data because of the marginal number of missing individuals (4 percent). A potential problem emerges, however, when comparing these estimates with those emerging 18 months after the program because the lack of information for the last cohort. For this reason, we re-estimate the impacts 6 and 12 months after the program without including the last cohort. None of the empirical implications of the paper changes and, therefore, we only provide the first set of estimates.

[27] We match on the predicted probability of attending s high-quality training course (top quartile) using the same set of regressors as before. These propensity score models are not reported but they are available upon request.

size issues. As expected, these estimates are much smaller than the overall mean treatment impacts that emerge when the group of non-participants constitutes the counterfactual group. These marginal gains are not explained, however, for differences in the types of jobs hold for beneficiaries attending high- and low-quality courses. In fact, the percentage of beneficiaries who hold unproductive jobs is similar between these two groups before and after the program. This may suggest that some productivity-enhancing effects explain these earnings differentials.

Second, the marginal impacts are consistent with our previous findings about the increasing trend of the returns to training quality over time. For instance, the estimates from the pooled sample reveal that the effect of increasing quality from q_1 to q_4 changes from 18 to 32 percent when one moves from 6 to 18 months after the program.

When the estimates from Tables 4 and 5 are taken together, three important 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 unproductive jobs toward productive ones in firms protected by international laws that guarantee minimum work conditions.[28] In addition, these average 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.[29]

Second, 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. Finally, this study highlights the importance of having multiple cohorts of participants for the

[28]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.

[29]Whereas the Peruvian public school system spent S./ 1,200 soles per-capita in 2001, the PROJOVEN program spends about S./ 2,600 soles per-capita.

same program design when assessing the effectiveness of active labor market programs. The sensitivity of some estimates to the sample used is very illustrative.

8 Robustness Checks

8.1 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). Figure 4 depicts the earnings trajectory for the treatment and comparison groups. Three clear patterns emerge. Fist, there is some evidence about the existence of Ashenfelter's Dip in the PROJOVEN program 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. Second, the pre-program earnings dip is similar for individuals attending both high- and low-quality courses. Thus, our marginal treatment impacts may be unaffected by Ashenfelter's dip. Third, the full post-program earnings trajectory is consistent with the point estimates emerging 6, 12, and 18 months after the program. In particular, individuals attending low-quality courses show the largest drop in earnings in the medium-term.

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 6 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 pooled sample are 46, 38, and 31 percent 6, 12, and 18 months after the program. For the same reference periods, the treatment impacts for those attending high-quality courses are 54, 43, and 47 percent, while for those attending low-quality courses are 36, 26, and 19 percent. All these OLS treatment impacts are statistically significant at the 5 percent level. These estimates corroborate the evidence of the strong effect of training quality on the earnings of program participants and are also consistent with the previous findings about the large drop in the medium-term earnings of beneficiaries attending low-quality courses.

8.2 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. If this is true, it may hamper the causal relationship we have been testing in this paper. To address this potentially confounding factor, we use two different approaches. First, we check whether individuals enrolled in high-quality courses have larger treatment doses than individuals enrolled in low-quality courses. Using the pooled sample, 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 67 and 63 percent complete at least a month out of the three months of on-the-job training experience.

Table 7 presents a second, more stringent test. We estimate average treatment impacts on the treated by using both 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. Each row describes a different parameter of interest and each column refers to impacts six, 12, and 18 months after the program.

Two patterns emerge. First, the treatment impacts for the formal training are positive although much smaller with respect to the overall mean program impacts. 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, half a year after the program, the matching (OLS) treatment impacts for individuals attending high- and low-quality courses are 49 (29) and 3 (-20) percent, respectively. These estimates, however, are not statistically significant.

Taken together, the estimates in Tables 4 and 7 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 (Row 1 in Table 7) with the overall impacts (Column 7 of Table 4), we observe large differences that suggest that formal training alone is 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.

8.3 The Gender Dimension

Columns 9 and 10 of Table 2 show a potential damaging effect for the identification of the training quality returns: a disproportional number of males attend high-quality courses. Thus, the returns to training quality may not follow from gains in productivity but from intrinsic labor-market returns to males' work. To purge this confounding factor from our estimates, we re-estimate both the difference-in-differences matching and OLS estimators separately for males and females by using the pooled data.

