

Labor Market Impact on Youth:

A meta-analysis of the Youth Employment Inventory¹

Olga Susana Puerto²

A. Introduction

Interventions to support young workers have been broadly applied in both developed and developing economies alike, as concerns for unemployment and inactivity rates drive public and private resources to invest in youth. Despite comprehensive experiences around the world, very little is known about the determinants of success of youth employment interventions. In particular, what type of interventions work best and what are the key features in implementation design and targeting that prompt positive impacts on employment and earnings under different economic and institutional conditions. To respond to these questions, this paper uses a meta-analytical framework that examines the evaluation evidence collected by the Youth Employment Inventory, a World Bank initiative that compiles world-wide interventions designed to integrate youth into the labor market. This meta-analytical framework combines information on program impact, program characteristics, and country context. It employs a probability model to measure the likelihood of obtaining positive impacts on employment and earnings, synthesizing the results from individual studies in a common field.

Empirical results from a sample of 172 evaluated studies – including net impact evaluations and evaluations with gross outcomes - indicate that program success is not determined by the type of intervention but rather by the program's targeting strategies toward disadvantaged youth, the country level of development, and the flexibility of the labor market regulations. Sensitivity tests show these results are stable under different specifications, particularly when the sample size is constrained to studies with net impact evaluations (i.e. with treatment and control groups).

The paper has five sections including this Introduction. Section B starts with a literature review of cross-country studies and meta-analyses of labor market programs. Whenever available, estimates on employment and earnings of young people are discussed. Section C provides a summary of the Youth Employment Inventory, including design, coverage, and main features. It describes the main type of interventions and findings based on a descriptive evaluation of the inventory's evidence. Section D displays the dataset of programs included in the meta-analysis, and describes the key variables into play. It also introduces the model, empirical results and sensitivity tests. Conclusions and implications are drawn in Section E.

¹ This is a background paper for the World Bank's 2007 Youth Employment Inventory, conducted by the Social Protection and Labor Unit of the Human Development Network. Main results from this paper can be found in the inventory's synthesis report (Betcherman et al. 2007).

² Correspondence: Olga Susana Puerto, World Bank, opuerto@worldbank.org.

B. Literature review of overview studies of labor market programs – Focus on Youth

There are few overview and cross-country studies on labor market programs around the world, and the evidence is even more scant when the focus is on specific target groups such as youth. Betcherman et al. (2004) and Dar and Tzannatos (1999) already exposed the substantial lack of solid information and evaluation evidence on active labor market programs (ALMP). In their review, Betcherman et al. (2004) looked at the evidence from 159 net impact evaluations of ALMP, of which solely 12 percent (19 programs) were youth-oriented. Their assessment on youth employment programs, particularly related to training systems, indicates often negative impacts on labor market outcomes (a finding more accentuated in developed economies), stressing the importance of early and sustained interventions to reduce school dropouts rates and improve educational attainment. Developing economies, on the contrary, reported positive impacts when training is offered as part of a comprehensive package.

Evaluation evidence is considerably richer in OECD countries, where there is a long-standing tradition in the design and implementation of impact evaluations and where the sustainability of publicly-funded programs relies greatly on the evaluation outcomes.³ Several evaluations have measured the effectiveness of ALMP in developed countries, and whenever possible, impacts on employability and earnings of young people have been estimated. Despite this amount of evidence, very few studies have put together cross-country analyses to compare the effectiveness of active measures and draw systematic lessons on what can potentially determine the success of employment interventions in OECD countries. Systematic cross-country evidence from econometrical studies stems from the works of Heckman *et al.* (1999), Kluve and Schmidt (2002), and Kluve (2006). Descriptive cross-country analyses for the region can be also found in Martin and Grubb (2001).

Based on a sample of evaluation studies of ALMP implemented in Europe and the U.S. before 1994, Heckman *et al.* (1999) observed the impacts of interventions such as job training, job search assistance, and wage subsidies on employability. The findings suggest a non-clear pattern of success across categories of intervention and very moderate and rather disappointing outcomes, especially for youth. The paper also examines the large heterogeneity in evaluation methods across countries, and particularly between the much less developed evaluation evidence in Europe and the large experience of the U.S. When drawing methodological lessons, the authors suggest there is not an optimal method of choice for conducting program evaluations, i.e. experimental and non-experimental methods as well as other econometric techniques may be – in general- equally convenient to measure labor market impacts so long as the quality of the underlying data is ensured.

Kluve and Schmidt (2002) collected a sample of 53 evaluations of European programs implemented between 1983 and 1999, and compared their results to the U.S. programs previously studied by Heckman *et al.* (1999). Their overview analysis suggests mixed program effects across categories of intervention and target population: while training and job search assistance may be effective to improve the participants' labor market prospects, direct job creation programs in the public sector may lead to negative gains and hinder employability of some target groups. Young workers were the most difficult group to assist among the unemployed.

Following the previous studies, Kluve (2006) set a meta-analytical framework to estimate the probability of success, in terms of positive treatment effects on employment, of ALMP in

³ As is particularly the case in the United States.

Europe, with special attention drawn to programs implemented in the late nineties and in the 2000s. The probability of success is modeled by (i) the category of intervention, (ii) the study design⁴, (iii) the institutional labor market context, and (iv) the economic country-context at the time a particular program was implemented. The meta-analysis used microeconomic evaluations studies of ALMP from 17 European countries, yielding a comprehensive dataset with 137 observations or ‘assessments’ on the net impacts of the interventions by type of program or target group.⁵ About 25 percent of these observations referred to youth employment programs. Results indicate that category of intervention is the only clear determinant of success of active labor market measures in Europe, and there is little if any evidence that study design or country-context factors explain the programs’ effectiveness.

Kluve (2006) defines the set of categories or program types as follows: labor market training, private sector incentive programs (e.g. wage subsidies), direct employment programs in the public sector (e.g. public works programs), and services and sanctions (e.g. job search assistance and compulsory programs to maintain unemployment benefits). Besides these typologies, programs were further disaggregated by target group: youths and disabled workers. Specific findings across categories of intervention suggest rather modest positive impacts from training programs on employability. Relative to training programs, the probability model indicates significantly higher returns from private sector incentive programs and services and sanctions programs: they increase the likelihood of positive labor market impacts in 40 to 50 percentage points more than training programs do. On the other hand, public sector employment programs are 30 to 40 percent less likely to yield positive impacts than training programs. On specific target groups, the model indicated that youths are still the hardest population to assist, reducing the probability of positive employment impacts in 40 to 60 percentage points.

