

# The Role of Training Programs for Youth Employment in Nepal

Impact Evaluation Report on the Employment Fund

*Shubha Chakravarty*

*Mattias Lundberg*

*Plamen Nikolov*

*Juliane Zenker*



**WORLD BANK GROUP**

Social Protection and Labor Global Practice Group

&

Gender Cross-Cutting Solutions Area

April 2016

## Abstract

The youth unemployment rate is exceptionally high in developing countries. Because the quality of education is arguably one of the most important determinants of youth's labor force participation, governments worldwide have responded by creating job training and placement services programs. Despite the rapid expansion of skill-enhancement employment programs across the world and the long history of training program evaluations, debates about the causal impact of training-based labor market policies on employment outcomes still persist. Using a quasi-experimental approach, this report presents the short-run effects of skills training and employment placement services in Nepal. Launched in 2009, the intervention provided skills training and employment placement services for more than 40,000 Nepalese youth over a three-year period, including

a specialized adolescent girls' initiative that reached 4,410 women ages 16 to 24. The report finds that after three years of the program, the Employment Fund intervention positively improved employment outcomes. Participation in the Employment Fund training program generated an increase in non-farm employment of 15 to 16 percentage points for an overall gain of about 50 percent. The program also generated an average monthly earnings gain of about 72 percent. The report finds significantly larger employment impacts for women than for men, but younger women ages 16 to 24 experienced the same improvements as older females. These employment estimates are comparable, although somewhat higher, than other recent experimental interventions in developing countries.

---

This paper is a product of the Social Protection and Labor Global Practice Group; and the Gender Cross-Cutting Solutions Area. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at [schakravarty@worldbank.org](mailto:schakravarty@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# The Role of Training Programs for Youth Employment in Nepal: Impact Evaluation Report on the Employment Fund<sup>1</sup>

Shubha Chakravarty<sup>2</sup>, Mattias Lundberg, Plamen Nikolov and Juliane Zenker

JEL Codes: I25 I26 J16 J24 J64 O15 J46

---

## <sup>1</sup> Acknowledgments

Many individuals and organizations provided support for this ongoing research. Chiefly, the Employment Fund and its staff have championed this study for several years, offering their time, resources, and data to the research team, the survey firm, and administrators of other programs under the Bank's Adolescent Girls Initiative. Mr. Siroco Messerli and Bal Ram Paudel, (Team Leaders), in particular, facilitated all aspects of the study, including liaising with the T&E providers. We would like to thank the T&E providers for allowing unrestricted access and support to the survey team.

The Employment Fund is implemented by Helvetas Swiss Intercooperation and is supported through the generous financial support of the Swiss Development Corporation (SDC) and the UK's Department for International Development (DFID). This research study has benefited from the financial support of the World Bank's Multi-donor Trust Fund for Adolescent Girls, the Umbrella Fund for Gender Equality, and DFID. The field surveys were led by New Era Limited, which provided exceptional survey design and implementation, data management and analysis under the leadership of Mr. Madhup Dhungana.

On behalf of the World Bank, the project implementation was led by Jasmine Rajbhandary, Venkatesh Sundararaman, Bhuvan Bhatnagar, and Maitreyi Das. Amita Kulkarni, Uttam Sharma and Jaya Krishna Upadhaya coordinated survey activities in the field in 2010 and 2011, 2012 respectively. Ali Ahmed, Marine Gassier, and Jennifer Heintz of the World Bank provided excellent research assistance. The views presented in this paper are the authors' and do not represent those of the World Bank or its member countries. All errors remain our own.

<sup>2</sup> Correspondence should be sent to: [schakravarty@worldbank.org](mailto:schakravarty@worldbank.org)

## Table of Contents

I.	Introduction.....	1
II.	Skills Training for Young Women.....	3
	a. Programs worldwide .....	3
	b. The Employment Fund and the Adolescent Girls Employment Initiative (AGEI) in Nepal .....	4
III.	Impact Evaluation Design.....	6
IV.	Methodology and Data.....	8
	a. Sample Description and Sampling Technique.....	8
	b. Estimating the EF Training Program Effects.....	9
	c. Intent-to-Treat (ITT) and Average Treatment Effect on the Treated (ATT) .....	11
	d. Heterogeneous Effects .....	12
V.	Internal Validity.....	13
	a. Time-varying trends and differences at baseline .....	13
	b. Survey response rates and attrition .....	14
	c. Uptake of EF-sponsored training courses .....	14
VI.	Short-Run Results of the EF Program.....	15
	a. First-stage Estimation of Propensity Scores .....	15
	b. Impacts on Employment and Earnings for the Full Sample .....	16
	c. Non-Employment Impacts for the Full Sample .....	17
	d. Trade-wise Program Impacts on the Full Sample .....	18
	e. Gender- and Age- disaggregated Impacts .....	19
VII.	Summary and Implications .....	21
VIII.	References.....	23
	<b>Annex 1. Figures and Tables</b> .....	26
	Figures .....	26
	Tables.....	27
	<b>Annex 2. Example Ranking Form</b> .....	43
	<b>Annex 3. Analysis of Average Treatment Effects on the Treated (ATT) Impacts</b> .....	50

## I. Introduction

Youth unemployment and underemployment across much of the developing world is extremely high. Two facts highlight the importance of youth's labor in the world economy: 17 percent of the world's population are youth (ages 16-24) and youth make up 40 percent of the world's unemployed. In developing countries, where labor frequently encompasses informal self-employment and small-scale agriculture, youth also struggle with high underemployment (they are not able to work as much as they would like to) and low productivity. Up to two-thirds (60%) of the young population is underutilized in some developing economies, meaning that they are unemployed, in irregular employment – most likely in the informal sector, or neither in the labor force nor in education or training (ILO, 2013). Youth unemployment and underemployment not only slow down economic growth but also negatively impact crime rates (Fella and Gallipoli, 2007), depression rates (Frese and Mohr, 1987), substance abuse rates (Linn, Sandifer and Stein, 1985), and rates of social exclusion (Goldsmith, Veum and Darity, 1997).

Nepal is similarly affected by youth unemployment and underemployment. With a per capita income of US\$700, Nepal is South Asia's second poorest country (ahead of only Afghanistan). Helped by remittances, the proportion of the population falling below the national poverty line has declined in recent years from 31 percent in 2003-04 to 25 percent in 2011 (Central Bureau of Statistics 2011). One underlying factor of poverty in Nepal is the lack of employment opportunities, and a reliance on self-employment in the agriculture sector, which accounts for 61 percent of the total labor force (Nepal Living Standard Survey 2011). The unemployment rate for those aged 15-29 is 19.2 percent, compared to just 2.7 percent for people older than 15 (ILO 2014). Faced with these prospects, young Nepalese are compelled to consider migrating overseas in search of better opportunities.

Governments worldwide have responded by creating job training and placement services programs. In 2013, European Union (EU) launched an eight billion euro initiative aiming to provide every young European with a job, apprenticeship, or training within four months of becoming unemployed.<sup>3</sup> In Latin America, job training programs (referred to collectively as the “*Jovenes*” programs) have been implemented since the early 2000s. To date, more than 700 youth employment programs in around 100 countries have been implemented and more than 80 percent of these programs offer some sort of skills training.<sup>4</sup>

Despite the rapid expansion of skill-enhancement employment programs across the world and the long history of training programs' evaluations,<sup>5</sup> debates about the causal impact of training based labor market policies on employment outcomes still persist. Based on US and European evidence, Card et al. (2009) review impacts of various training programs. Their analysis suggests that classroom and on-the-job training programs are not particularly effective in the short run, but have larger positive impacts after two years. They also find that youth programs, on average, tend to yield less positive impacts than untargeted programs. The evidence regarding impacts by gender is mixed. Kluve (2006) reviewed a number of

---

<sup>3</sup>See <http://www.economist.com/news/leaders/21582006-german-led-plans-tackling-youth-unemployment-europe-are-far-too-timid-guaranteed-fail> and <http://www.dw.de/the-lost-generation-youth-unemployment-in-the-eu/a-16925070>

<sup>4</sup> See <http://www.youth-employment-inventory.org/>

<sup>5</sup> See Heckman (1999) and many others, including randomized controlled trials, like the Job Training Partnership Act in the US.

employment program evaluations in Europe and found that the programs tended to have larger impacts for women than men. However, Card (2009) found that programs tended to work equally well for men and women.

In general, program evaluations from developing countries show larger impacts than programs conducted in other regions. Based on 289 youth employment interventions in 84 countries, Betcherman et al. (2007) show higher impact in developing countries than in developed ones.<sup>6</sup> Most of the rigorous evidence on training programs in developing countries is from Latin America, where positive impacts are particularly pronounced (Gonzalez-Velosa et al., 2012; Attanasio et al., 2008).<sup>7</sup> Attanasio et al. (2008) evaluate the *Jovenes en Acción* job training program in Colombia. *Jovenes en Acción* provided three months of classroom training followed by a three month unpaid internship at a company. Attanasio et al. (2008) detect positive employment effects for women (4 to 7 percentage points), no employment effects for men and positive earnings effects for both men (8 percent) and women (18 percent). The study argues that the increase in earnings is due to increased employment in formal sector jobs upon training completion.<sup>8</sup>

In Nepal, a wide variety of public and private technical education and vocational training (TEVT) programs are available to youth. Youth unemployment is frequently cited as a driver of the ten-year conflict from which Nepal emerged in 2006 (Macours 2011, ILO 2014), and since then international donors have invested heavily in the TEVT sector. Additionally, a combination of age, low educational attainment, and norms around marriage and childbirth confers multiple additional disadvantages to young women in the labor market. Their labor force participation rate is 43.0 percent compared to 51.7 percent for young men (ILO 2014). Concerns about the social exclusion of women, ethnic minorities, indigenous peoples, and other historically disadvantaged groups have also spurred advocacy and investment in training opportunities for these groups. The sector remains fragmented, however, with many public, private, and non-governmental training providers of varying quality and weak links to the job market. Among development partners, the World Bank and the Asian Development Bank are particularly active in this area, with the ADB's US\$25 million Skills for Employment project and the World Bank's US\$30 million Enhanced Vocational Education and Training Project. To the best of our knowledge, no rigorous evidence yet exists on the impacts of these programs on the labor market outcomes of participants.