Three basic patterns emerge from Table 8. First, the large overall treatment effects found in the PROJOVEN program are driven by the performance of female participants. They show large and statistitically significant effects in both the short-run and the medium-term. In contrast, the male participants show positive but smaller effects in the short-run and no effects in the medium-term. Second, the returns to training quality are again positive and statistically significant when looking only the estimates emerging from the female subsample. We observe that the OLS average treatment impacts for females attending high-quality courses are 87, 90, and 97 percent after six, 12, and 18 months of participation in the program. In contrast, the treatment gains for males are smaller and not statistically significant, although we still observe important differences between males attending high- and low-quality courses. The large number of female participants who moves from unproductive jobs toward productive ones explains these striking differences. In fact, 43 percent of female participants were working as either unpaid family workers or housekeepers before the program. This number reduces to 10 percent 6 months after the program. In contrast, only 19 and 3.5 percent of males hold these types of jobs before and after the program, respectively. Finally, the point estimates from our matching estimators do not differ substantially from the corresponding OLS estimates, which indicate that the treatment impacts reported in this paper are robust to alternative identifying assumptions.

9. Conclusions

The adoption of market-based approaches that ensure both quality and relevance in the provision of training services has been shown to effectively increase the earnings of disadvantaged young individuals, who frequently emerge from public schools operating far from any efficient frontier. The overall mean gains are mainly explained by the ability of the program to relocate individuals from unproductive jobs to productive ones in firms protected by international laws that guarantee minimum work conditions. The size of the point estimates are heightened because the per capita expenditures on participants are not small relative to the deficits that this program is being asked to address.

We also find strong heterogeneity of the treatment impacts by considering the quality of the training services. Individuals attending high-quality training courses show much higher impacts than those attending low-quality courses. The fact that the distribution of types of jobs between individuals attending high and low-quality courses is similar before and after the program suggests that some productivity-enhanced effects of high-quality training may explain the marginal gains.

This entire set of positive average and marginal training impacts is largely driven by the performance of female beneficiaries, who demonstrate much larger treatment effects than male participants. In addition, these earning differentials are larger in the medium-term than in the short-run, which is explained by a sharp drop in the medium-term earnings of individuals attending low-quality services.

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 productive gains between people attending training courses of varying quality. Thus, a policy implication that follows from this result is that on-the-job training experience may mitigate the outcomes of low-quality training services.

This favorable assessment of the PROJOVEN program should be tempered by the existent 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 strong quality premiums observed in this paper are based on a sample of training institutions that pass a minimum quality threshold imposed by the program administrator. 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

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

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)

Bitler, M., J. Gelbach, and H. Hoynes. 2003. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." Cambridge, Mass.: National Bureau of Educational Research (NBER) Working Paper 10121.

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..

Brewer, D., E. Eide, and R. Ehrenberg. 1999. "Does it Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings." *Journal of Human Resources* 34(1): 104-123.

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

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.

Chacaltana, J. and D. Sulmont. 2003. "Políticas Activas en el Mercado Laboral Peruano: El Potencial de la Capacitación y los Servicios de Empleo." Lima, Peru: Consorcio de Investigaciones Económicas y Sociales (CIES).

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.

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. 2006. "Bandwidth Selection and the Estimation of Treatment Effects with Non-Experimental Data." Manuscript.

Glewwe, P. and M. Kremer. 2005. "Schools, Teachers, and Education Outcomes in Developing Countries." In: E. Hanushek and F. Welch, eds. Handbook of the Economics of Education. Amsterdam: North Holland.

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, J. Smith, and P. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66(5): 1017-1098.

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., 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., Layne-Ferrar, A., and Todd, P. 1995. "Does Measured School Quality Really Matter? An Examination of the Earnings Quality Relationship, NBER, Working Paper W5274