Additional evidence of meta-analytical frameworks of labor market programs is available in the United States. Greenberg *et al.* (2003) synthesized findings from 31 evaluations of 15 publicly-funded training programs to measure the programs’ effects on participants’ earnings⁶. Their model regresses the reported earning effects against (i) type of training, (ii) demographic characteristics of the target population (e.g. gender and race), (iii) economic conditions of the area where the program was implemented, (iv) evaluation method, (v) number of years since training was received, and (vi) year in which the program was implemented. Of a total of 315 observations on earnings outcomes, 99 (31 percent) were related to training programs for disadvantaged youth. Results suggest highly heterogeneous earning effects among assisted groups, i.e. men, women and youths. Findings for youth were particularly discouraging. Across program components, classroom training yielded consistently positive effects relative to on-the-job training. Training appeared to be less effective for whites and female beneficiaries than for all other participants.⁷ The study found a statistically significant positive correlation between

⁴ Study design identifies whether the evaluation followed a randomized experiment and also classifies programs by the experiment’s sample size.

⁵ These 137 observations originated from 95 different evaluation studies.

⁶ Programs included in the analysis are: Manpower Development and Training Act (MDTA), Neighborhood Youth Corps (NYC), Job Corps, Comprehensive Employment and Training Act (CETA), the National Supported Work Demonstration, the JOBSTART Demonstration, the New Chance Demonstration, and the Job Training Partnership Act (JTPA). They were all implemented in the U.S. between 1962 and 1998.

⁷ These findings are consistent with an early paper by Gay and Borus (1980). In their review, Gay and Borus (1980) use administrative data on participants’ earnings from a 1969-72 longitudinal study of four federally sponsored employment and training programs to validate the usefulness of performance indicators in assessing the effectiveness of the programs. The study identified net positive impacts on earnings of out-

training effects and program cost (e.g. a \$1,000-increase in program cost raised the training effect by about \$108).

This is, to the best of our knowledge, the extent to which overview analyses and more concretely meta-analytical frameworks have been applied to the study of labor market programs. In fact, meta-analyses have been only recently introduced to economic discussions⁸ to quantitatively combine and synthesize results from individual studies. Previous applications of meta-analyses include fields such as education, medicine, and psychology (Hunt, 1997).

This paper builds on the set of studies and methodologies described above, and offers a cross-country, cross-program analysis of employment interventions for youth, with the aim of providing systematic and quantitative lessons with far-reaching implications for both developed and developing countries alike. The paper relies on the information gathered by the World Bank's Youth Employment Inventory, a comprehensive initiative that compiles policies and programs designed to integrate young people into the labor market.

C. Evidence on Youth: The Youth Employment Inventory

1. Designing the Youth Employment Inventory

The Youth Employment Inventory is a World Bank initiative that offers a wealth of information on employment interventions to support young workers around the world. It is based on available documentation of current and past programs and includes evidence from 289 interventions from 84 countries in all regions of the world. The interventions included in the inventory have been analyzed in order to (i) document the types of programs that have been implemented to support young workers to find a job; and (ii) to identify what appears to work in terms of improving employment outcomes for youth.

The inventory includes completed and ongoing interventions aiming to facilitate the transition of young people into the labor market, with a particular focus on disadvantaged youth. The inventory is limited to post-formal-schooling interventions and allows the inclusion of both labor market policies and programs with the following caveats: (i) Policy interventions are eligible for inclusion if they specifically target young people; while (ii) labor market programs are included even if they did not explicitly target youth but the documentation indicates that young people were the primary participants. These limitations and further details on the selection criteria are discussed to greater extent in the inventory's synthesis report (Betcherman *et al.* 2007).

The inventory groups interventions into nine categories, displayed in Table 1. These categories set a clear framework for the inclusion and classification of interventions. They are based on a two-fold approach to youth employment problems, consisting of: (1) increasing labor

of-school and black NYC beneficiaries; while there appeared to be significant negative effects for non-black NYC participants and all Job Corps beneficiaries.

⁸ For instance, Jarrell and Stanley (1990) use a meta-analysis to measure wage differentials across union and non-union workers. Also, Sirmans et al. (2006) apply a meta regression analysis to determine the value of housing.

demand in general in relation to supply; and (2) increasing 'integrability' of disadvantaged youth, so that they can take advantage of opportunities that arise when the labor demand increases⁹.

The categories of intervention are largely self-explanatory but a few comments may be useful. Category 1, “making the labor market work better for young people”, includes interventions that improve information (counseling, job search skills), increase labor demand for youth (wage subsidies and public works), and remove discrimination. Category 2, “improving chances for young entrepreneurs”, covers interventions that provide assistance (financial, technical, and training) to youth who are starting their own business. Categories 3 and 4 both deal with training: the former includes the full range of post-formal schooling training programs while the latter includes interventions intended to address training market failures by providing information, credit, and other financial incentives.

Table 1: Categories used to classify interventions in the Youth Employment Inventory

<p>1. Making the labor market work better for young people</p> <p>1a. Counseling, job search skills 1b. Wage subsidies 1c. Public works programs 1d. Anti-discrimination legislation 1e. Other</p> <p>2. Improving chances for young entrepreneurs</p> <p>3. Skills training for young people</p> <p>3a. Vocational training including apprenticeship systems 3b. Literacy & numeracy – young adult literacy programs 3c. 2nd chance & equivalency programs 3d. Other</p> <p>4. Making training systems work better for young people</p> <p>4a. Information 4b. Credit (to individuals or enterprises) 4c. Financial incentives (subsidies, vouchers) 4d. Other</p> <p>5. Programs to counteract residential segregation of disadvantaged young people</p> <p>5a. Transportation 5b. Others</p> <p>6. Improving labor market regulations to the benefit of young people</p> <p>7. Programs for overseas employment of young people</p> <p>8. Comprehensive interventions</p> <p>9. Other (e.g. voluntary national service programs)</p>

Location can also be a barrier for young people if where they reside isolates them from learning or employment opportunities, or even a secure living environment. Category 5 is meant to include interventions (e.g., transportation services or residential mobility) that can help young

⁹ Integrability can be increased by (a) remedying or counteracting market failure, (b) improving labor market regulations, and (c) improving the skills of disadvantaged youth (Godfrey 2003).

people overcome this form of barrier. Category 6 covers regulatory reforms (e.g., changes in labor law, minimum wage, etc.) that are designed to enhance employment opportunities for young people. Category 7 includes programs to provide job opportunities outside the country. Interventions that provide multiple types of services, and thus cannot be included in one of the other groups, are included in Category 8. Finally, Category 9 is a residual grouping.

Several tools were used in the process of collecting and compiling interventions. These tools include a questionnaire to systematize and standardize the information, and the creation of a database to facilitate analysis and interpretation of the evidence (Box 1).

Box 1: Compiling information for the Youth Employment Inventory

Template

A questionnaire template was designed to ensure consistency and uniformity in the collection and recording of information for the inventory. Information collected on each program includes intervention category, country, time period in which it was implemented, current status, the specific labor market problems it sought to address, main objectives, a detailed description of the program (scale, financing, etc.), as well as several performance indicators to understand the program's impact, summary measures on the quality of the evaluation evidence and the quality of the intervention (described below), and sources for further information on the intervention. To allow for quantitative analysis of the data, variables included in the template were coded on the basis of multiple choice measures wherever feasible.