Using a quasi-experimental approach, this report presents the short-term effects of skills training and employment placement services in Nepal. Founded in 2008, the Employment Fund (EF) is operated by Helvetas, a Swiss NGO, in partnership with the Government of Nepal. It is currently one of the largest youth training initiatives in the country, serving almost 15,000 youth annually. In partnership with the

---

<sup>6</sup> Betcherman et al. (2007) also found that only one-quarter of the 289 studies reviewed included an experimental or quasi-experimental design. Among this particular subset of studies, 40 percent showed zero or negative impact on employment and earnings outcomes. Betcherman et al. (2007) concludes that observational studies were almost 50 percent more likely to exhibit positive impacts than programs with experimental or quasi-experimental designs.

<sup>7</sup> Only Card et al. (2007) detects no statistically significant effect. Card et al. (2007) evaluate a job training program in the Dominican Republic which included classroom training and a private sector internship. The impact evaluation shows no effect on employment and a marginally significant earnings effect of about 10%.

<sup>8</sup> Two previous evaluations of *Jovenes*-type programs in Argentina and Peru similarly found larger impacts for women than for men, with employment impacts for women of 9 to 12 percentage points in Argentina (Aedo et. al. 2004) and 6 to 15 percentage points in Peru (Ñopo et. al. 2007).

Employment Fund's donors,<sup>9</sup> the Adolescent Girls Employment Initiative (AGEI) was launched in 2009 to expand the program's reach to an additional 4,410 Nepali women aged 16-24 over a three-year period.

Unique characteristics of this impact evaluation include:

- A large study sample (4,677 individuals in the pooled 2010-2012 cohorts examined in this report)
- Particular focus on the outcomes of women aged 16-24 as well as the outcomes of a broader category of youth (men and women aged 16-35)
- Socioeconomic survey data of participants and a control group of non-participants as well as administrative follow-up data on program participants
- Exhaustive tracking of program participants over time that produced high response rates.
- Examination of a large set of outcome domains, including employment and earnings, empowerment and self-confidence, risky behaviors, and impacts on the household.

We find, approximately two years into the program, the EF program positively improved employment outcomes. EF training program participation generated an increase in non-farm employment of 15 to 16 percentage points for an overall gain of 46 percent. The program also generated an average monthly earnings gain of 921 NRs monthly ( $\approx$ 12 USD). Given that the average monthly income at baseline was 1272 NRs ( $\approx$  17 USD), the EF program impact generated an economically meaningful income gain of about 72 percent for the combined 2010-2012 cohorts.). We find that the program's impacts on employment were larger for women than for men, though the impacts on other economic outcomes were similar for both sexes. We also detect no difference in impacts when comparing older (24-35) and younger women (age 16-24), indicating that the program is equally effective for the younger women targeted under the AGEI.

This report's employment estimates are comparable, though somewhat higher, than other recent experimental interventions in developing countries. Estimates of employment impacts in job training programs in Latin America by Attanasio et al. (2008) and Alzua et al. (2013) show respectively gains of 14 percentage points and 10 percentage points. In terms of earnings, this evaluation finds substantially higher impacts than the increase in earnings of 18 percent found by Attanasio et al. (2008).

The report is organized as follows: In Section II, we describe the global Adolescent Girls Initiative and the Employment Fund program in Nepal. Section III describes the design of the AGEI impact evaluation, Section IV describes our empirical approach, and Section V discusses the potential threats to the study's validity. Section VI presents results and Section VII concludes.

## **II. Skills Training for Young Women**

### **a. Programs Worldwide**

This impact evaluation was launched under the aegis of the World Bank's Adolescent Girls Initiative (AGI) launched in 2009 with the aim of facilitating the transition to productive employment for young females. Comprised of eight pilot projects – in Afghanistan, Haiti, Jordan, Lao People's Democratic Republic, Liberia, Nepal, Rwanda, and South Sudan– the initiative's launch was motivated by two factors: (1) the

---

<sup>9</sup> The Employment Fund is financed by the United Kingdom's Department for International Development (DFID), the Swiss Agency for Development and Cooperation (SDC) and the World Bank.

particular challenges faced by young women in developing countries during the school-to-work transition, including high burden of domestic work, competing pressures to bear and raise children, negative social norms regarding occupational choice and mobility, and gender discrimination, and (2) the potential benefits -- to household members, and to current and future children -- of empowering young women.

AGI interventions offered skills training and various ancillary services tailored to local context, such as childcare, mentoring, job placement assistance, and links to microcredit, in order to facilitate young women's transition to productive employment. Five of the eight pilots included an experimental component and two of the impact evaluations have reported results. Using a randomized evaluation, the AGI project in Jordan provided short-term soft-skills training and six-month employment vouchers to recent female graduates of technical universities. Although the vouchers assisted young women to obtain short-term internships, the evaluation detected no effect on the recipients' employment or earnings after the end of the voucher period (Groh, 2010). In contrast, the AGI pilot in Liberia found strong positive impacts on employment and earnings. Using a randomized "pipeline" design, individuals were randomly assigned to receive training in two sequential batches. In addition to positive impacts on participants' self-confidence, savings, and household food security, the program had sizable impacts on employment (47 percent) and weekly earnings (80 percent) for women aged 16-27 (Adoho 2014).

Two other skills training programs for young women have also been rigorously evaluated. In India, a training program in stitching and tailoring offered to young women in poor slum communities of New Delhi had a unique feature, introduced to increase commitment and encourage regular attendance (Maitra & Mani, 2014). Maitra & Mani (2014) found that the program increased the likelihood of casual or permanent wage employment by more than 5 percentage points, self-employment by almost 4 percentage points, and any employment by 6 percentage points. The program increased hours worked in the post-training period by around 2.5 hours. Finally, a related project implemented by the international NGO BRAC in Uganda implemented village-level girls' clubs to provide life skills, reproductive health, and livelihood skills to young women aged 14 to 20. A randomized impact evaluation found substantial increases in employment (72 percent) and improved empowerment and reproductive health outcomes (Bandiera 2012).

## **b. The Employment Fund and the Adolescent Girls Employment Initiative (AGEI) in Nepal**

Started in 2008, the Employment Fund (EF), now one of the largest skills training programs in the country, provides vocational training and placement services under a unique governance structure. Since 2011, the Employment Fund has operated under an agreement between the Government of Nepal and the Swiss Agency for Development and Cooperation (SDC). A Steering Committee, chaired by the Joint Secretary of the Ministry of Education, governs the Employment Fund. In addition to providing training and placement services, the Employment Fund engages in support activities, capacity-building, and cooperative activities with the government's Council for Technical Education and Vocational Training (CTEVT).

Each year, the Employment Fund authorizes training programs under a competitive bidding system with various training providers. First, the Employment Fund issues a call for proposals to Training and Employment (T&E) providers seeking to provide skills training and employment services. The range of T&E provider types is enormous: from formal technical education and vocational training (TEVT)



institutions, public and private providers, to skilled artisans offering apprenticeships. The second step, after the call for proposals, is for each T&E provider to complete a Rapid Market Assessment (RMA) outlining viable and potential employment opportunities.<sup>10</sup> The third step in the process is for the EF to evaluate submitted proposals according to preset criteria; the EF weighs the capacity and experience of each T&E provider, the market demand for the proposed trades being offered, and the proposed costs. Finally, the EF issues a contract to selected providers: the contract specifies the number of training courses (hereafter called “events”) to be conducted and the number of individuals to be trained and employed in each event. The T&E providers are then free to recruit and select their own trainees for each of their training events, according to guidelines established by EF.

Table 1 provides the total number of T&E providers, number of training events, and number of trainees. Table 1 shows an increase of total number of program beneficiaries between 2010 and 2012. Training courses in technical skills vary across a wide range of trades (*e.g.*, incense stick rolling, carpentry, tailoring, welding and masonry). All females receive 40 hours of life skills training (beginning in 2011) and a subset of EF trainees receive a short course in basic business skills. In addition, each trainee is encouraged to complete a skills certification test offered by the National Skills Testing Board (NSTB).<sup>11</sup>

Upon completion of the classroom-based training, the EF places emphasis on job placement services. EF verifies trainees’ employment status three months and six months after the completion of the training.<sup>12</sup> Upon verification, T&E providers receive an outcome-based payment from the EF that is higher for trainees who are employed. The outcome-based payment system creates strong incentives for the T&E providers to provide placement assistance and provides graduates with an opportunity to put their new skills to work immediately after the training. The EF emphasizes the placement of trainees into “gainful” employment in which they earn a minimum of 3,000 NRs ( $\approx$ 40 USD) per month.<sup>13</sup>

In 2010, the EF partnered with DFID and the World Bank’s Adolescent Girls Initiative to improve the EF’s reach and impact for young women aged 16 to 24. Training under this Adolescent Girls Employment Initiative (AGEI) proceeded in the same way as it did for other EF trainees, except that certain events had been flagged in advance as likely to attract female trainees. This was done in order to ensure that the EF reached adequate numbers of young women.<sup>14</sup> For the purpose of this evaluation, we designate all female participants aged 16 to 24 in EF-sponsored training courses as “AGEI”.

The Employment Fund struggled in 2010 to recruit young women to training events and in 2011 launched an enhanced communication and outreach strategy to recruit more female trainees.<sup>15</sup> In addition to the T&E

---

<sup>10</sup> The RMA process is facilitated via EF-supported capacity building workshops.

<sup>11</sup> In some cases, the Employment Fund has even cooperated with the NSTB to develop skills standards for tests in trades where no test existed.

<sup>12</sup> The employment status of a sample of graduates is verified by EF field monitors three to six months after the completion of the training event.

<sup>13</sup> The definition of “gainful” employment was increased in 2012 to 4,600 NRs ( $\approx$ 60 USD). Throughout this report, we use the prevailing exchange rate during 2010 and 2011 of 75 NRS to 1 USD.

<sup>14</sup> The T&E providers running these training events are not given hard quotas for how many AGEI trainees to reach, but any female trainees aged 16 to 24 that they do train are designated as “AGEI” trainees for accounting and monitoring purposes. Employment Fund records indicate that 4410 young women were reached over the three year initiative, as shown in Table 1, exceeding the target of 4375.