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

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

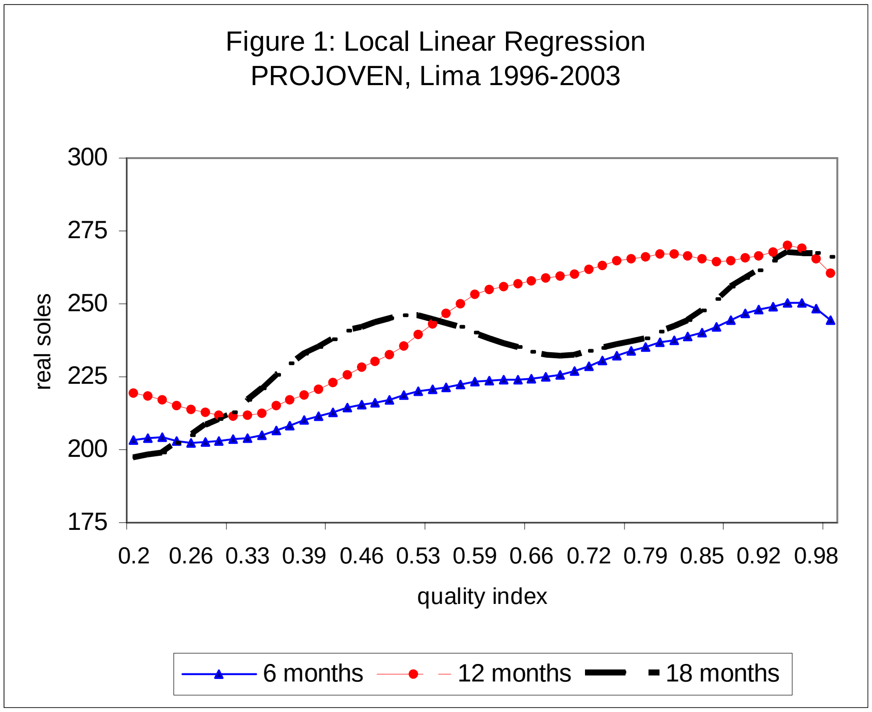
Ñopo, H., J. Saavedra, and M. Robles. 2001. "Una Medición del Impacto del Programa de Capacitación Laboral Juvenil PROJOVEN." Lima, Peru: Group for the Analysis of Development (GRADE).

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.



Notes: Pooled data. Epanechnikov kernel with bandwidth=0.20. Dependent variable is soles (in real terms). The quality index is constructed by using first principal component of factor analytic methods.

Figure 2A: Training and Earnings Over Time

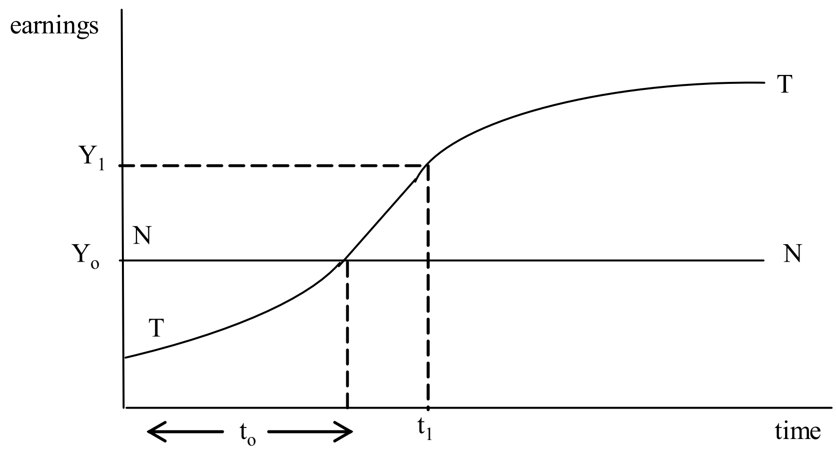


Figure 2B: Training Quality and Earnings Over Time

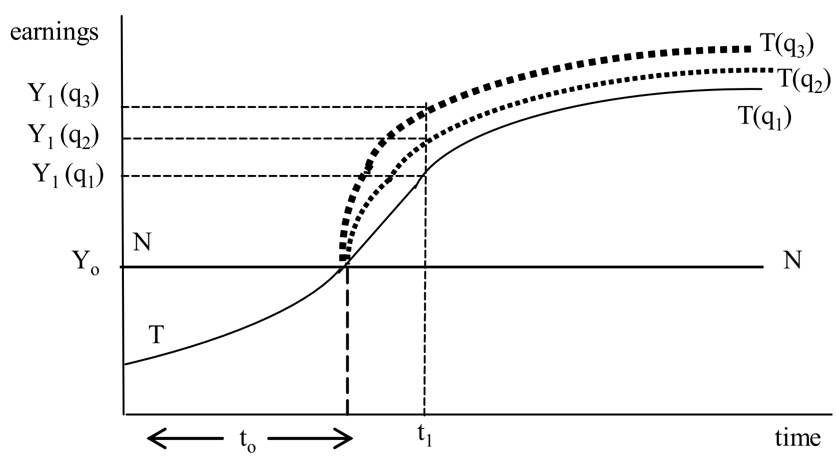


Figure 3. Beneficiary Selection Process
Youth Training Program PROJOVEN, Lima 1996 to 2003.