Database

In the project design stage, a decision was made to organize the data collection so that the inventory could be in an electronic format in order to facilitate search capabilities, updating, and to do quantitative analysis. The template was built into an Excel worksheet and an independent machine-readable file was created for each intervention included in the inventory. After the data-collection phase ended, an Excel macro was designed on Microsoft Visual Basic to read every file and construct a searchable database where the number of observations (rows) matched the number of interventions (files or worksheets). Data collected in the questionnaire – both plain text and codes -- are displayed in the columns, creating a database of program-specific information (Database 1).

Simultaneously, a database of country-specific information (Database 2) was created to contextualize the economic conditions of the country. This information includes level of development, level of income, and a characterization of the labor market regulatory/institutional situation. Sources of information for the country database are the World Development Indicators and the Doing Business Database. The Excel macro links databases 1 and 2 through a common key-variable, namely country name, creating a comprehensive database for the analysis of the inventory.

The database can be accessed at <http://go.worldbank.org/48Z06GMD70>

Two critical variables figure prominently in the analysis of lessons learned from the inventory. These are the *quality of intervention* (QOI) and the *quality of evaluation* (QOE). The QOI is the measure of program effectiveness, whose possible values are described in Table 2. The primary performance indicators that are considered in establishing a QOI rating are the effects of the program on the employment and earnings of participants. At one level, the QOI value can be used to identify impact – i.e., to distinguish those programs that actually help participants in the labor market (QOI = 1, 2, or 3) from those that appear to have no effect, or even a negative effect (QOI = 0). This distinction is similar, but not identical, to determining program success. To be specific, interventions can have a positive employment impact but are not cost-effective (i.e., QOI = 1). These programs cannot be considered successful.

Assessing the quality of the intervention relies greatly on the information quality, which varies widely from basic descriptive information to solid evaluation evidence. The QOE variable identifies the evaluative basis for assessing program quality. The QOE measure is described in Table 3. With this variable, then, assessments of the effectiveness of interventions can be judged with knowledge of the underlying evidence. For example, one can consider only those programs that meet the most exacting burden of proof (i.e., QOE = 3), with the tradeoff that sample size will be reduced. On the other hand, accepting a less demanding basis of evidence will increase the pool of programs under consideration, but at the expense of rigor. As we will see later in this paper, assessments of program impact are correlated with the quality of the evaluation evidence.

Table 2: Measuring the Quality of the Intervention (QOI)

QOI value	Description
0	Intervention had negative or zero impact on labor market outcomes.
1	Intervention had positive impact on labor market outcomes but is not cost-effective.
2	Intervention had positive impact on labor market outcomes and there is no evidence on costs.
3	Intervention had positive impact on labor market outcomes and is cost-effective.
99	Missing value. Not enough evidence to make an assessment.

Table 3: Measuring the Quality of Evaluation (QOE)

QOE value	Description
0	Intervention has no evaluation information available on outcomes or impact.
1	Evaluation includes basic information on the gross outcomes of the intervention (e.g. number of participants/ young people who found a job after the intervention, improvement in earnings of participants) without considering net effects (i.e., there is no control group).
2	Evaluation includes estimate of net impact on, e.g., employment and earnings in the labor market (using control groups to measure impact) but no cost-benefit analysis.
3	Evaluation includes net impact plus cost-benefit analysis.

2. Coverage of the Youth Employment Inventory

What are the most popular employment interventions for youth and where have been implemented? Is there any particular orientation of the interventions towards disadvantaged youth? How are youth employment interventions being financed? This section approaches these questions based on the evidence from the inventory.

First, regarding coverage of the inventory by category of intervention, skills training programs – particularly vocational training and apprenticeships systems – and comprehensive programs¹⁰ are the most popular interventions supporting the entry of young people into the labor market. They account for 38 and 33 percent of all interventions covered by the inventory, respectively (Table 4). Other prevalent categories are interventions to make the labor market

¹⁰ Comprehensive interventions have in general very strong training components. They encompass job and life skills training (in classroom and/or on-the-job), apprenticeship systems, entrepreneurship schemes, information, counseling, placement, financial incentives, and other support services.

work better for young people and entrepreneurship schemes. All other intervention categories have very small samples.¹¹

Across regions, most programs covered by the inventory were implemented in OECD countries (42 percent) and Latin America and the Caribbean (24 percent). Europe and Central Asia, Sub-Saharan Africa, South and East Asia and the Pacific, and Middle East and North Africa had relatively low participation in the total number of documented interventions. Advanced and middle income countries dominate the picture with nearly 90 percent of collected interventions carried out in these countries. The distribution of interventions varies by the countries' income level. While OECD countries tend to have a more diverse and comprehensive set of policies and programs to approach youth, middle income countries have focused greatly on training systems, and low income countries in entrepreneurship schemes.

Second, regarding targeting and the orientation of the programs toward disadvantaged groups, the inventory shows that most interventions are addressed to youth only while nearly 20 percent are open to people of all ages. Rural youth are targeted but received relatively less services than urban youth: only 10 percent of interventions served exclusively rural populations, 28 percent were implemented in the cities and the remaining 62 percent worked in both, rural and urban areas.

To measure the extent to which interventions are oriented towards the disadvantaged, the inventory identified the following target groups: women, the disabled, ethnic groups, and individuals with certain income and education levels. The incidence of gender targeting is relatively low: only 16 percent of all programs are oriented towards young women and 2 percent are targeted explicitly at young men. There are 32 programs (11 percent) for disabled young people, most of which provide comprehensive services. Regarding ethnicity, there is only a small number of interventions (20 in total, with 9 of these providing training) targeted towards particular ethnic groups.

Just over half (52 percent) of all programs in the inventory are oriented towards low-income youth. Training programs and comprehensive programs are especially likely to be targeted on income. Regarding education, 49 percent of all programs were targeted to youth with low educational attainment. Comprehensive programs are again most likely to have education-related targeting. Nine programs in OECD countries were found to assist educated youth.

Finally, regarding financing, the major sponsor of youth employment programs is the government, but joint public-private ventures with international organizations and bilateral donors play an important role in the delivery of the programs nevertheless. About 56 percent of programs are primarily government-funded, and 33 percent are financed by a mix of institutions, such as central and local governments, international organizations, bilateral donors, civil society and the private sector. Government-funded programs provide most comprehensive interventions as well as wage subsidies, financial incentives for training and public works programs. Public programs are more prevalent in advanced economies and more likely to have rigorous evaluation information and net impact studies than programs sponsored by other institutions. On the other hand, the provision of training programs and entrepreneurship schemes in developing countries relies greatly on donors' resources and the participation of non-government organizations.

¹¹ There is no evidence on programs related to anti-discrimination legislation, young adult literacy, counteracting location barriers, or programs to promote overseas employment of young people.