<sup>15</sup> They reached only a bit more than half of their annual target of 1500.

advertisement, the EF sponsored radio and newspaper ads specifically geared towards young women. Many of these ads specifically encouraged women to sign up for non-traditional trades for women, such as mobile phone repair, electronics, or construction.<sup>16</sup> The Employment Fund also partnered with women's and community-based organizations to attract applications from women and marginalized groups – if a referred applicant gained entry to an EF-sponsored training event, the partner organization was paid a small finder's fee equivalent to about 1.25 US dollar per person.

The Employment Fund uses a differential pricing mechanism that awards a higher incentive to service providers who agree to train (and place) more disadvantaged groups, according to established vulnerability criteria.<sup>17</sup> The highest incentive is awarded for training and placing the most disadvantaged (highly vulnerable women including AGEI trainees, widows, ex-combatants, disabled women, etc.), and incentives are gradually lowered for less prioritized groups. Training providers that are able to cater to these higher priority target groups are therefore eligible to receive a higher outcome price, but they also face a higher risk of failing to achieve the outcome (gainful employment). The combination of a results-based system with a progressive incentive scheme ensures that training providers with the capacity to work with vulnerable groups will likely opt to do so.

### III. Impact Evaluation Design

This evaluation estimates the impact of the EF training program by comparing the outcomes of participants, who comprise the “treatment” group, to a control group of individuals who applied, but were not selected for, an EF-sponsored training course. Isolating the causal effects of the EF training program on employment and other outcomes is complicated by the fact that some T&E providers have at least some degree of choice over who they choose to train or training participants could seek EF training for reasons we cannot fully measure.<sup>18</sup> In other words, comparing post-program participation outcomes of EF program participants versus non-participants may confound training program influences on outcomes with those of hard-to-measure individual-level, family-level or T&E-level attributes that affect both training program participation and the outcome of interest.

This estimation problem would be greatly reduced if we could compare the outcomes of individuals from truly comparable backgrounds except for their EF training program status. The primary concern in being able to detect the causal effects of the training program is that the two sets of individuals (*i.e.*, ones receiving training and ones who do not) may have had different characteristics to begin with and it may be those characteristics rather than the EF training program that explain the difference in outcomes between the two groups.<sup>19</sup> The *unobserved* differences in characteristics are particularly concerning.

---

<sup>16</sup>Anecdotal evidence also suggests that the introduction of life skills training, covering topics such as communication, leadership, and reproductive health, was particularly popular among young women and may have contributed to the growing numbers of AGEI trainees.

<sup>17</sup> For more details on the differential pricing scheme for vulnerable groups and the outcome-based payment system, see [http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2014/07/18/000470435\\_20140718142756/Rendered/PDF/894760BRI0141100Box385283B00PUBLIC0.pdf](http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2014/07/18/000470435_20140718142756/Rendered/PDF/894760BRI0141100Box385283B00PUBLIC0.pdf)

<sup>18</sup> Systematic differences in unobservable characteristics that cannot be measured quantitatively, such as motivation, confidence, and natural ability, will lead to biased estimates of program impact.

<sup>19</sup> We perform a number of statistical tests (presented in section V) to examine the similarities and differences between the two groups in our evaluation sample.

T&E providers can select which beneficiaries enter their training programs, and this selection process drives the difference between trainees and non-trainees. Each T&E provider advertises, collects applications, shortlists and interviews applicants, and selects participants for each of training events following Trainee Selection guidelines established by EF. These guidelines stipulate a two week minimum period of advertisement, the eligibility criteria for training events, and the procedures for shortlisting and interviewing. There are three eligibility criteria for all EF-sponsored training programs: age (from 16 to 35), education (below SLC,<sup>20</sup> or less than 10 years of formal education), and self-reported economic status.<sup>21</sup> Only applicants who meet all three criteria are eligible to be shortlisted. T&E providers are advised to shortlist at least 50 percent more candidates than the number of spaces in the training event. The guidelines also outline a uniform process for interviewing shortlisted candidates, including a detailed scoring rubric, instructions for ranking the shortlisted candidates by score, and selecting the top-scoring candidates for participation.<sup>22</sup> Figure 1 in Annex 2 shows a sample ranking form used by T&E providers. This scoring and ranking procedure forms the basis of our sampling strategy for this evaluation and is described in detail in the next section.

We use quasi-experimental methods to solve the above evaluation concerns. Our general evaluation strategy is to observe an individual before and after an EF training program and to compute a simple difference in outcome for that individual over time. The average difference over time of the individuals enrolled in the training program (*i.e.*, the treatment group) is then compared to the average difference over time of the individuals who are not enrolled (*i.e.*, the comparison group). Conceptually, this so-called difference-in-differences method (Campbell, 1969; Meyer, 1995), cancels out the effect of all of the characteristics that are unique to a specific individual and that do not change over time.<sup>23</sup> Therefore the difference in outcomes across the two groups can (under certain conditions, discussed in Section Va) be attributed solely to the treatment status, *i.e.* whether the individuals received training or not. To further purge bias in our estimates arising from observable differences between trainees and non-trainees, we employ a combination of this difference-in-difference estimation with propensity score matching and propensity score weighting techniques. We describe the specific estimation and matching methods in the next section.

---

<sup>20</sup> The School Leaving Certificate (SLC) is obtained after successfully passing examinations after the 10<sup>th</sup> grade. To be eligible, EF applicants must have not taken, or not passed, their SLC exams. This criterion has been loosened for some trades starting in 2012. In practice, because the educational status reported on the application is not verified, this criterion was not perfectly adhered to.

<sup>21</sup> An applicant is considered “economically poor” if they report a non-farm per capita household income of less than 3000 Nepali rupees (NRs) per month or, in the case of farming families, less than 6 months of food sufficiency. Since these self-reports are not verified, and applicants know in advance that they must be “poor” in order to be eligible for the program, it is unclear how well this criterion is adhered to.

<sup>22</sup> In addition, and as mentioned earlier, the guidelines also set out a progressive payment structure to incentivize T&E providers to select trainees from particular socially disadvantaged groups. The guidelines also allow for T&E providers to select up to 2 “alternates” per event, in case a selected trainee declines to join the program.

<sup>23</sup> The difference-in-differences method helps resolve the problem of selection to the extent that many characteristics of units or individuals can reasonably be assumed to be constant over time (or *time-invariant*). Think, for example, of *observed* characteristics, such as a person’s year of birth, a region’s location close to the ocean, a town’s level of economic development, or a father’s level of education. Most of these types of variables, although plausibly related to outcomes, will probably not change over the course of an evaluation. Using the same reasoning, we might conclude that many *unobserved* characteristics of individuals are also more or less constant over time.

## IV. Methodology and Data

### a. Sample Description and Sampling Technique

Our primary source of data comes from a survey covering three consecutive cohorts of EF trainees (from 2010 to 2012), with two rounds of data collection for each cohort.<sup>24</sup> Figure 1 depicts the impact evaluation timeline.

We sample at the training event and the applicant level. The main sampling frame for data used in this study consisted of all EF training courses sponsored in a given year. The number of training events comprising the sample frame ranges from 598 (in 2010) to 711 (in 2012). Table 1 reports the number of events and participants by year.

Sampling into this study included a combination of stratified, random and convenience sampling. First, we selected a subset of training events occurring between the months of January through April.<sup>25,26,27</sup> Second, from the universe of training events offered during these four months, we randomly selected up to 15 districts. Third, from that list of training events occurring in these districts, we randomly selected 20 percent of the training events. Finally, a survey team visited each sampled training event on the day when applicant selection took place. Each event's ranking sheet listed the shortlisted applicants from the top-scorer to the bottom and indicated the threshold, or minimum score needed to gain admission to the course. From this ranking sheet, the survey team selected applicants whose scores were within 20 percent of the threshold for admission to training events. The sampled applicants above the threshold comprise this study's treatment group, while those below the threshold make up the control group. Immediately following the sampling of applicants, a baseline survey was conducted on the treatment and control groups, before the results of the selection process were announced.

Table 2 shows the resultant sample of events for the three cohorts. The 2010 event sample comprised 64 events across 30 districts. The 2011 sample comprised 182 events, of which 113 events were dropped from the baseline survey, either because the survey team could not reach the event on the day of applicant selection or because the event was not "oversubscribed".<sup>28</sup> The remaining 69 events in 34 districts were included in the 2011 baseline sample. The sampling process was much improved in 2012, with 85 out of 112 sampled events successfully included in the final evaluation sample.

This non-random (in 2010) and partially-random (in 2011 and 2012) sampling of training events and applicants introduces bias in two potentially important ways. First, the training events in our study may not be representative of all EF-sponsored trainings. In all three cohorts, training events that enter our sample

---

<sup>24</sup> For the 2010 cohort, a second follow-up was conducted on half of the cohort. Future analysis will examine the longer-term outcomes for this group.

<sup>25</sup> Because of the AGEI focus of this study, we prioritize AGEI training events (identified by the T&E as we described earlier). Because the selection into the study population was based on an individuals' proximity to the threshold score, it was not possible to stratify on AGEI status. However, events that were likely to include more AGEI candidates were purposely oversampled in 2011 and 2012 so as to increase the number of AGEI candidates in the study population.

<sup>26</sup> Eighty percent of EF training events occurred during these four months.

<sup>27</sup> In 2010, because a complete event listing was not available in advance, the events were chosen by convenience, based on scheduling and accessibility.

<sup>28</sup> The survey team was instructed to drop the event from the sample if there were not at least 3 rejected candidates that fell within 20% of the threshold score. In other words, if there were not at least 3 people who could be sampled for the control group, the event was dropped from the sample.

are more likely to be based in district centers, are more likely to be more sought after (and hence oversubscribed), and are more likely to be run by high-capacity T&E providers.<sup>29,30</sup> Second, selected trainees could differ in characteristics, other than being offered training, from non-trainees, since T&E providers purposely interview and select the applicants they think will perform best. We mitigate this bias by selecting our study participants within a narrow range of each event's threshold score, which we hypothesize will limit the differences between the treatment and control groups. We examine this hypothesis further in section V.