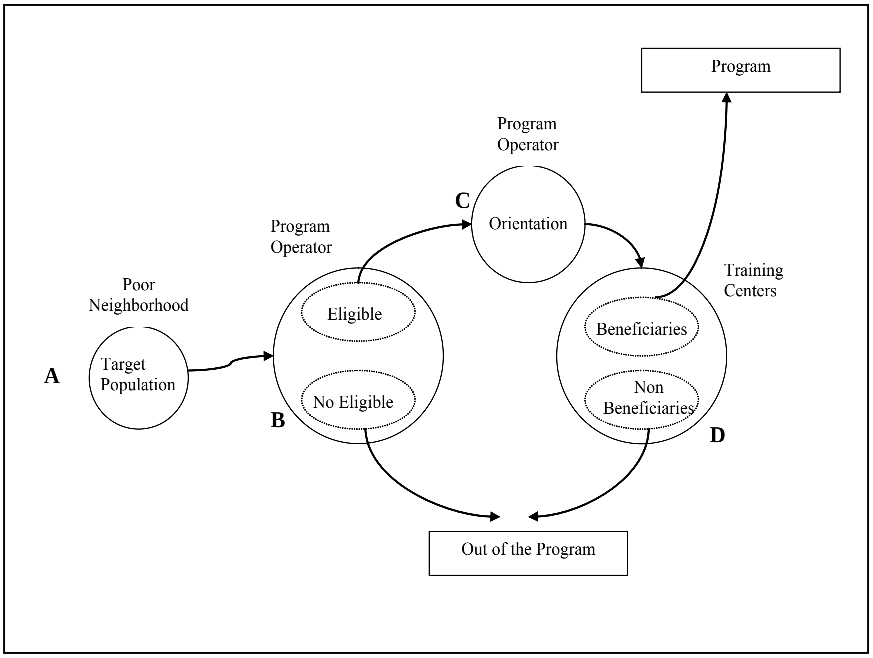
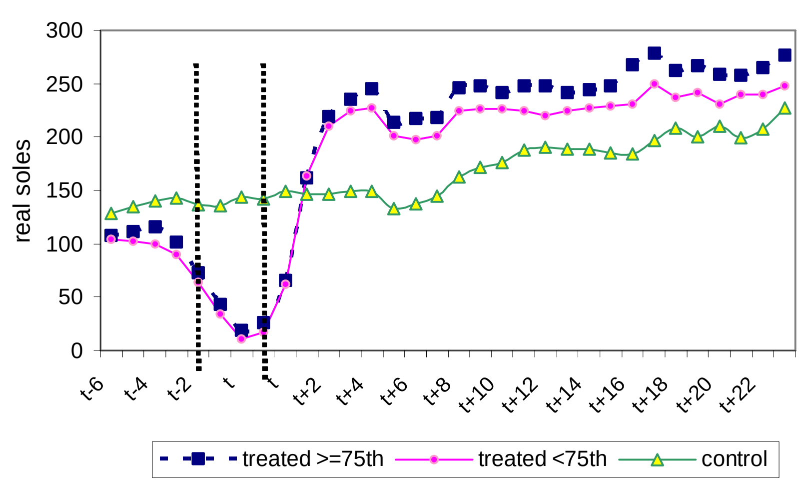


Figure 4: Monthly Earnings for Pooled Data
PROJOVEN, Lima 1996-2003



Note: Pooled means are unweighted. The quality index is constructed by using first principal component of factor analytic methods.

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

Quality Categories	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5
Class size	0.28	0.35	0.41	0.43	0.35
	(0.20)	(0.26)	(0.23)	(0.29)	(0.29)
Expenditures per trainee	0.39	0.39	0.55	0.48	0.50
	(0.23)	(0.26)	(0.20)	(0.23)	(0.16)
Human resources	0.72	0.57	0.64	0.65	0.57
	(0.25)	(0.24)	(0.22)	(0.18)	(0.22)
Infrastructure	0.85	0.94	0.97	0.95	0.96
	(0.27)	(0.19)	(0.15)	(0.16)	(0.12)
Equipment	0.54	0.67	0.65	0.81	0.79
	(0.26)	(0.30)	(0.21)	(0.16)	(0.26)
Curricular structure	0.56	0.85	0.78	0.69	0.68
(contents and activities)	(0.28)	(0.26)	(0.24)	(0.29)	(0.27)
Market Knowledge	0.74	0.68	0.83	0.71	0.64
(worker and occupational analysis)	(0.22)	(0.21)	(0.15)	(0.19)	(0.21)
PCA Quality Index	0.62	0.68	0.70	0.50	0.50
	(0.18)	(0.15)	(0.15)	(0.16)	(0.18)
# competing institutions	30	33	35	33	48
# competing courses	154	158	215	204	363
# funded courses	75	98	118	148	169