Table 4: Coverage of the Inventory by Category of Intervention and Region

Category of intervention	Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	OECD	South and East Asia and the Pacific	Sub-Saharan Africa	Total	%
1. Making the labor market work better for young people	13	3	1	17	0	1	35	12%
1a. counseling, job search skills	2	1		3			6	
1b. wage subsidies	8			9			17	
1c. public works programs	3		1	3		1	8	
1d. anti-discrimination legislation							0	
1e. other		2		2			4	
2. Improving chances for young entrepreneurs	3	5	1	11	6	7	33	11%
3. Skills training for young people	18	38	2	38	9	6	111	38%
3a. vocational training including apprenticeship systems	13	36	2	33	8	6	98	
3b. literacy & numeracy – young adult literacy programs							0	
3c. 2nd chance & equivalency programs	3	1		3	1		8	
3d. other	2	1		2			5	
4. Making training systems work better for young people	0	0	0	6	1	4	11	4%
4a. information				1		2	3	
4b. credit (to individuals or enterprises)				1			1	
4c. financial incentives (subsidies, vouchers)				2	1	1	4	
4d. other				2		1	3	
5. Programs to counteract residential segregation of disadvantaged young people	0	0	0	0	0	0	0	0%
5a. transportation							0	
5b. others							0	
6. Improving labor market regulations to the benefit of young people				1	1		2	1%
7. Programs for overseas employment of young people							0	0%
8. Comprehensive interventions	6	22	4	47	4	11	94	33%
9. Other (e.g. voluntary national service programs)	1			2			3	1%
Total	41	68	8	122	21	29	289	100%

3. Main Findings from the Youth Employment Inventory: A qualitative analysis

This section summarizes the main findings from the inventory based on descriptive statistics and an overview analysis of interventions across regions.¹² Important evidence-based lessons are also extracted from interventions with impact evaluations – i.e. those with treatment and control groups (Puerto, 2007b).

Training is the dominant form of intervention used to integrate young people into the labor market. The inventory shows that interventions with strong training components, specifically skills training and comprehensive programs are the most common approach to youth. These programs are highly popular in OECD and middle income countries (particularly in Latin America) and account for more than two-thirds of interventions around the world. This suggests a clear interest in building up skills development strategies to counteract un/under-employment among the young.

Interventions are often targeted at low-income or poorly-educated young people, particularly in non-developed countries. A vast majority of interventions in Latin America focus on young people from low income families, while educated youth are most targeted in Europe and Central Asia. Ratings on the quality of the interventions indicate that programs targeting low-income and low-educated youth have higher possibilities of yielding positive labor market impact than programs without any income or education orientation.

The overall evaluation evidence on youth employment interventions is weak. Only one-fourth of interventions in the inventory have estimates of net impact, and just one in ten has evidence on cost-effectiveness (Table 5). Available information on youth employment interventions is actually stronger in developed countries. Industrialized economies have a long-standing experience in the implementation of ALMP, which has grown over time with the need for documenting and analyzing program effects. Nearly 60 percent of interventions with net impact evaluations were carried out in OECD economies.

Table 5: Coverage of the Inventory by Quality of Intervention and Quality of Evaluation

Quality of Evaluation (QOE)	Quality of Intervention (QOI)					Total
	0	1	2	3	99	
0					114	114
1	9	2	85	3	3	102
2	22	1	21	1		45
3	7	8	3	10		28
Total	38	11	109	14	117	289

Note: See Tables 2 and 3 for definitions on Quality of Intervention and Evaluation.

¹² Further information on regional analyses can be found in the inventory's regional reports: Rother and Puerto (2007) for OECD countries, Stavreska (2006a) for Eastern Europe and Central Asia, Puerto (2007a) for Latin America and the Caribbean, Rother (2006) for Sub-Saharan Africa, and Stavreska (2006b) for Asia, including both East Asia and the Pacific and South Asia.

The assessed impact of an intervention is affected by the quality of the underlying evaluation evidence. As displayed in Table 5, when information on gross outcomes only is available (QOE = 1), 90 of the 99 programs where a QOI assessment was made were judged to have positive impact¹³. However, when a net impact evaluation has been carried out (QOE = 2 or 3), the possibility of finding a positive employment impact decreases significantly, to 60 percent (44 out of 73 programs). This has two important implications:

- First, within the context of this study, an **overall assessment of what interventions can do for the employment and earnings of young workers is much more favorable when the standard of acceptable evidence is relatively light than when a higher standard is set** (i.e., net impact evaluation).
- Second, because of the lack of serious evaluations especially in developing countries, **policy-makers – who tend to focus on gross outcome measures – are generally overestimating how useful their interventions are in helping young people find employment or increasing their earnings.**

When cost-effectiveness is taken into account, only about one-third of interventions are successful (10 out of 28 programs where QOE = 3).¹⁴ This finding holds even after interpolating programs for which there is no evidence available on cost. This simulation exercise assumes that the programs without cost information have the same probability of being cost effective as programs with cost information. The final outcome indicates that **33.2 percent of interventions with net impact evaluation are successful (i.e. have positive impact and are cost-effective).**

Across categories of intervention, a descriptive analysis of the interventions with impact evaluations (i.e. with treatment and control groups) shows relatively better results from (i) interventions to make the labor market work better for youth, (ii) comprehensive programs, and (iii) entrepreneurship schemes. Training programs were less successful than average, with relatively better outcomes in developing and transition economies than in developed countries. Sample sizes in these categories become a consideration especially when we impose the restriction that programs need to have an impact evaluation.¹⁵ Applying the simulation exercise mentioned before, the results show relatively little variation across programs.

Intervention quality ratings indicate that the **labor market impact tends to be more favorable in developing and transition countries than in industrialized countries.** While only 60 percent of programs in the OECD region had positive impact, the corresponding rates in Europe and Central Asia and Latin America and the Caribbean – the two other regions with significant samples – were 90 percent and 92 percent, respectively. Although the sample sizes are too small in South and East Asia and the Pacific and Sub-Saharan Africa to draw firm conclusions, the limited evidence in these regions offers additional support for the conclusion that youth programs have been more successful in developing countries. This is a surprising finding, given the more extensive experience of OECD countries with employment programs, their greater capacity and resources, more available information and analysis, and generally better functioning labor markets.

¹³ Where impact is defined as the effect of the programs on the future employment prospects of participants, measured by post-program employment and/or earnings.

¹⁴ Where success implies having a positive impact on the labor market prospects of participants and being cost-effective (i.e. costs do not exceed benefits).

¹⁵ The number of interventions with impact evaluations ranges from 34 for comprehensive programs to just 3 for entrepreneurship.

Three hypotheses have been drawn in the analysis to explain why industrialized economies have less effective interventions. The *first hypothesis* suggests that this finding is due to measurement problems stemming from the fact that programs in industrialized countries tend to be more rigorously evaluated. The data show that the availability of net impact evaluation reduces the frequency of positive employment effects, so it may be that the differences are due to the fact that the positive impact of interventions in non-industrialized countries is overstated and that, if they were evaluated as rigorously as OECD programs, these differences would disappear. This hypothesis will be challenge by the meta-analysis model in section D.