The sampling procedures described above resulted in a study population of 4677 over all three cohorts. For the pooled sample (*i.e.*, 2010 through 2012 cohorts), the study population is about 64 percent female and on average 24.5 years old. Fifty-eight percent are married while 51 percent have at least one child. Approximately 59 percent of the sample has engaged in any income-generating activity in the month prior to the survey, a figure which may seem high, but includes those who are working without pay on their own household farms. When we restrict to non-farm income-generating activities, the employment rate falls to 27 percent. At baseline, the average earnings of the pooled sample were 1272 NRs per month (equivalent to about 17 USD). This figure may seem low, since it represents the average earnings over the entire study population of 4677 individuals, including those with zero earnings. Only 17 percent of the 2010-2012 pooled sample earned more than 3000 NRs per month, a level deemed to represent “gainful” employment. Interestingly, about 15 percent of the sample was already engaged in the same trade for which training they applied (denoted as “trade-specific IGA”), indicating that a significant minority of applicants were looking to upgrade their existing skills. Though they are not older than the men in the sample, women are more likely to be married and have a child, and have lower employment and earnings at baseline.<sup>31</sup>

As discussed in Section II, the EF provides financial incentives for T&E providers to recruit and train people from the Dalit and Janajati ethnic groups. For Janajatis, the T&E providers appear to respond to these incentives, as 44 percent of the applicants are Janajatis, and they are statistically more likely to be in the treatment group than control.<sup>32</sup> The T&E providers appear to have less success with attracting Dalit applicants: only 8 percent of the applicants are Dalits (a bit less than the population average), and they are equally divided between the treatment and control groups.

## **b. Estimating the EF Training Program Effects**

To estimate causal effects of the EF training program on various outcomes, first we employ a “difference in difference” (DID) technique. We will refer to this as the OLS specification in the remainder of this report. The main equation we estimate is:

---

<sup>29</sup> The survey team deployed a fixed number of staff to each district center, with a schedule of all events in that district. Events conducted by high-capacity T&E providers and in popular trades were more likely to keep to their scheduled start date. Because the T&E providers were not required to wait more than 5 days between the interviews and the start of training, and the unpredictability of the interview dates, it was impossible for the survey team to reach all of the events with enough time to conduct the baseline survey.

<sup>30</sup> If these characteristics also determine the quality of the training, this non-random sampling of events may bias our estimates upward and overstate the true impact of the program. One mitigating factor is that in every cohort our sample includes a high fraction of the T&E providers contracted by EF.

<sup>31</sup> More details about the characteristics of the study population are found in a separate baseline report, available upon request from the authors.

<sup>32</sup> For reference, the 2001 census indicated that Janajatis made up 37% and Dalits made up 13% of the total Nepali population.

$$Y_{ijt} = \alpha + \delta (Post_t \times Treat_i) + \beta_1 Post_t + \gamma_i + \varepsilon_{ijt} \quad (1)$$

This regression relates a given outcome to EF program training status.  $Y_{ijt}$  is the outcome of interest for individual  $i$  from training event  $j$  at time  $t$ .  $Treat_i$  is an indicator which is equal to 1 for the treatment group and 0 for control.  $Post_t$  is an indicator equal to 1 for follow-up observations and 0 for baseline. Its coefficient captures aggregate factors that would cause changes in  $Y_{ijt}$  even in the absence of a training program. The term  $\gamma_i$  represents an individual fixed effect. This individual fixed effect is critical to our identification strategy, as it controls for differences in time-invariant observable and unobservable characteristics at baseline, as described in section III. The final term,  $\varepsilon_{ijt}$ , is an idiosyncratic error term that is clustered by training event, in order to account for the likely correlation of outcomes among applicants to the same training course.

The coefficient of interest,  $\delta$ , defines the impact of the program comparing the change in  $Y_{ij}$  of treatment group individuals with control group individuals over time. If outcomes of individuals assigned to EF training are similar to individuals not trained (that is, if the training has no impact), then we should find  $\delta = 0$ . If individuals trained by EF have better labor market outcomes than non-participants, we should find  $\delta > 0$ .

To further purge remaining differences between observable characteristics among trainees and non-trainees that could influence the difference in impacts between the two groups, we augment the “difference-in-difference” technique with propensity score matching and weighting approaches. Both methods rely on first estimating each individual’s likelihood of being offered training (*i.e.*, *propensity score*), based on individual baseline characteristics, such as age, education, and family background.

We implement the propensity score method by first employing the following probit model:

$$\pi_{ij} = \beta X_i + Z_j + \mu_{ij} \quad (2)$$

In this equation,  $\pi_{ij}$  is equal to 0 (for non-trained individuals) or 1 (for trained individuals) for individual  $i$  in event  $j$ , and  $X_i$  is a vector of individual and household level explanatory variables, all measured at baseline. The error term, clustered by event, is given by  $\mu_{ij}$ . To predict likelihood of being trained, we use age, sex, education, ethnicity, employment status, marital and parental status, analytical ability (as measured by the commonly used Raven’s progressive matrices and one financial literacy question), and an entrepreneurial orientation score based on a set of 11 questions. At the household level, we include household size, education level of the household head, and the quintile of the household’s wealth based on an index of ten household assets. At the district level, the model controls for the district in which each event is held, the T&E provider, and the trade of the training (*e.g.*, hospitality), all represented by the vector  $Z_j$ . The predicted value of  $\pi_{ij}$  is the estimated propensity score, or likelihood of being in the treatment group.

After estimating this propensity score, we derive the estimated treatment effect using two methods. The first method is “inverse propensity score weighting”, in which individuals are weighted according to the

inverse of their estimated propensity to participate in the program. The weighted observations are then used in a DID regression, as given by equation (1). We will refer to this as the IPSW specification from here on. We implement IPSW following Hirano et al (2003). This particular weighting method, as opposed to matching approaches, has the nice property of including all the data (unless weights are set to 0) and does not depend on random sampling, thus providing for replicability. We use a weighted least squares regression model, with weights of  $1/\hat{\pi}$  for the treatment group and  $1/(1-\hat{\pi})$  for the control group, where  $\hat{\pi}$  is the estimated propensity score from (2). Standard errors are clustered by training event.

The second method is “propensity score matching”, in which individuals in the treatment group are matched to individuals in the control group who have similar propensity scores. We use a nearest-neighbor matching algorithm, in which each individual in the treatment group is compared to a fixed number of control observations (in our case, four) with the closest propensity scores. We will refer to this as the NN specification in the remainder of this report. Following Smith and Todd (2005), we estimate the difference-in-difference matching estimator for the training program effect  $\delta$  as follows:

$$\widehat{\delta}_M = \frac{1}{N_T} \sum_{i \in T} [(y_{it_1} - y_{it_0}) - \sum_{j \in C} W_{ij} (y_{jt_1} - y_{jt_0})] \quad (4)$$

$N_T$  is the number of treatment observations, the subscript  $t_1$  denotes follow-up observations and  $t_0$  denotes baseline observations;  $W_i$  is a matrix of weights. Weights for nearest-neighbor matching are computed by:

$$W_{ij}(y_{jt_1} - y_{jt_0}) = \frac{1}{x} \sum_{j \in A_x} (y_{jt_1} - y_{jt_0}) \quad (5)$$

$A_x$  is a set of  $x$  observations with the lowest values of  $|\hat{\pi}_i - \hat{\pi}_j|$ . As in the two previous models outlined in this section, the dependent variable is the first difference of a given outcome between the baseline observation and the follow-up observation. The statistical software package we use for this specification does not allow for clustering of standard errors.<sup>33</sup>

### c. Intent-to-Treat (ITT) and Average Treatment Effect on the Treated (ATT)

Estimates of the EF program’s effects correspond to two different questions. The first is the effect of the intervention on the average outcomes of those assigned to one of the EF training events, regardless of whether they used the training services. In the experimental literature, this is known as the *intention to treat* (ITT) effect. Angeluci and Orazio (2006) discuss the quasi-experimental counterpart, which is the method we apply in this study.<sup>34</sup>

If one is interested in the effectiveness of a program among the entire class of youth who are eligible for it, then the ITT estimates are the appropriate results to examine. Note that this ITT estimate is not biased by the fact that only some individuals choose to participate the EF program because we derive the ITT by comparing the average outcomes of everyone assigned to one of the EF training groups, whether they use

<sup>33</sup> We implement the *psmatch2* package in Stata (Leuven 2003).

<sup>34</sup> We retain the use of this term to be consistent with the same concept in the experimental design literature.

the program or not, with the average outcomes of everyone assigned to the control group (*i.e.*, the non-trainees).

If instead one is interested in the effects among individuals who have actually completed the program, then one should consult the average treatment effects on the treated (ATT) results, which are based on actual program participation, rather than program assignment.

We present both intent-to-treat (ITT) and average treatment effects on the treated (ATT) estimates of the impact of the Employment Fund’s training program. In our view, the ITT estimates are more relevant from the policy maker’s perspective.<sup>35</sup> If one were to invest in scaling up the EF’s training programs to more Nepali youth using the same eligibility criteria, the ITT estimates indicate the level of impact one could expect to achieve *on those target youth*. The ITT estimates account for the foreseeable uptake and dropout rates that one would expect if the program were expanded. The ATT estimates are relevant from a program implementer’s point of view, as they indicate how well the program worked for the participants who completed the course. They compare everyone who completed the training to everyone who did not, regardless of why they did not complete the training (*e.g.*, they were not offered a space, they were not eligible, or they dropped out or declined to join).<sup>36</sup> Because neither the direction nor the extent of bias can be determined precisely, and to address both the policy and programmatic perspectives, this report presents ITT results in Section VI. ATT estimates are available in a separate Annex upon request from the authors.

#### **d. Heterogeneous Effects**

In seeking to examine the effects of the Employment Fund’s training program, it is important to keep in mind that different people within the intervention may respond differently to the same policy intervention—a possibility that researchers refer to as “treatment heterogeneity.” We might be concerned that people with particular social or demographic characteristics may fare differently than the average program participant.

Because treatment heterogeneity has important policy implications, we estimate heterogeneous treatment effects by estimating the impact both for the full sample of participants as well as for several sub-populations. For example, we may want to know how well the program works for the average participant or how well the program works specifically for women, young women, and different ethnic groups that are especially targeted by the EF. Different effects for various demographic groups can assist us in informing the policy debate regarding whether specialized investments are paying off or whether further strategies might be needed to ensure that various groups participate and benefit equally from the program. The AGEI sub-group (*i.e.*, young women aged 16-24), in particular, merits special attention because of the Employment Fund’s outreach activities towards this particular age group.