Notes: The available bidding data have aggregate scores for each category. These 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

	Pooled data		cohort 1	cohort 2	cohort 3	cohort 4	cohort 5	Quality Index	
	treated	comparison	p-value					treated >50th	treated <50th
A. Socio-Demographic									
age	19.67	19.73	0.02	0.51	0.22	0.92	0.84	19.69	19.64
sex (%)	42.70	42.60	0.93	0.97	0.91	0.88	0.94	46.00	40.00
schooling (%)									
none	0.23	0.25	0.08	0.31	0.97	0.31	----	0.12	0.34
incomplete primary	1.04	0.69	0.26	0.16	0.96	0.77	1.00	0.60	1.49
complete primary	4.82	6.27	0.38	0.22	0.29	0.46	0.85	4.64	5.05
incomplete high school	8.76	8.00	0.97	0.57	0.98	0.77	0.87	7.62	9.87
complete high school	85.14	84.70	0.53	0.92	0.88	0.79	1.00	87.02	83.24
marital status (%)									
single	91.19	77.44	0.00	0.00	0.00	0.00	0.00	92.02	90.38
married and/or cohabitating	5.12	14.87	0.00	0.00	0.00	0.00	0.00	5.24	5.27
other	3. 69	8.79	0.56	0.65	0.14	1.00	0.25	2.76	4.35
have children (%)	14.66	25.83	0.00	0.00	0.00	0.00	0.00	13.61	16.00
number of children	1.21	1.28	0.71	0.28	0.21	0.42	0.12	1.18	1.24
B. Labor information									
work status (%)									
have a job	52.17	52.11	0.54	0.84	0.88	0.94	0.94	54.52	50.00
unemployed	25.80	26.58	0.36	0.98	0.9	1.00	0.87	23.57	27.84
out of labor force	22.03	21.30	0.16	0.82	0.96	0.93	0.92	21.90	22.45
kind of work (%)									
self-employed	19.89	21.04	0.59	0.22	0.28	0.41	0.88	19.00	20.74
worker in private sector	53.67	62.23	0.00	0.71	0.98	0.42	0.03	55.46	51.18
worker in public sector	0.66	0.88	0.31	0.56	0.14	0.16	0.56	0.66	0.67
unpaid family worker	17.67	10.24	0.00	0.12	0.63	0.00	0.00	17.03	18.66
housekeeper	7.33	5.29	0.47	0.15	0.03	0.30	0.02	7.21	7.60
monthly earnings	91.54	126.00	0.00	0.02	0.08	0.00	0.00	99.34	83.40
experience	2.88	2.71	0.06	0.06	0.00	0.67	----	2.84	2.93
participation in training courses	23.03	23.00	0.37	0.32	0.02	0.03	0.00	24.70	21.50
hours of training	56.87	56.02	0.51	0.17	0.05	0.05	0.00	57.40	57.03
C. Household characteristics									
number of persons	6.23	6.00	0.00	0.44	0.00	0.10	0.28	6.26	6.22
household income	828.00	959.00	----	----	0.12	0.00	0.00	894.31	767.45
number of bedrooms	2.09	2.15	0.00	0.37	0.24	0.31	0.04	2.15	2.03
household density rate	3.12	2.87	0.05	0.90	0.00	0.72	0.00	3.10	3.13
floor: earthen	65.04	72.02	0.00	0.84	0.03	0.87	0.04	65.92	61.00
ceiling: concrete	35.07	23.98	0.00	0.29	0.00	0.00	0.00	34.13	35.85
walls: concrete/bricks	67.03	62.33	0.11	0.91	0.34	0.01	0.19	67.33	66.55
water: piped into the home	72.57	60.22	0.00	0.16	0.11	0.11	----	73.28	72.04
water sewage: flush toilet	65.95	61.28	0.00	0.33	0.15	0.01	0.10	65.87	65.95
D. Parent's schooling									
father (%)									
none	2.10	1.94	------	0.77	0.87	0.13	0.59	2.46	1.66
primary	37.52	32.78	------	0.63	0.35	0.35	0.00	37.77	37.76
incomplete high school	20.76	20.51	------	0.24	0.48	0.86	0.78	21.71	20.06
complete high school	26.72	32.22	------	0.77	0.94	0.00	0.00	26.77	26.83
higher education	7.57	5.20	------	0.29	0.37	0.01	0.29	6.37	8.85
mother (%)									
none	9.05	7.69	------	0.39	0.16	0.59	0.01	9.99	8.16
primary	47.27	41.93	------	0.89	0.04	0.00	0.34	47.90	47.16
incomplete high school	18.86	19.75	------	0.03	0.68	0.01	0.31	19.10	18.95
complete high school	18.09	21.90	------	0.82	0.69	0.00	0.07	16.79	19.36
higher education	3.72	3.19	------	0.41	0.97	0.00	0.80	3.47	4.01
#	1725	1742	599	627	720	732	764	840	873

Notes: Pooled means are unweighted. There are some observations with missing information for some covariates and thus the means are not based on the same number of observations. p-values refers to the test for differences in means for the treatment and comparison samples.