A *second hypothesis* implies real differences between youths in developed and non-developed countries. For example, disadvantaged young people – the dominant clientele of youth programs everywhere – may be at such a deficit in OECD countries given the high levels of human capital and the skills intensity of their labor demand that employment interventions are simply not enough to compensate. In developing countries, on the other hand, the boost that these programs can give young people may be enough to give them the advantage they need to improve their situation. Unfortunately, this hypothesis cannot be tested with the information collected by the inventory.

At last, the finding might be explained by differences in labor market institutions or policies, such as employment protection laws (EPL). In general, employment protection rules could limit the effectiveness of youth programs since it is well documented that, where such rules are strict, young people are likely to experience difficulty in entering the labor market (e.g., OECD 2004). Within the OECD region, the evidence is at least consistent with this hypothesis: youth programs have a higher positive impact rate in Anglo-Saxon countries (74 percent) where EPL is more flexible than in the rest of the OECD, largely continental Europe (38 percent) where rules are more protective.¹⁶ However, observing the World Bank’s *Doing Business* “employment rigidity index” there is no clear pattern between flexibility and program effectiveness. In fact, developed countries have some of the highest levels of labor market flexibility around the world.

D. Meta-analysis of the Youth Employment Inventory: A quantitative analysis

1. Dataset and variables

The inventory offers a considerable collection of studies and programs with estimates on labor market outcomes on youth. In order to draw data-supported lessons on what can potentially work for this target population, the sample of interventions collected by the inventory has been filtered to examine only those programs with relatively good quality of evaluation, i.e. QOE = 1, 2, and 3. This set of programs ensures enough evidence to assess the quality of the interventions and determine whether the programs yield a positive impact on employability and earnings of participants. As a result, the dataset for the meta-analysis starts with a sub-sample of 172 observations of interventions where an assessment was made, including both programs with gross outcomes and with net impact evaluations, as shown in Table 6. A subsequent specification constrains the sample to interventions with only net impact evaluations, i.e. QOE = 2 and 3.

Based on the measures of intervention quality (or QOI, described in detail in Section B), a binomial variable has been constructed to identify the occurrence of positive labor market impacts. This is the dependent variable of the model, which will measure the probability of

¹⁶ The differences between the Anglo-Saxon and continental Europe (and other) countries in the OECD are discussed in some detail in the OECD regional paper (Rother and Puerto, 2007).

program success. The variable takes value 1 in 78 percent of the sample (134 cases), where the evaluation reported positive effects on employability and/or earnings of beneficiaries (i.e. QOI = 1, 2, or 3), and value 0 in the remaining 22 percent of observations (38 cases) where negative or zero outcomes were reported (QOI = 0). As can be seen, the construction of the dependent variable offers no consideration on cost-effectiveness, given the substantial lack of evidence on cost benefit analyses.

**Table 6: Sample of interventions for the Meta-analysis
(QOE = 1, 2, 3 and QOI ≠ 99)**

Quality of Evaluation (QOE)	Quality of Intervention (QOI)				Total
	0	1	2	3	
1	9	2	85	3	99
2	22	1	21	1	45
3	7	8	3	10	28
Total	38	11	109	14	172

Note: See Tables 2 and 3 for definitions on Quality of Intervention and Evaluation.

Explanatory variables are part of four groups: (i) category of intervention, (ii) evaluation quality, (iii) economic and institutional country context, and (iv) specific characteristics of the program.

Category of Intervention. Given the evaluation evidence available in the set of 172 programs, categories of intervention have been clustered in five types, as displayed in Table 7. Program type 1 comprises interventions to make the labor market work better; this corresponds to 15 percent of the sample. Program type 2 includes all entrepreneurship schemes, prevalent in 9 percent of the cases. Training-related interventions (i.e. categories 3 and 4) are clustered under program type 3, with 36 percent of the cases. The highest participation is that of comprehensive programs, classified under program type 4. The last type clusters the remaining categories (categories 6 and 9) with fairly low evidence on outcomes. *Program types are introduced in the model as five independent dummy variables, where training-related programs represented the omitted category.*

Table 7: Classification of categories of intervention by labor market impact

Type	Category of intervention	Negative or Zero impact	Positive impact	Total
1	1. Making the labor market work better for young people	5	21	26
2	2. Improving chances for young entrepreneurs	0	15	15
3	3. Skills training for young people	13	45	58
	4. Making training systems work better for young people	2	2	4
4	8. Comprehensive approach	18	47	65
5	6. Improving labor market regulations	0	1	1
	9. Other	0	3	3
Total		38	134	172

Evaluation Quality. Within the sample of interventions with evaluation evidence a further distinction has been drawn between evaluations with only gross outcomes (i.e. QOE = 1) and net impact evaluations (i.e. QOE = 2 and 3). This classification seeks to test whether the type of evaluation affects the reported labor market outcomes.

Economic and institutional country context. Country characteristics have been also considered in other analyses (e.g. Kluge (2006) and Greenberg *et al.* (2003)) to capture the effect of the macroeconomic conditions and labor market regulations on labor market outcomes. We distinguish between developed and non-developed economies, in order to test whether youth employment programs are affected by the country's income level. About 58 percent of evaluated interventions took place in non-developed countries (Annex, Table A.1). In addition, we use the rigidity of employment index (as reported by the Doing Business Report) to measure the effect of employment regulations flexibility on youth's labor market prospects. The rigidity of employment index is a composite measure of difficulty of hiring and firing, and rigidity of hours. The higher the index, the more stringent the labor market regulations are. It varies from 26.2 in East Asia and the Pacific to 53.1 in Sub-Saharan Africa. The OECD area reported the second lowest level with a rating of 35.8.

Specific characteristics of the program refer mainly to the features of the target population, in particular the programs' specific focus on women, the disabled, specific ethnic groups, as well as low-income and low-educated youth. Dummy variables were created for each target group to test whether focalization strategies on the most disadvantaged allow better outcomes (Annex, Table A.2). The model also controls by participant's age, given that some programs (30 percent in this sample) were also open to workers of all ages.

Additional program controls include: the decade when the intervention was first implemented and the current status of the program. Most interventions, nearly 72 percent, have been implemented during the nineties and 2000s, and over 60 percent are already completed. The location of the program in rural and urban areas has also been considered in the model. The data shows that just about one third of the interventions have specifically targeted a certain area, while two thirds of the interventions are implemented in both rural and urban settings. The last variable under consideration is the programs' primary source of financing, which takes value 1 for government-sponsored interventions (two thirds of observations) and 0 for others.