---

<sup>35</sup> Conceptually, the ITT estimates are equal to or lower than the ATT estimates, since they include all the people to whom the program was targeted but chose not to participate, or who did not complete the program in the treatment group. ATT estimates may overestimate true program impact, since they represent the impact among a self-selected group of people who chose to take part and complete the program. It may be that those who are more motivated and capable are more likely to remain in the program, and the less capable drop out; on the other hand, it may be that the more capable decide that they don’t really need what the program provides.

<sup>36</sup> Data on training program completion were obtained from the Employment Fund’s administrative records, and cross-checked with self-reported training program participation collected at follow-up.



To test whether the impacts of the EF training program vary by certain pre-defined sub-groups, we employ a triple difference specification:

$$Y_{ijt} = \alpha + \psi(Post_t \times Treat_i \times Group_i) + \beta_1 Post_t \times Treat_i + \beta_2 Post_t \times Group_i + \beta_3 Post_t + \gamma_i + \varepsilon_{ijt} \quad (6)$$

$Group_i$  represents the subgroup characteristic. The parameter of interest  $\psi$ , indicates the differential impact of the EF training on that subgroup (relative to everyone else), while  $(\psi + \beta_1)$  indicates the overall impact on the treated individuals in that subgroup. The remaining terms are as defined in (1).

## V. Internal Validity

Although we have sought to maximize the scientific quality of this impact evaluation, it remains possible that we have missed some important program impacts. Alternative explanations, other than assignment to a training program, could account for differences in outcomes between individuals offered training and individuals who are not trained. These explanations are the so-called threats to the internal validity. In this section, we address and dismiss such potential explanations.

### a. Time-Varying Trends and Differences at Baseline

We start by addressing concerns about pre-existing differences and time-varying trends that could account for observed training effects when comparing trainees and non-trainees.<sup>37</sup> Towards that objective, we present “*balancing tests*” which capture the degree of similarity between the two types of participants. Table 4 presents baseline participant characteristics (*i.e.*, balancing tests) for a set of 38 demographic indicators.<sup>38</sup> These tests are based on “ITT” comparisons of the treatment group (*i.e.*, individuals whose scores qualify them for admission to an EF training event) and the control group. The baseline balance tests for the pooled sample (2010-2012) indicate that significant differences exist between treatment and control groups for baseline observable characteristics and pre-treatment outcome variables.<sup>39</sup> Relative to rejected candidates, treated individuals are more likely to be Janajati and are less likely to have finished SLC (10<sup>th</sup> grade), characteristics which reflect the eligibility criteria and the EF’s differential pricing scheme for vulnerable groups. Further, the likelihood of treated individuals being engaged in non-farm and trade specific employment before take up of training was higher, as well as their working hours and ability to earn more than 3000 NRs a month. These differences are consistent with T&E providers’ incentives to select candidates they think will perform best. Finally, individuals in the treatment group are also less likely to have control over savings and money of their own at baseline. To address these differences (and potential differences in unobservable characteristics) we applied a difference-in-difference approach in our analysis. However, growth in the outcome variables may not follow a common trend, particularly when starting off at very different initial levels. Although it does not resolve the parallel trend assumption, we additionally

---

<sup>37</sup> Characteristics of these two groups need not be the same for a difference-in-difference identification strategy to yield causal effects; however, similar characteristics facilitate the empirical estimation.

<sup>38</sup> Using a statistical tool called a student’s T test, we examine whether the difference in means between the two groups is statistically significant.

<sup>39</sup> The balance tests in Table 4 use standard errors that are clustered by training event.

applied propensity score weighting and matching techniques to achieve a higher degree of baseline comparability across groups.<sup>40</sup>

## **b. Survey Response Rates and Attrition**

One potential problem with any study such as ours is sample attrition. Approximately one year after each baseline survey, we conducted a follow-up survey.<sup>41,42</sup> The response rates were quite high for both follow-up surveys (see Table 5). Overall, the survey firm was able to track and successfully interview 88 percent of the baseline survey respondents, yielding a final sample for analysis of 4,101 individuals.<sup>43</sup> High response rates such as these limit the degree to which any differences between respondents and non-respondents can affect our estimates.

Particularly worrisome is the possibility of so-called “*differential attrition*”, a situation in which we re-interview trainees and non-trainees at statistically different rates. Even with low attrition rates for a panel survey, differential attrition may influence the scientific validity of results. Suppose, hypothetically, that very motivated individuals in the control group (*i.e.*, non-trainees) are more likely than trainees to migrate to a district where they find employment. As a result of this migration, these individuals will not show up in our analyses and this scenario could compromise the scientific validity of this study’s estimates.

In Table 19, we explore this possibility and show no evidence to support it. Table 19 shows a series of regressions on the correlates of survey attrition. We regress the panel status of respondents on their treatment status, depending on how the treatment indicator is defined. This specification provides information on whether treatment or control individuals are more likely to be lost to follow-up. Next, we add a series of covariates, such as gender, age, marital status, parental status, ethnicity, and employment, a set of district and T&E provider-specific control dummy variables. The results indicate that attrition is not correlated with treatment status, and hence differential attrition between treatment and control groups is not an issue in this evaluation.

## **c. Uptake of EF-Sponsored Training Courses**

Another threat to our study is imperfect compliance, which occurs when those offered treatment choose not to participate, or when people from the control group gain access to the program. These deviations from treatment assignment may bias our estimates of program impact, because the factors that determine one’s actual participation, such as motivation and tenacity, cannot be observed and are likely correlated with the outcomes of interest.

---

<sup>40</sup> We have redone the baseline balance tests after matching, and the results indicate that though the similarity of treatment and control groups has improved, significant differences remain. The post-matching balance tests are available upon request.

<sup>41</sup> Because the EF-sponsored training courses vary in length from 1 to 3 months, the follow-up survey examines outcomes 9 to 11 months after the end of the training.

<sup>42</sup> The EF itself conducts follow-up with a sample of participants up to 6 months after the training to verify employment and earnings. Hence, the impact evaluation follow up survey occurs 3-5 months after the treatment group’s last contact with the program.

<sup>43</sup> The reasons given for loss to follow-up for the 2010 and 2011 cohorts include: inability to track the household (11%), no one in the household during multiple visits (15%), refusal (8%), respondent migrated for work within Nepal or abroad (8%), respondent migrated after marriage (10%), or other (40%).

Using administrative data from EF, we examine the rate of program take-up by the treatment and control groups for the 2010, 2011 and 2012 cohorts in Table 6.<sup>44</sup> The table shows a high degree of uptake (65 to 74 percent) among the treatment group, but also a high rate of participation among the control group.<sup>45</sup> Between 26 percent and 36 percent of the individuals in the control group participated in the EF training course that they applied for, even though their scores did not qualify them for admission.<sup>46,47</sup>

In the presence of imperfect compliance, standard impact evaluation methods produce intent-to-treat (ITT) estimates, as described in section IV. However, the relatively high degree of noncompliance in our study leads to a likely downward bias in our ITT estimates, which compare a treatment group in which not everyone received treatment to a control group in which some people did receive treatment, hence diluting the impact of the program. We present an alternative set of ATT estimates in Annex 3 (available upon request from the authors) in which we compare those who actually complete training to those who do not, irrespective of treatment assignment.<sup>48</sup> In our view, the ITT results serve as a lower bound for the estimated program impact while the ATT results serve as an upper bound.

## **VI. Short-Run Results of the EF Program**

This section describes the impact of the Employment Fund's training programs for the combined 2010, 2011 and 2012 samples. We measure outcomes approximately one year after the start of training.<sup>49</sup> We start with a description of our analysis on likelihood of being in the treatment group. We proceed with the short-term impacts of the EF program on the full sample and on various demographic and socioeconomic sub-groups.

### **a. First-Stage Estimation of Propensity Scores**

At the heart of the propensity score matching method is pairing trainees with non-trainees who are similar in terms of their observable characteristics. This pairing between a trainee and comparable non-trainees is done via a propensity score, as described in section IV.

---

<sup>44</sup> An individual is recorded in the EF monitoring database as a trainee only when the T&E provider submits the person's name at the end of the training. Since the EF only reimburses T&E providers at the conclusion of the training, they do not record nor do we have any way to track how many people enrolled, but did not complete, the training in 2010 to 2012.

<sup>45</sup> The monitoring records indicate that the control group individuals who received training did so in the original training course that they applied for. There is no evidence that they applied for and participated in a different EF training course sponsored by a different T&E provider, nor that they participated in EF training courses in a different year.

<sup>46</sup> In practice, while we cannot be sure what happened in each case, the T&E providers probably dipped into the pool of lower-scoring applicants in an attempt to fill up the training slots as people dropped out.

<sup>47</sup> The stable unit treatment value assumption (i.e., SUTVA) based on Rubin (1980) assumes that (1) the treatment status of any unit does not affect the potential outcomes of the other units (i.e., non-interference) and (2) the treatments for all units are comparable (i.e., no variation in treatment). Note that take-up of training among control group units is not a violation of the SUTVA, as take-up among non-beneficiaries was not directly a result of the take-up of training among the individuals selected for training.

<sup>48</sup> The ATT estimates suffer from an additional source of bias, since it is likely the "best" members of the control group who gained access to the program and are reassigned to treatment in the ATT analysis, and the least motivated among those assigned to treatment who fail to participate and are reassigned to control. This is discussed further in Annex 3, available upon request.

<sup>49</sup> Because the EF-sponsored training courses vary in length from 1 to 3 months, the follow-up survey examines outcomes 9 to 11 months after the end of the training.

Results from the propensity score estimation (shown in Table 7) exhibit only three variables correlated with an individual being offered training (*i.e.*, treatment status).<sup>50</sup> This finding reflects the high degree of similarity among short-listed candidates. The only characteristics correlated with an individual being assigned to a training program are being Muslim, the likelihood of having children, and the number of children one has. No other variables are correlated with training assignment, meaning that within the pool of shortlisted candidates, there are few observable differences between those selected for training and those who are not.

Based on the estimated propensity score, we match each treated individual (*i.e.*, a person assigned to a training program) to a group of control observations (*i.e.*, individuals who are non-trainees), as described in section IV.