Table 3: Coefficient Estimates from Balanced Probit Models for Program Participation
PROJOVEN, Lima 1996-2003

	Coefficients				
covariates	cohort 1	cohort 2	cohort 3	cohort 4	cohort 5
A. Socio-demographic					
constant	-1.49	0.02	6.67	1.82	-0.20
age	0.02	-0.03	-0.05	-0.04	0.06**
sex	-0.01	-0.06	-0.07	0.05	0.23**
schooling					
none	-----	-0.44	-0.52	----	----
incomplete primary	1.53**	----	-0.87	----	----
complete primary	----	-0.20	-0.65**	0.17	-1.09
incomplete high school	0.37	----	----	0.71	-0.88
complete high school	0.13	-0.13	-0.22	0.36	-1.11*
marital status					
single	0.26	0.62	-5.67**	-0.25	-0.56
married and/or cohabitating	-0.69	-0.06	-6.47**	-0.53	-1.12**
have children	-0.23	-0.02	-0.12	-0.35	0.38
number of children	0.05	-0.44	-0.01	-0.09	-0.68
B. Labor information					
work status					
have a job	-0.20	-0.51	-1.05**	-0.14	0.47
unemployed	-0.45**	-0.04	-0.19	0.01	-0.16
kind of work					
self-employed	0.38	1.23**	0.95**	0.62	-0.60
worker in private sector	0.18	1.04**	1.15**	0.63**	-0.88
worker in public sector	0.26	1.99*	----	----	-0.97
unpaid family worker	1.56**	0.74**	0.97**	0.72**	0.57
housekeeper	0.27	0.63	1.89**	0.92**	-0.15
monthly earnings	-0.00**	-0.00**	-0.00**	-0.00**	0.00
experience	0.03	0.34**	0.16**	0.10**	----
participation in training courses	-0.80**	0.34	-0.55**	0.20	0.25
hours of training	0.00**	-0.00**	0.00**	0.00	-0.00**
C. Household characteristics					
number of persons	0.06**	-0.05**	-0.25**	0.02	0.02
household income	----	----	----	-0.00*	-0.00**
number of rooms/ number of persons	0.06*	0.01	0.09**	0.00	0.16**
participation in welfare programs	0.14	----	----	----	----
floor : earthen	1.00**	0.31**	-0.11	-0.04	0.09
ceiling					
concrete	0.86**	0.01	0.36**	0.26**	-0.39
matting	----	-0.12	0.09	0.35**	-0.27**
walls: concrete /brick	-0.71**	0.15	-0.09	0.17	0.03
water: piped into the home	0.52**	0.81**	0.28**	----	----
water sewage: flush toilet	-0.24*	-0.55*	-0.38**	0.27**	0.28**
D. Parent's schooling					
father					
no information	----	----	0.63	-0.63	----
none	----	----	----	-0.25	-0.09
primary	----	-0.27	0.20	-0.42	0.31
incomplete high school	----	-0.02	-0.07	-0.36	0.24
complete high school	----	-0.17	0.00	-0.46*	0.09
higher education	----	-0.01	0.16	----	0.49
mother					
no information	----	----	-1.43**	----	----
none	----	0.70**	-0.30	-0.79**	0.74
primary	----	0.30	-0.18	-0.65*	0.33
incomplete high school	----	0.72**	0.00	-1.14**	0.17
complete high school	----	0.78**	-0.11	-1.19**	0.27
higher education	----	0.18	0.22	----	0.29
#	585	604	679	690	705
R^2	0.34	0.23	0.17	0.15	0.17

Note: * statistically significant at 5 percent, ** statistically significant at 10 percent. The specification of each probit model follows Dahejia and Wahba's (1999) balancing test. Not all covariates are observed for all cohorts.