2. The model and results

This analysis uses a probit model to estimate the probability that a certain youth employment program yields positive impacts on the labor market outcomes of its beneficiaries. Probit is a binary choice model that estimates the probability of an event as a function of a set of attributes, assuming a normal distribution in the data. Following Hayashi (2000), in the probit model, a scalar dependent variable y_t is a binary variable, $y_t \in \{0, 1\}$. In our case $y_t = 1$ indicates that a certain program reported positive labor market impacts on youth, while $y_t = 0$ indicates negative or zero impacts. This event is determined by a vector of regressors \mathbf{x}_t , namely category of intervention, evaluation quality, country characteristics and program characteristics. As a result, the conditional probability of y_t given \mathbf{x}_t is given by

$$\begin{cases} f(y_t = 1 | \mathbf{x}_t; \boldsymbol{\beta}_0) = \Phi(\mathbf{x}_t' \boldsymbol{\beta}_0), \\ f(y_t = 0 | \mathbf{x}_t; \boldsymbol{\beta}_0) = 1 - \Phi(\mathbf{x}_t' \boldsymbol{\beta}_0), \end{cases}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. Given the binary features of y_t , this can be written compactly as

$$f(y_t | \mathbf{x}_t; \boldsymbol{\beta}_0) = \Phi(\mathbf{x}_t' \boldsymbol{\beta}_0)^{y_t} [1 - \Phi(\mathbf{x}_t' \boldsymbol{\beta}_0)]^{1-y_t}$$

Table 8 reports the results of two model specifications. Specification 1 includes the larger sample where QOE = 1, 2, 3, and Specification 2 restricts the sample to net impact evaluations, i.e., QOE = 2, 3. The explanatory variables are the same, with the exception of the quality of evaluation variable which is not needed in the second specification. Marginal effects are displayed for each variable. These *marginal effects report the change in the probability of a positive program impact for an infinitesimal change in each independent continuous variable or for a discrete change in the case of dummy variables*. The computation of marginal effects fits maximum-likelihood probit models, where the maximum-likelihood estimator of $\boldsymbol{\beta}_0$ is given by the function $m(\mathbf{w}_t; \boldsymbol{\beta}) = \log f(y_t | \mathbf{x}_t; \boldsymbol{\beta}) = y_t \log \Phi(\mathbf{x}_t' \boldsymbol{\beta}) + (1 - y_t) \log [1 - \Phi(\mathbf{x}_t' \boldsymbol{\beta})]$, where \mathbf{w}_t is the t-th observation in the dataset. The models' estimated coefficients on which the marginal effects are based are presented in the Annex, Table A.3.¹⁷

On the first set of variables regarding category of intervention, the estimates suggest **there are no statistically significant differences among program types** – labor market work better and comprehensive programs – in terms of the likelihood that they deliver positive impact on the labor market, relative to training programs. This result holds for both specifications. The unclear pattern of success across categories of intervention was also reported by Heckman *et al.* (1999) for a sample of OECD programs. In the estimation process, two categories (entrepreneurship and others) were dropped due to collinearity effects of their small sample size on the predicted variable¹⁸.

On the other hand, the meta-analysis confirms that **evaluation quality matters**. This is shown in Specification 1, where the statistically significant negative coefficient for quality of evaluation variable indicates that assessments of program impact are likely more negative when proper net impact studies have been carried out. Having a net impact evaluation reduces the likelihood of success by 35 percentage points (compared to the event of not having a net impact evaluation). This reflects an over optimistic reading of results from evaluations with gross outcomes, and emphasizes the importance of conducting rigorous evaluations to capture the real effects of programs. This result supports the qualitative finding discussed in Section C according to which the assessed impact of an intervention is affected by the quality of the underlying evaluation evidence.

¹⁷ A logit model was also estimated to test whether a logistic distribution better fitted the data than a normal distribution. The logit regression reported very similar estimates than the probit. Results are shown in the Annex, Table A.4.

¹⁸ It is worth noticing the changes in sample size once all variables are considered into play. Specification 1 goes from 172 observations to 95, while Specification 2 goes from 73 to 59. This is mainly due to missing values in the programs' characteristics, particularly targeting on gender, disabled youth, and specific ethnic groups.

Table 8: Probit model reporting marginal effects of youth employment programs

	Specification 1 QOE=1, 2, 3		Specification 2 QOE = 2, 3	
	Marginal effect	z-stat ⁶	Marginal effect	z-stat ⁶
<i>Category of intervention¹</i>				
Labor market work better	-0.032	-0.19	0.011	0.04
Comprehensive	-0.124	-1	-0.312	-1.41
<i>Quality of the evaluation²</i>				
Net impact evaluation	-0.347	-2.53 *		
<i>Economic and institutional country context³</i>				
Non-developed countries	0.527	2.77 **	0.791	2.61 **
Rigidity of employment index	-0.013	-2.88 **	-0.021	-2.48 *
<i>Specific characteristics of the program</i>				
Time period and status⁴				
Program implemented before the nineties	-0.422	-2.36 *	-0.539	-1.7
Completed programs	-0.348	-3.02 **	-0.683	-2.82 **
Targeting⁵				
Programs target only youths	-0.121	-1.11	-0.204	-0.92
Programs located in specific areas	-0.328	-1.87	-0.549	-1.84
Programs focus on women	-0.125	-0.75	-0.172	-0.71
Programs focus on specific ethnic groups	0.152	0.77	0.312	0.7
Programs focus on poor youth	0.47	2.33 *	0.753	2.21 *
Programs focus on low-educated youth	-0.232	-1.41	-0.539	-1.56
Financing				
Government-sponsored	-0.107	-0.55	0.597	1.48
	Observations = 95		Observations = 59	
	Pseudo R2 = 0.46		Pseudo R2 = 0.42	

Notes:

1. Training-related programs (including skills training and programs to make the training systems work better) are the omitted category. 2. Programs with evaluations reporting only gross outcomes are the omitted category. 3. Developed countries are the omitted category. The rigidity of employment index is a continuous variable. 4. On decade of implementation, programs implemented during the nineties and 2000s are the omitted category. On current status of the interventions, ongoing programs are the omitted category. 5. Omitted categories on targeting reflect none specific orientation toward disadvantage people within those groups. 6. The z-statistics test the null hypothesis of a zero coefficient following a standard normal distribution.

The values of the z-statistics are reported in the third column: * significant at 5%; ** significant at 1%.

Economic and institutional country context variables have highly significant effects on program impact. The regressions show that youth employment programs are more effective in developing countries and transition countries than in developed economies. The likelihood of success is between 53-79 percentage points (depending on specification) higher when the program is implemented in a developing or transitional setting than when it is carried out in a developed country. Given that the quality of evaluation is controlled for, this result

cannot be explained by the fact that impact evidence is more rigorous in developed countries. Another possible explanation, which cannot be tested here, is that the skills disadvantage of participants in developed countries may often be too large to be tackled by these programs; while in developing countries where skills are scarcer, programs may provide enough of a boost to make a measurable difference.