### **b. Impacts on Employment and Earnings for the Full Sample**

In this section we present impacts of the EF program as estimated with the three identification strategies described in Section IV. Table 8 shows the ITT results on employment and earnings for the pooled 2010, 2011 and 2012 cohorts. We find (results in the first row of Table 8) strong evidence of consistent impact on the employment rate across all specifications.<sup>51</sup> All three models indicate a positive and significant effect, despite the high employment rate (*i.e.*, 61 percent) at baseline. Restricting the employment to non-farm activities, we also find a significant increase: the rate of participation in non-farm income-generating activities increases by 15 to 16 percentage points (from a base of 29.6 percent). Translating the results in percentage change terms, we find that the program increased non-farm employments by 50 to 54 percent. These impacts are not only statistically significant but also economically meaningful.

We also examine the trade-specific income generating activity (IGA) rate – the percent of individuals who find employment in the same trade as the training that they applied for – and we find impacts ranging from 18 to 19 percentage points. The trade-specific IGA impacts are larger than the non-farm employment impacts, suggesting that members of the control group, even when able to find employment, were less able than the treatment group to find employment in the trade in which they sought training.

Our results are considerably lower than estimates obtained by a simple before-and-after comparison within the group of trainees. For the 2010 cohort, for example, the treatment group had a non-farm employment rate of 30 percent at baseline and 55 percent at midline. A simple difference in employment rates would have suggested that the program increased employment rate by 25 percentage points (or 83 percent), substantially larger than our quasi-experimental estimate of 16 percentage points (or 50 percent).

The EF program also leads to persistent improvements in the underemployment rate (*i.e.*, cases in which people are working fewer hours than they wish). Table 8 shows that EF-sponsored training courses increased hours worked in IGAs for the pooled cohorts by 19-21 hours per month (*i.e.*, 28-30 percent). All three model specifications exhibit a statistically significant and positive impact.

---

<sup>50</sup> We report the ITT results, for both all three cohorts. For the ITT first-stage probit, treatment is defined as having a score that qualifies the respondent for admission to the EF training course to which s/he applied.

<sup>51</sup> We measure employment by whether the respondent reported any income-generating activities in the past month or not.

We detect strong program impacts on monthly earnings. We measure earnings as an individual's total earnings in the past month, including income from all IGAs, but not including unearned income.<sup>52</sup> We observe a statistically significant (at the 1 percent level) increase in monthly earnings for the treatment group by 850 to 921 NRs ( $\approx$  12 USD), from a baseline average of 1,272 NRs ( $\approx$  17 USD).<sup>53</sup> In percentage terms, this earnings increase translates to a 72 percent for the pooled sample.

With alternative measurements of earnings, we detect even larger program impacts. To account for the highly skewed nature of earnings distributions, we examine the impact on logged earnings and we find impacts of over 100 percent. A third way to examine the impact on earnings is to consider the proportion of participants who earned a "decent living." The Employment Fund considers 3,000 NRs per month ( $\approx$  40 USD) as "gainful employment" and considers this amount as "being productively employed." At baseline, only about 20 percent of the sample was "gainfully employed". The EF training program increased the "gainful employment" rate increases by 13 to 14 percentage points, a result statistically significant across all three models. The surveys of the 2011 and 2012 cohorts provide data on self-employment and work outside of home, conditional on having non-farm employment. We find an increase in self-employment (significant at the 5 percent level) and no increase in the proportion of people working outside the home.

### **c. Non-Employment Impacts for the Full Sample**

The last four rows of Table 8 present the estimated impacts on savings and borrowing behavior. We find no systematic impacts of the EF training program on loans. The program had a significant and positive impact on individual total savings – a year into the program, savings increased by 901 to 1,171 NRs ( $\approx$  12 to 15 USD).

Table 9 presents program impacts on a range of empowerment outcomes. Because empowerment is a multi-dimensional concept, we examine various aspects of empowerment – from psychological empowerment,<sup>54</sup> entrepreneurial self-confidence,<sup>55</sup> and financial literacy<sup>56</sup> to control over resources.<sup>57</sup> The results show a number of significant impacts of the EF training program. Relative to the control group, EF training participants report having more money of their own, more control over household spending, and more

---

<sup>52</sup> If an individual did not work in the past month, his/her earnings are recorded as zero.

<sup>53</sup> This average is based on the entire study cohort, including those with zero earnings at baseline. The average earnings among those with non-zero earnings were 2928 NRs, translating to a percentage increase in earnings of 30%.

<sup>54</sup> Psychological constructs related to self-confidence and self-regulation were measured using a series of questions to which the respondent indicates their level of agreement (using variations of a Likert scale), which were then aggregated into a single score. The self-regulation scale included questions on goal-setting and self-control. Preliminary reliability testing yielded a Cronbach's alpha of 0.77 for the self-regulation scale.

<sup>55</sup> The entrepreneurial score is based on the respondent's self-confidence to perform a series of tasks related to running a business or searching for a job. The specific tasks asked were: Find information about job opportunities in your community, Apply and interview for a job, Run your own business, Work in a team with 3-4 other people to accomplish a task, Identify income generating activities to start up a new business, Obtain credit from a bank, microfinance institution, CBO or NGO, Manage financial accounts, and Collect the money someone owes you. Respondents rated their ability to do these tasks on a scale of 1 to 5, and their responses were summed to form a single numerical score.

<sup>56</sup> Financial literacy is defined as responding correctly to at least one out of three questions indicating the respondent's familiarity with interest rates and ability to perform a mathematical computation (about one-third of respondents got the question right at baseline)

<sup>57</sup> Control over monetary resources is asked in a variety of ways in different sections of the survey questionnaire, including a direct question on whether the respondent has any money of his own that he alone can decide how to use, whether he/she can decide how to spend any earnings from employment or whether he/she needs permission, and whether he/she controls her own savings, and the extent to which he/she participates in decisions regarding common household expenditures such as food and medicine.

access to mentors who can advise them on work-related matters. These measures indicate strong gains in economic empowerment. We also see strong gains on psychological empowerment, including significant increases in self-confidence both in life and with regard to entrepreneurial activities. The gains are on the order of 0.1 to 0.15 standard deviation. We also find marginal impacts, equal to about 0.13 of a standard deviation, on the self-regulation ability of trainees, which reflects the respondent's self-reported ability to control impulses, delay gratification, stick to difficult or detail-oriented tasks, and exert control over what happens in their life, all of which are predictors of labor market success. Taken together, these positive impacts on six out of 10 empowerment indicators demonstrate how a skills training program and subsequent employment can benefit youth in a number of realms.

The final set of outcomes examined relate to reproductive health and household-level outcomes.<sup>58</sup> Table 10 shows that the EF program had no average effects on desired fertility. The program had no impacts on pregnancy. The EF training had no detectable impact on HIV knowledge, on household food insecurity,<sup>59</sup> or on protein consumption.<sup>60</sup>

Understanding the full array of potential advantages and disadvantages of job training entails considering the EF program impacts on trainees' out-migration. Nepal's youth labor force out-migration rates hover around 20 percent every year, with as many as 350,000 Nepalese leaving the country to look for employment opportunities elsewhere, including mainly in Persian Gulf countries and Malaysia. On the one hand, out-migration may improve outcomes for the migrants' family through remittances. On the other hand, investing in skills development for youth only to have those youth depart the country may cause concern for policy makers. This evaluation finds no evidence that the EF program increased out-migration, at least in the short run. Not only do we not find any increase on the receipt or amounts of remittances received by households of trainees (as shown in the last two rows of Table 10), but the fact that we were able to successfully re-interview 89 percent of the baseline survey respondents approximately a year after program enrollment indicates that almost all of them remained in the country.

#### **d. Trade-Wise Program Impacts on the Full Sample**

The Employment Fund sponsors about 600 training courses annually -- from short four-week courses on incense-stick rolling to three-month technical courses. The breadth of course offerings attracts individuals of varied training and career needs.<sup>61</sup> Our evaluation sample includes a bit more than 10 percent of the courses on offer each year (shown in Tables 1 and 2). Panel 1 of Table 11 shows the breakdown of courses by trade.

---

<sup>58</sup> The rationale for studying the effects of a job training program on reproductive health outcomes comes from the strong links between employment and fertility decisions, particularly for female youth. When first entering the labor market, many youth are making concurrent decisions about whether to pursue paid work outside the home, whether and when to start a family, and how to balance the two. Previous impact evaluations have shown that empowerment in one realm (such as employment) can affect decisions in other areas of life.

<sup>59</sup> Measured using 4 questions adapted from the standard FANTA scale, with a higher score representing more insecurity.

<sup>60</sup> Measured as whether the household consumes meat, fish, or eggs 5 or more times per week.

<sup>61</sup> The Employment Fund mandates that all graduates sit for a skills certification test overseen by the National Skills Testing Board (NSTB) for any trade in which such a test exists. Most of the EF training courses offer "Level 1" certification; in a few trades the EF's training providers also offer more advanced, or "Level 2" courses.

We grouped training courses into seven categories.<sup>62</sup> The most common categories of training in our sample are Electrical/Electronics/Computer (*e.g.*, electric wiring, computer hardware technician, and mobile phone repair), Construction/Mechanical/Automobile (*e.g.*, arc welding, brick molding, furniture making, motor bike service), and Tailoring/Garment/Textile (*e.g.*, *galaicha* weaving, garment fabrication, hand embroidery, tailoring and dressmaking). Because we sample approximately the same number of applicants per event, the breakdown of applicants by trade (Panel 2 of Table 11) is very similar to the event-wise breakdown.

Table 12 shows that the impacts of the skills training program differed markedly by type of trade for the pool 2010-2012 samples. Training in electronics, beautician, and tailoring consistently show strong ITT impacts on employment—graduates of these training programs are more likely to have employment in general and are also more likely to be working in the trade in which they were trained. Beautician training shows large impacts on both employment and earnings.

We detect no significant impacts on employment or earnings outcomes for the remaining four trades. Results for food and hospitality (*e.g.*, cooking and wait service) show no significant ITT impacts; results for construction show no significant impacts except for a marginal impact on trade-specific employment and on earning more than 3000 NRs per month.<sup>63</sup> For the remaining three trades (*i.e.*, poultry technician, handicrafts and farming), we detect some ITT impacts but they are not consistent across models.