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

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Pooled Data
A. 6 months after program						
$\Delta = Y_1(q) - Y_0(0)$	76 (35) [70]	58 (34) [52]	43 (24) [40]	42 (24) [33]	72 (27) [67]	67 (16) [59]
$\Delta = Y_1(q_4) - Y_0(0)$	143 (49) [131]	89 (49) [79]	27 (45) [26]	22 (49) [17]	58 (41) [54]	72 (21) [64]
$\Delta = Y_1(q_1) - Y_0(0)$	61 (53) [56]	35 (46) [31]	30 (54) [28]	36 (40) [28]	80 (41) [74]	61 (22) [54]
B. 12 months after program						
$\Delta = Y_1(q) - Y_0(0)$	25 (34) [24]	82 (41) [73]	12 (32) [11]	-6 (26) [-5]	94 (32) [89]	49 (13) [43]
$\Delta = Y_1(q_4) - Y_0(0)$	38 (60) [36]	125 (84) [110]	21 (48) [19]	-24 (44) [-19]	93 (35) [79]	54 (27) [48]
$\Delta = Y_1(q_1) - Y_0(0)$	-17 (44) [-15]	75 (60) [67]	-29 (44) [-26]	-44 (41) [-35]	127 (48) [118]	33 (26) [29]
C. 18 months after program						
$\Delta = Y_1(q) - Y_0(0)$	63 (35) [58]	68 (38) [61]	36 (32) [34]	25 (44) [20]	----	44 (22) [39]
$\Delta = Y_1(q_4) - Y_0(0)$	108 (66) [100]	105 (64) [85]	101 (58) [94]	-17 (66) [-13]	----	68 (35) [60]
$\Delta = Y_1(q_1) - Y_0(0)$	39 (70) [36]	75 (49) [67]	-3 (48) [-2]	2 (54) [1]	----	27(48) [24]

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 probit model. Difference-in-differences matching is applied to the sample of individuals inside the overlapping support region. The matching variable is the log of the odd-ratio. We use Epanechnikov kernel function with the bandwidths determined by weighted cross-validation. Pooled data estimates are based on the propensity scores from individual cohort estimates. The optimal bandwidths for the pooled data are 1.3, 1.3, and 1.2 when estimating the treatment impacts 6, 12, and 18 months after the program, respectively. These data include 1,541, 1,453, and 1,118 observations in the comparison group and 1,436, 1,407, and 1,031 observations in the treatment group. In the fourth quartile, there are 347, 336, and 264 observations 6, 12, and 18 months after the program. In the first quartile there are 364, 362, and 224 observations, respectively.

Table 5. Marginal Treatment Impacts on Monthly Earnings
Difference-in-Differences Local Linear Matching Estimator
PROJOVEN, Lima 1996 to 2003

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Pooled Data
A. 6 months after program						
$\Delta = Y_1(q_4) - Y_1(q_1)$	122 (63)	51 (95)	-29 (102)	29 (58)	-38 (46)	19 (35)
	[113]	[46]	[-27]	[23]	[-35]	[18]
$\Delta = Y_1(q_4) - Y_0(q_3)$	125 (60)	6 (58)	-110 (66)	17 (81)	12 (51)	31(29)
	[115]	[5]	[-101]	[20]	[11]	[28]
B. 12 months after program						
$\Delta = Y_1(q_4) - Y_1(q_1)$	72 (55)	5 (83)	116 (80)	46 (52)	-54 (84)	13 (45)
	[67]	[4]	[107]	[36]	[-50]	[12]
$\Delta = Y_1(q_4) - Y_0(q_3)$	-33 (38)	43 (69)	-25 (73)	11 (56)	27 (70)	-17 (38)
	[-31]	[38]	[-23]	[9]	[25]	[15]
C. 18 months after program						
$\Delta = Y_1(q_4) - Y_1(q_1)$	88 (63)	14 (93)	69 (88)	12 (52)	----	36 (37)
	[81]	[12]	[64]	[9]		[32]
$\Delta = Y_1(q_4) - Y_0(q_3)$	118 (49)	50 (76)	11 (86)	-1 (55)	----	38 (41)
	[109]	[45]	[10]	[-1]		[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, q_3, and q_1 are the fourth, third, and first quartiles of the quality index distribution. The propensity scores are estimated using a probit model. The matching variable is the log of the odd-ratio. Difference-in-differences matching is applied to the sample of individuals inside the overlapping support region. Pooled data estimates are based on the propensity scores from individual cohort estimates. We use Epanechnikov kernel function with the bandwidths determined by weighted cross-validation. The optimal bandwidths for the pooled data are 2.7, 5, and 1.2 when estimating the treatment impacts 6, 12, and 18 months after the program, respectively. These data include 369, 358, and 265 observations in the fourth quartile; 339, 332, and 253 observations in the third quartile; and 359, 358, and 261 observations in the first quartile.