Likewise, the models suggest that **interventions for youth tend to be more successful in countries with higher labor market flexibility**. In particular, the higher the rigidity of employment index the lower the probability of obtaining positive labor market impacts. Note that the size of the coefficient in both specifications is very small and, while the effect may be statistically significant, the importance seems relatively minor.

Among program characteristics, the period of implementation and the current status of the program have significant effects on the probability of success. First, although statistical significance is borderline, the models suggest **there is a learning process, where programs developed during the 1990s and after tend to yield better outcomes than older programs**. This is the case in Latin America where there has been a move towards demand-oriented programs that match effectively the needs of the productive sector, as well as open participation of the private sector and other agents in the provision and financing of programs. Second, both specifications indicate that **ongoing programs perform better than completed programs**.

In terms of the programs orientation toward particular groups, of all characteristics of the target population included in the model, only income level explains with statistical significance program success. Programs targeting low-income youth perform significantly better than programs without this orientation, suggesting that **interventions do have promise for improving the labor market situation of low-income young people**. Other considerations toward a particular gender, the disabled, specific ethnic groups, and youth with low education levels do not affect the outcomes. Similarly, the model tested whether publicly-funded programs perform better than otherwise, but the marginal effect of source of financing lacked statistical significance.

Results of Specification 1 were tested to ensure the best fit of the model and to rule out the possibility of outliers. The first test checked the stability of the explanatory power in a smaller sample. After splitting the sample randomly in two, the R-squared increases slightly from .46 to .51, suggesting a steady fit in the model. Marginal effects of this model are reported in Table A.5, Annex. An additional test was performed to ensure the stability of the explanatory power by ruling out the possibility of outliers. Specification 1 is run iteratively by sequentially and randomly dropping one observation with replacement. Ninety-five models resulted from this exercise, and the R-squared reported ranged from .45 to .51 (Figure A.1, Annex) verifying the stability of the specification's best fit, and eliminating the possibility of outliers.

E. Conclusions

The Youth Employment Inventory (YEI) is based on available documentation of current and past programs and includes evidence from 289 studies of interventions from 84 countries in all regions of the world. These studies have been analyzed based on the evaluation evidence available in order to (i) document the types of programs that have been implemented to support young workers to find work, and (ii) identify what appears to succeed in terms of improving employment outcomes for youth.

This paper uses a meta-analytical framework to examine simultaneously all interventions collected by the inventory with evaluation evidence on labor market outcomes. It employs a probit model to estimate the probability of obtaining positive program impacts based on the type of intervention, quality of the evaluation evidence, characteristics of the programs, and country-specific characteristics.

Empirical results from a sample of 172 evaluated studies – including net impact evaluations and evaluations with gross outcomes - indicate that **program success is not determined by the type of intervention but rather by the program’s targeting strategies toward disadvantaged youth, the country level of development and the flexibility of the labor market regulations**. Data from the youth employment inventory shows there are no major differences across types of interventions in terms of impact, suggesting that policy-makers should consider which type of intervention best addresses the problem of concern. This unclear pattern was also found by Heckman *et al.* (1999) when analyzing active labor market measures in the OECD area. However, it differs from Kluve’s (2006) findings according to which some interventions work in fact better than others.

In addition, the meta-analysis of the inventory shows that country context matters when assessing the impact of youth employment programs. An employment program implemented in a developing or transitional country has at least a 50 percent higher probability of yielding positive impact for youth than a developed-country program. The analysis proved this is not a measurement problem, since the estimates hold even when the sample is constrained to studies with net impact evaluation. Other explanations may come into play, such as the human capital gap between these two groups of countries. Labor market institutions appear to have small but significant effects on program impact. The model shows that less flexible employment protection rules, measured by the rigidity of employment, slightly lower the probability of obtaining positive outcomes from youth employment programs.

Certain characteristics of the programs show interesting effects. Ongoing programs and those carried out since the 1990s have significantly better performance than earlier interventions, suggesting a potential learning process. Moreover, targeting interventions on economically disadvantaged youth appears to have substantial positive impact on participants’ labor market prospects. Sensitivity tests show these results are stable under different specifications, particularly when the sample size is constrained to studies with net impact evaluations.

References.

- Betcherman, Gordon, Karina Olivas, and Amit Dar. 2004. "Impact of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and transition Countries." Washington, D.C.: World Bank, Social Protection Discussion Paper Series 0402.
- Betcherman, Gordon; Martin Godfrey; Olga Susana Puerto; Friederike Rother; and Antoneta Stavreska. 2007. "Global Inventory of Interventions to Support Young Workers: Synthesis Report." Preliminary draft. Washington, D.C.: World Bank.
- Dar, Amit and P. Zafiris Tzannatos. 1999. "Active Labor Market Programs: A Review of the Evidence from Evaluations," Social Protection Discussion Paper no. 9901, January. The World Bank. Washington, D.C.
- Gay, Robert, and Michael Borus. 1980. "Validating performance indicators for employment and training programs," *Journal of Human Resources*. Winter 1980, 15, 29-48.
- Greenberg, David H.; Charles Michalopoulos; Philip K. Robins. 2003. "A Meta-Analysis of Government-Sponsored Training Programs". *Industrial & Labor Relations Review*. Volume 57, Issue 1 2003 Article 2.
- Hayashi, Fumio. 2000. *Econometrics*, Princeton University Press.
- Heckman, J.J., R.J. LaLonde and J.A. Smith (1999), "The economics and econometrics of active labour market programs", in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* 3, Elsevier, Amsterdam.
- Hunt, Morton. 1997. *How Science Takes Stock: The Story of Meta-Analysis*. New York: Russell Sage Foundation.
- Jarrell, S. B., and T. D. Stanley. 1990. "A Meta-Analysis of the Union-Nonunion Wage Gap." *Industrial and Labor Relations Review*, Vol. 40, No. 2 (January), pp. 54-67.
- Kluve, J. 2006. "The Effectiveness of European Active Labor Market Policy", IZA Discussion Paper, No. 2018, Bonn.
- Kluve, J. and C.M. Schmidt (2002), "Can training and employment subsidies combat European unemployment?", *Economic Policy* 35, 409-448.
- Martin, J.P. and D. Grubb (2001), "What works and for whom: a review of OECD countries' experiences with active labour market policies", IFAU Working Paper 2001:14.
- Puerto, Olga Susana. 2007a. *Interventions to Support Young Workers in Latin America and the Caribbean*. World Bank: Washington D.C.
- Puerto, Olga Susana. 2007b. *Learning from International Experiences, The Youth Employment Inventory; Background paper for the Sierra Leone Youth and Employment ESW*, World Bank, Washington D.C.

Rother, Friederike and Olga Susana Puerto. 2007. Interventions to Support Young Workers in OECD countries. World Bank: Washington D.C.

Rother, Friederike. 2007. Interventions to Support Young Workers in Sub-Saharan Africa. World Bank: Washington D.C.

Sirmans, G. Stacy; Lynn MacDonald; David A. Macpherson; Emily Norman Zietz. 2006. The Value of Housing Characteristics: A Meta Analysis. The Journal of Real Estate Finance and Economics. Volume 33, Number 3 / November, 2006

Stavreska, Antoneta. 2006a. Europe and Central Asia, Youth Employment Inventory Summary Report. World Bank: Washington D.C.

Stavreska, Antoneta. 2006b. Interventions to Support Young Workers in South and East Asia and the Pacific. World Bank: Washington D.C.

Annex.

**Table A.1: Classification of countries' level of development by labor market impact
(for a sample of programs with QOE=1, 2, 3)**

	Negative or Zero impact	Positive impact	Total
Developing and Transition Countries	9	91	100
OECD Countries	29	43	72
Total	38	134	172

**Table A.2: Number of interventions targeting disadvantaged youths by labor market
impact
(for a sample of programs with QOE=1, 2, 3)**

	Negative or Zero impact	Positive impact	Total	%
Women	6	24	30	17%
Disabled	1	13	14	8%
Ethnicity	1	9	10	6%
Income	18	78	96	56%
Education	23	81	104	60%

Table A.3: Probit model: simple coefficients (Table 8)

	Specification 1 QOE=1, 2, 3		Specification 2 QOE = 2, 3	
	Marginal effect	z-stat ⁶	Marginal effect	z-stat ⁶
Category of intervention¹				
Labor market work better	-0.118	-0.19	0.029	0.04
Comprehensive	-0.464	-1	-0.811	-1.41
Quality of the evaluation²				
Net impact evaluation	-1.586	-2.53 *		
Economic and institutional country context³				
Non-developed countries	2.149	2.77 **	2.808	2.61 **
Rigidity of employment index	-0.051	-2.88 **	-0.053	-2.48 *
Specific characteristics of the program				
Time period and status⁴				
Program implemented before the nineties	-1.438	-2.36 *	-1.484	-1.7
Completed programs	-1.848	-3.02 **	-2.441	-2.82 **
Targeting⁵				
Programs target only youths	-0.500	-1.11	-0.528	-0.92
Programs located in specific areas	-1.129	-1.87	-1.538	-1.84
Programs focus on women	-0.426	-0.75	-0.435	-0.71
Programs focus on specific ethnic groups	0.896	0.77	0.983	0.7
Programs focus on poor youth	1.583	2.33 *	2.359	2.21 *
Programs focus on low-educated youth	-0.982	-1.41	-1.769	-1.56
Financing				
Government-sponsored	-0.459	-0.55	2.184	1.48
Constant	5.120	3.15 **	1.609	0.83
	Observations = 95		Observations = 59	
	Pseudo R2 = 0.46		Pseudo R2 = 0.42	

Notes:

1. Training-related programs (including skills training and programs to make the training systems work better) are the omitted category. 2. Programs with evaluations reporting only gross outcomes are the omitted category. 3. Developed countries are the omitted category. The rigidity of employment index is a continuous variable. 4. On decade of implementation, programs implemented during the nineties and 2000s are the omitted category. On current status of the interventions, ongoing programs are the omitted category. 5. Omitted categories on targeting reflect none specific orientation toward disadvantage people within those groups. 6. The z-statistics test the null hypothesis of a zero coefficient following a standard normal distribution.

The values of the z-statistics are reported in the third column: * significant at 5%; ** significant at 1%.

Table A.4: Logit model: using Specification 1

	Coefficient	z-stat⁶	
Category of intervention¹			
Labor market work better	-0.171	-0.16	
Comprehensive	-0.687	-0.87	
Quality of the evaluation²			
Net impact evaluation	-2.737	-2.42	*
Economic and institutional country context³			
Non-developed countries	3.592	2.59	**
Rigidity of employment index	-0.086	-2.75	**
Specific characteristics of the program			
Time period and status⁴			
Program implemented before the nineties	-2.472	-2.25	*
Completed programs	-3.106	-2.89	**
Targeting⁵			
Programs target only youths	-0.900	-1.16	
Programs located in specific areas	-1.849	-1.76	
Programs focus on women	-0.605	-0.62	
Programs focus on specific ethnic groups	1.585	0.84	
Programs focus on poor youth	2.687	2.29	*
Programs focus on low-educated youth	-1.623	-1.35	
Financing			
Government-sponsored	-0.934	-0.61	
Constant	8.792	3.06	**
Observations = 95 ; Pseudo R ² = 0.4515			

Notes:

1. Training-related programs (including skills training and programs to make the training systems work better) are the omitted category. 2. Programs with evaluations reporting only gross outcomes are the omitted category. 3. Developed countries are the omitted category. The rigidity of employment index is a continuous variable. 4. On decade of implementation, programs implemented during the nineties and 2000s are the omitted category. On current status of the interventions, ongoing programs are the omitted category. 5. Omitted categories on targeting reflect none specific orientation toward disadvantage people within those groups. 6. The z-statistics test the null hypothesis of a zero coefficient following a standard normal distribution.

The values of the z-statistics are reported in the third column: * significant at 5%; ** significant at 1%.

Table A.5: Probit model: Specification 1 dropping randomly 50 percent of the sample

	Marginal effect	z-stat ⁶
<i>Category of intervention</i> ¹		
Labor market work better	-0.017	-0.79
Comprehensive	-0.005	-1.28
<i>Quality of the evaluation</i> ²		
Net impact evaluation	-0.010	-1.62
<i>Economic and institutional country context</i> ³		
Non-developed countries	0.594	1.46
Rigidity of employment index	0.000	-1.53
<i>Specific characteristics of the program</i>		
Time period and status ⁴		
Program implemented before the nineties	-0.004	-0.69
Completed programs	-0.001	-0.38
Targeting ⁵		
Programs target only youths	0.004	0.93
Programs located in specific areas	0.001	0.47
Programs focus on women	-0.164	-1.99 *
Programs focus on poor youth	0.019	0.84
Programs focus on low-educated youth	-0.274	-1.1
Financing		
Government-sponsored	0.000	0.05
Observations = 47 ; Pseudo R ² = 0.5112		

Notes:

1. Training-related programs (including skills training and programs to make the training systems work better) are the omitted category. 2. Programs with evaluations reporting only gross outcomes are the omitted category. 3. Developed countries are the omitted category. The rigidity of employment index is a continuous variable. 4. On decade of implementation, programs implemented during the nineties and 2000s are the omitted category. On current status of the interventions, ongoing programs are the omitted category. 5. Omitted categories on targeting reflect none specific orientation toward disadvantage people within those groups. 6. The z-statistics test the null hypothesis of a zero coefficient following a standard normal distribution.

The values of the z-statistics are reported in the third column: * significant at 5%; ** significant at 1%.

**Figure A.1: R-squares of 95 models featuring Specification 1
(Repeatedly dropping observations with replacement)**