Table 6. Treatment Impacts on Monthly Earnings
Parametric Least Square Estimator
PROJOVEN, Lima 1996 to 2003

	Treatment Impacts for Pooled data		
	6 months after program	12 months after program	18 months after program
$\Delta = Y_1(q) - Y_0(0)$	53 (9) [46]	43 (9) [38]	35 (13) [31]
$\Delta = Y_1(q_4) - Y_0(0)$	61 (13) [54]	48 (14) [43]	52 (18) [47]
$\Delta = Y_1(q_1) - Y_0(0)$	43 (13) [36]	30 (15) [26]	23 (18) [19]

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. 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, household' density rate, floor, ceiling, and walls house characteristics, 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 1,539, 1,453 and 1,118 observations in the comparison group 6, 12, and 18 months after the program; and 1,436, 1,407, and 1,031 observations in the treatment group, respectively. In the fourth quartile, there are 347, 336, and 244 observations for the same reference periods. In the first quartile, there are 364, 362, and 224 observations for the same reference periods.

Table 7. Average Treatment Impacts on Monthly Earnings for Formal Training
Difference-in-Differences Local Linear Matching Estimator
PROJOVEN, Lima 1996 to 2003

	Treatment Impacts for Pooled data		
	6 months after program	12 months after program	18 months after program
$\Delta = Y_1(q) - Y_0(0)$			
Matching	21 (16)	13 (12)	31 (29)
	[19]	[12]	[28]
OLS	6 (10)	8 (11)	21 (15)
	[5]	[6]	[19]
$\Delta = Y_1(q_4) - Y_0(0)$			
Matching	55 (35)	40 (37)	59 (40)
	[49]	[36]	[53]
OLS	33 (17)	24 (19)	37 (24)
	[29]	[21]	[33]
$\Delta = Y_1(q_1) - Y_0(0)$			
Matching	3 (31)	-14 (24)	45 (73)
	[3]	[-13]	[40]
OLS	-20 (16)	-25 (17)	32 (27)
	[-18]	[-21]	[28]

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, household' density rate, 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 optimal bandwidths are 1.3, 2.5, and 2.6 when estimating treatment impacts 6, 12, and 18 months after the program. The matching variable is the log of the odd ratio. There are 1,547, 1,453, and 1,118 observations in the control group 6, 12, and 18 months after the program, and 577, 566, and 429 observations in the treatment group, respectively.

Table 8. Treatment Impacts on Monthly Earnings by Gender
Matching and OLS Estimators
PROJOVEN, Lima 1996 to 2003

	Treatment Impacts for Pooled Data					
	6 months after the program		12 months After the program		18 months after the program	
	male	female	male	Female	Male	female
$\Delta = Y_1(q) - Y_0(0)$						
Matching	21 (20) [18]	88 (19) [78]	7 (22) [6]	71 (26) [63]	2 (21) [2]	71 (33) [63]
OLS	23 (13) [20]	75 (11) [66]	12 (15) [11]	69 (11) [61]	-16 (17) [-14]	75 (18) [67]
$\Delta = Y_1(q_4) - Y_0(0)$						
Matching	21 (33) [19]	108 (25) [96]	-9 (39) [-8]	100 (35) [89]	13 (39) [12]	106(48) [94]
OLS	25 (19) [22]	98 (17) [87]	-1 (21) [1]	101 (20) [90]	-2 (23) [-1]	110 (32) [97]
$\Delta = Y_1(q_1) - Y_0(0)$						
Matching	0 (32) [0]	88 (26) [78]	2 (37) [2]	52 (25) [46]	-36 (52) [-31]	62 (48) [54]
OLS	7 (23) [6]	66 (16) [58]	12 (27) [10]	46 (17) [40]	-39 (27) [-35]	61 (22) [53]

Notes: Pooled data estimates from males and females sub-samples. 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. 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, household' density rate, house infrastructure (floor, ceiling, and walls), whether have 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 optimal bandwidths are 3.4, 3.4, and 1 when estimating treatment impacts for the male subsample 6, 12, and 18 months after the program; and 1.2, 5.2, and 4.2 for the female subsample, respectively. The matching variable is the log of the odd ratio. There are 675, 621, and 478 comparison observations, and 668, 604, and 439 treated observations for the male subsample 6, 12, and 18 months after the program, respectively. There are 884, 832, and 640 comparison observations, and 812, 803, and 592 treated observations for the female subsample 6, 12, and 18 months after the program.