Overall, the results in Tables 12 reveal substantial heterogeneity in employment outcomes across the various types of training. The positive and significant impacts discussed in Section VIb are driven almost entirely by three categories of trades: electronics, beautician training, and tailoring trades show positive and significant impacts on employment and earnings across both cohorts. We find no impacts for the food/hospitality and farming training. Construction-related trainings showed positive and significant impacts, but the effects are not consistent across outcomes.<sup>64</sup>

#### **e. Gender- and Age-Disaggregated Impacts**

An important objective of this report is to examine program impacts for men and women, and in particular for the “AGEI” population (*i.e.*, young women aged 16 to 24). To that end, Tables 13-18 disaggregate the results to compare outcomes for men versus women, and for younger women (the “AGEI” population) versus older women.<sup>65</sup>

Table 13 shows the differences in program impacts by gender. The employment impacts are significantly larger for women than for men, indicating that the EF program is more effective at increasing both overall

---

<sup>62</sup> NSTB categories dictated groupings.

<sup>63</sup> Construction is a unique case in which some labor demand comes from the Middle East and other foreign countries; the fact that the construction trainings do not increase employment and earnings, regardless of location of employment, may help to explain why no impacts on remittances were observed in Table 15.

<sup>64</sup> The difference in impacts may be driven by the types of training falling under this category (the 2010 sample includes more brick-molding courses while 2011 includes more furniture making), the characteristics of the trainees (brick-molding is female-dominated), or may reflect differences in the demand for construction labor in those two years. With the exception of poultry training, all of the training categories were offered in several districts throughout the country, so it is unlikely that local market differences are driving the observed differences in results.

<sup>65</sup> For simplicity, we present only the Inverse Propensity Score (IPSW) regression model results.

employment and non-farm employment for female participants than for male. The results for other economic outcomes, such as hours worked, earnings, and type of employment, are similar for both sexes.

Several factors could account for the differential in employment impacts between men and women. First, when asked, T&E providers suggested that female students attend classes more and are more diligent than male students. We lack data on the attendance or completion rates of EF trainees; however, the rate of non-compliance with treatment assignment is equal for men and women, suggesting that this mechanism is unlikely to exert much influence on outcomes. Second, the Employment Fund introduced life skills training for women in 2011 in all of its training courses.<sup>66</sup> Because all women received life-skills training, we cannot disentangle the influence of this factor, from other program elements, on outcomes. A third explanation is simply that men start with a higher level of non-farm employment (47 percent compared to 20 percent for women at baseline) and hence it may be easier to make large gains on the extensive margin for women. A fourth possible explanation relates to the difference between the types of trades that men and women apply for. Although the Employment Fund specifically tries to encourage female participation in non-traditionally female trades, most of the training courses tend to be heavily gender-segregated. For example, men tend to dominate electronics and construction courses, while the tailoring and beautician trainings are comprised almost exclusively of females. As shown earlier, the tailoring and beautician trainings exhibit the largest employment impacts.<sup>67</sup>

The results in Table 14 indicate that the impacts on the employment outcomes for the AGEI sample (women aged 15 to 24) resemble the ones for older women (aged 25 to 35). We detect almost no statistically significant differences between younger and older women, meaning that the EF program appears to work equally well for the young women trained under the AGEI. The only difference we find between these two groups is that, conditional on employment, young women saw larger gains on obtaining work outside the home than older women.

Table 15 shows the differential impacts by gender on empowerment outcomes. Some of the average impacts shown in Table 9 are shown to be stronger for men or for women. For men, we detect a significantly larger impact on control over household spending, though the difference is only significant at the 10% level. For women, we detect a significantly larger impact on having a mentor, as well as a marginally larger gains in control over earnings and mobility relative to men. In Table 16, there are no significant differences in empowerment impacts between younger and older women.

The final set of results, in Table 17, examines the impacts on reproductive health and family-level outcomes by gender. While the full sample results in Table 10 showed no impact on desired fertility, disaggregating by gender reveals that the program had significant, though countervailing, impacts on men and women. For men, the EF program increased the desired number of children while for women the desired number of children went down. Despite this change in fertility preferences, we do not detect significant impacts on

---

<sup>66</sup> The forty-hour curriculum covered topics such as negotiation skills, workers' rights, sexual and reproductive health, and dealing with discrimination. Female students overwhelmingly responded positively to the life skills training, often claiming that it was one of their favorite parts of the course. The skills learned and the positive experience in this life skills training may contribute to the increased employment impact for women, which is line with the advice from experts in vocational training from around the world, who increasingly advocate for the inclusion of life skills in technical training programs

<sup>67</sup> It is entirely possible, in addition to the gender-related explanation, that occupations for which one obtains skills in the tailoring and beautician courses are more profitable (relative to other occupations) in the Nepali labor market.



pregnancy or actual fertility. Table 17 also shows a marginally larger impact on protein consumption for men relative to women. We detect no statistically significant difference between the effects for family and reproductive health outcomes (Table 18) for younger women in the AGEI sample compared to older women.

## **VII. Summary and Implications**

In this section, we summarize the impacts of the Employment Fund’s training program. The EF program aims to increase employment among Nepal’s disadvantaged youth through skills training and job placement support. It is one of the largest providers of Technical Education and Vocational Training (TEVT) in Nepal, and it works with local non-governmental training providers to offer courses of up to three months in a variety of trades, followed by six months of employment placement support.

To understand the effects of the EF program, we compared the experience of EF program trainees with the experience of comparable non-trainees, approximately one year after the program. Our sample included both male and female trainees, with a particular focus on the AGEI population (*i.e.*, females aged 16-24).

Approximately one year after each baseline survey, we conducted a follow-up survey. The response rates were quite high for both follow-up surveys. Overall, the survey firm was able to track and successfully interview 89 percent of the baseline survey respondents. We detect a high degree of training uptake (65 to 74 percent) among the group offered training, but also a high degree of participation among the control group. Between 26 percent and 36 percent of the individuals in the control group participated in the EF training course that they applied for, even though their scores did not qualify them for admission.

Using a combination of difference-in-difference and propensity score matching methods, we find positive and statistically significant effects on labor market outcomes for EF trainees on the following outcomes:

- employment rates (any or non-farm);
- finding employment related to the skill they learned;
- hours worked
- earnings; and
- the proportion of people earning more than 3,000 NRs per month

Individuals selected for EF training programs experience an increase in non-farm employment of 15 to 16 percentage points for an overall gain of 50 percent. We detect an increase in average monthly earnings of approximately 72 percent. The gains for individuals who actually complete the training (not just those who are selected for training) are even larger. These results are available upon request from the authors.

We also find strong impacts of EF training programs on economic and psychological empowerment. Trainees had more control over economic resources and stronger self-assessed self-confidence. In contrast to the findings on employment and empowerment outcomes, we find limited evidence of impacts on reproductive health or household level outcomes.

Alongside the sizable general impacts on employment outcomes, we find that training courses in electronics, beautician services, and tailoring underpin most of the EF program’s impacts. These three

categories of training are much more effective in consistently increasing employment and earnings than construction, poultry rearing, handicrafts, and food preparation and hospitality.

We also find larger impacts on employment for women than for men. Women selected for training in 2010 to 2012 experience overall and non-farm employment gains of 13 and 19 percentage points respectively, while the corresponding impacts for men are 2 and 10 percent. However, we find no significant differences by gender on other economic indicators such as earnings, hours worked, trade-specific employment, or savings and loans.

The impacts for young women aged 16 to 24 are not significantly different from those for older women aged 25 to 35. This finding highlights the suitability of TEVT programs for younger women, although it may require specialized outreach strategies to recruit them. The larger impacts for women overall are consistent with the findings of rigorous evaluations of several training programs in Latin America, though uncovering the reasons for these differences requires further research.

The EF training program impacts are among the largest for training programs that have been evaluated around the world. We posit two factors as possible explanations of these large impacts. First, the EF had time to become established and to develop systems prior to the launch of the impact evaluation in 2010. Most impact evaluations are conducted during the pilot stage of project implementation, and consequently they do not always measure program impacts at their best. In this evaluation, the program was already operating at scale, and the service delivery processes already road-tested, which surely contribute to the large impacts. Second, the EF program is designed around employment outcomes. For instance, training providers must complete market assessments as part of their proposals to ensure future employability in the trades in which they propose to train individuals. Training providers are rewarded with outcome-based payments that are higher if they can demonstrate that their graduates have found work, a policy which fosters an incentive to not just train young people, but to find jobs for them. The fact that many training providers return year after year to work with the EF despite this unusual contracting arrangement speaks to the feasibility of this approach.

These unique program features, along with the results shown in this evaluation, provide valuable guidance and evidence for researchers, practitioners, and policy makers interested in improving employment opportunities for young people in Nepal and around the world.

## VIII. References

- Adoho, Franck, Shubha Chakravarty, Dala T. Korkoyah Jr., Mattias Lundberg, and Afia Tasneem.** 2014. "The Impact of an Adolescent Girls Program: The EPAG Project in Liberia" World Bank Policy Research Working Paper 6832.
- Aedo, Cristián and Sergio Nuñez.** 2004. "The Impact of Training Policies in Latin America and the Caribbean: The Case of Programa Joven" , Inter-American Development Bank Working Paper #R-483.
- Alzúa, María L., Guillermo Cruces, and Carolina Lopez Erazo.** 2013. "Youth training programs beyond employment. Evidence from a randomized controlled trial." Working Paper.
- Anderson, M. L.** 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects" *Journal of the American Statistical Association*, 103(484).
- Angelucci, Manuela and Orazio P. Attanasio.** 2006. "*Estimating ATT Effects with Non-Experimental Data and Low Compliance.*" , No. 2368. IZA Discussion Papers.
- Attanasio, Orasio, Adriana Kugler, and Costas Meghir.** 2008. "Training Disadvantaged Youth in Latin America: Evidence from a Randomized Trial". NBER Working Paper 13931.
- Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman.** 2012. "Empowering Adolescent Girls: Evidence from a Randomized Control Trial in Uganda". World Bank Working Paper No. 75566.
- Betcherman, Gordon, Martin Godfrey, Olga Susana Puerto, Friederike Rother, and Antoneta Stavreska.** 2007. "Global Inventory of Interventions to Support Young Workers: Synthesis Report" *World Bank*, Global Inventory of Interventions to Support Young Workers: Synthesis Report.
- Bryson, A., R. Dorsett, and S. Purdon.** 2002. "The use of Propensity Score Matching in the Evaluation of Labour Market Policies" , Working Paper No. 4, Department for Work and Pensions.
- Campbell, Donald T.** 1969. "Reforms as Experiments" *American Psychologist*, 24: 407–429.
- Card, D., J. Kluve, and A. Weber.** 2009. "Active Labour Market Policy Evaluations: A Metaanalysis" , Austrian Center for Labour Economics and the Analysis of the Welfare State Working Paper, (0902).
- Card, David, Pablo Ibararán, Ferdinando Regalia, David Rosas, and Yuri Soares.** 2007. "The Labor Market Impacts of Youth Training in the Dominican Republic: Evidence from a Randomized Evaluation." , NBER Working paper 12883.
- Fella, G. and G. Gallipoli.** 2007. "Education and Crime Over the Life Cycle" , Rimini Center 15-07.
- Frese, Michael and Gisela Mohr.** 1987. "Prolonged Unemployment and Depression in Older Workers: A Longitudinal Study of Intervening Variables" *Social science & medicine*, 25(2): 173-178.

**Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Vermeersch. Christel M.J.** 2011. *Impact Evaluation in Practice*: World Bank.

**Goldsmith, Arthur H., Jonathan R. Veum, and William Darity.** 1997. "The Impact of Psychological and Human Capital on Wages" *Economic inquiry*, 35(4): 815-829.

**González-Velosa, C., Ripani, L. and Rosas-Shady, D. (2012).** ¿Cómo mejorar las oportunidades de inserción laboral de los jóvenes en América Latina? Banco Interamericano de Desarrollo, Unidad de Mercados Laborales y Seguridad Social, Nota Técnica IDB-TN-345. Available online at: <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=36976183>.

**Groh, Matthew, Nandini Krishnan, David McKenzie, and Tara Vishwanath.** 2012. "Soft Skills or Hard Cash? The Impact of Training and Wage Subsidy Programs on Female Youth Employment in Jordan.". World Bank Policy Research Working Paper Series.

**Heckman, J., R. LaLonde, and Smith J.** 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics*, ed. O. Ashenfelter and David Card, 1865-2097. Vol. III. Elsevier, Amsterdam.

**Hirano, Keisuke, Guido W. Imbens, and Geert Ridder.** (2003). "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica*, 71(4): 1161–89.

**Hopenhayn, Martin.** 2002. "Youth and Employment in Latin America and the Caribbean: Problems, Prospects and Options", ECLAC Working Paper.

**International Labor Organization (ILO).** Global Employment Trends for Youth. 2013.

**International Labor Organization (ILO).** Youth Employment Statistics in Nepal. 2014. [http://apyouthnet.ilo.org/stats/youth-employment-statistics-in-nepal/at\\_download/file1](http://apyouthnet.ilo.org/stats/youth-employment-statistics-in-nepal/at_download/file1) (accessed September/1).

**International Labor Organization (ILO).** Labour market transitions of young women and men in Nepal. Work 4 Youth Publication Series No. 12. May 2014.

**Kling, Jeffrey R. and Jeffrey B. Liebman.** 2004. *Experimental Analysis of Neighborhood Effects on Youth*. John F. Kennedy School of Government, Harvard University.

**Kluge, Jochen.** 2006. "The Effectiveness of European Active Labor Market Policy", IZA Discussion Paper No. 2018; RWI Discussion Paper No. 37.

**Leuven, E. and B. Sianesi.** 2003. "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing and Covariate Imbalance Testing", 4.0.6 17may2012. (<http://ideas.repec.org/c/boc/bocode/s432001.html>).

**Linn, M. W., R. Sandifer, and S. Stein.** 1985. "Effects of Unemployment on Mental and Physical Health" *American Journal of Public Health*, 75(5): 502-506.

**Macours, Karen.** 2011. "Increasing inequality and civil conflict in Nepal". *Oxford Economic Papers*, 63 (1).

**Maitra, Pushkar and Subha Mani.** 2014. "*Learning and Earning: Evidence from a Randomized Evaluation in India.*" GCC Working Paper Series, GCC 14-05.

**Meyer, Bruce D.** 1995. "Natural and Quasi-Experiments in Economics" *Journal of Business and Economic Statistics*, 13(2): 151–161.

**Ñopo, Hugo, Jaime Saavedra-Chanduví, and Miguel Robles.** 2007. "Occupational Training to Reduce Gender Segregation: The Impacts of ProJoven" , Inter-American Development Bank Working Paper #623.

**O'Brien, Peter C.** 1984. "Procedures for Comparing Samples with Multiple Endpoints" *Biometrics*: 1079-1087.

**Robalino, D., Friederike Rother, David Newhouse, and Mattias Lundberg.** 2013. "Youth Unemployment: A Development Agenda for the Next Decade" , World Bank Social Protection and Labor Discussion Paper no.1308.

**Rosenbaum, P., and D. Rubin.** 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–50.

**Rubin DB.** 1980. "Randomization Analysis of Experimental Data in the Fisher Randomization Test." *Journal of the American Statistical Association*, 75: 591–593.

**Smith, Jeffrey and Petra E. Todd.** 2005. "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators" *Journal of Econometrics*(125): 305–353.

**Todd, Petra E.** 2013. "How to Improve Women's Employability and Quality of Work in Developing and Transition Economies"  
[http://www.womeneconroadmap.org/sites/default/files/Todd\\_Improve%20Women%20Employability%20and%20Quality%20of%20Work.pdf](http://www.womeneconroadmap.org/sites/default/files/Todd_Improve%20Women%20Employability%20and%20Quality%20of%20Work.pdf).

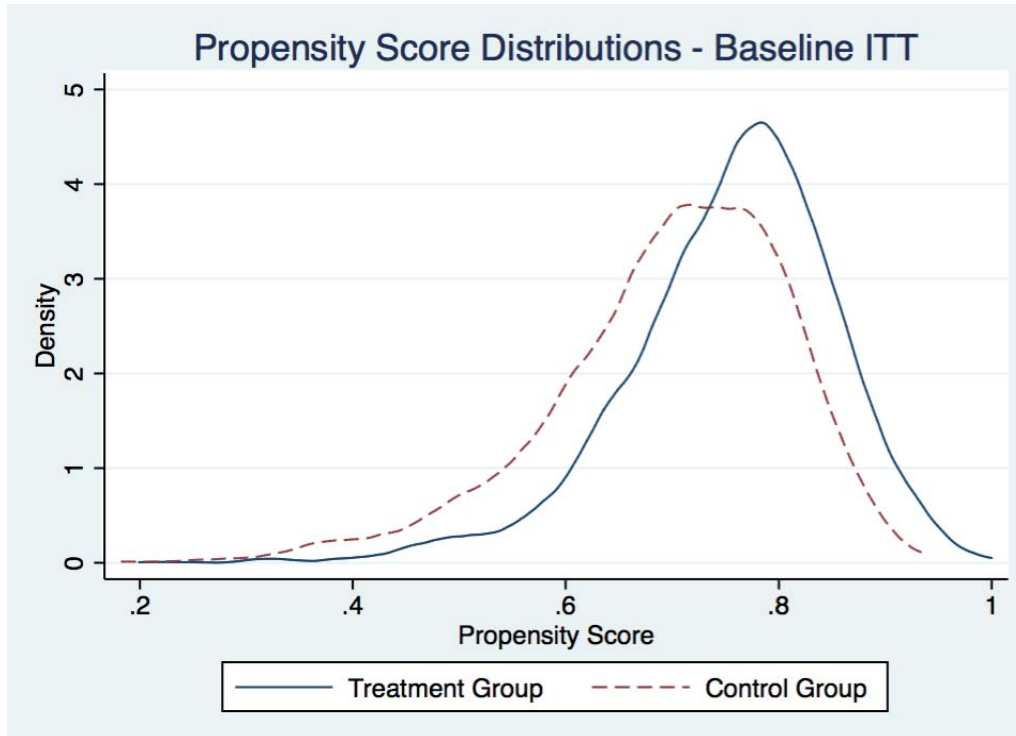
# Annex 1. Figures and Tables

## Figures

Figure 1. Timeline and Sample sizes of the AGEI impact evaluation

	2010			2011			2012		
<b>2010 Cohort</b> (n=1556; 1184 Treatment, 372 Control)	Baseline Survey	Training		First Follow-up Survey					
<b>2011 Cohort</b> (n=1586; 1237 Treatment, 349 Control)				Baseline Survey	Training		First Follow-up Survey		
<b>2012 Cohort</b> (n=1535; 1044 Treatment, 491 Control)							Baseline Survey	Training	

Figure 2. Distribution of Estimated Propensity Scores for the pooled 2010, 2011 and 2012 cohorts, ITT



## Tables

**Table 1. Scale of Employment Fund Programs and AGEI Sub-group**

All EF programs	2010	2011	2012
Total T&E providers	21	32	35
Total Events	598	645	711
Total trained	11750	12869	14255
AGEI only			
T&E providers working with AGEI	11	13	13
Total Events	110	218	246
Total trained	808	1664	1936

**Table 2. Summary of Sample of Events for Baseline Surveys**

	2010	2011	2012
Total # events conducted by EF in Jan-Apr	110	142	143
# events randomly sampled	N/A	182	112
# events included in baseline survey	65	69	85
# districts covered	30	34	29
# T&E providers covered	18	26	28

Notes: More events were sampled than conducted in Jan-Apr 2011 because some events that were scheduled for Jan-Apr were delayed and did not start on time.

**Table 3. Summary of Sample of Individuals for 2010, 2011 and 2012 Baseline Surveys**

	2010	2011	2012	Pooled
<i>Treatment group</i>				
Treatment- Male	478	447	458	1383
Treatment- Female	706	790	586	2082
Treatment- all	1184	1237	1044	3465
<i>Control group</i>				
Control- Male	138	139	203	480
Control- Female	234	210	288	732
Control- all	372	349	491	1212
<i>Pooled (treatment and control)</i>				
Male	616	586	661	1863
Female	940	1000	874	2814
<b>TOTAL</b>	1556	1586	1535	4677

**Table 4. Baseline balance tests for 2010-2012 Pooled Cohorts (ITT), Full Sample**

	Control	Treatment	Difference	p-value	N
<i>Demographics</i>					
Female (%)	0.640	0.630	-0.010	0.610	4101
AGEI (women aged 16-24) (%)	0.319	0.336	0.017	0.350	4101
Dalit (%)	0.090	0.077	-0.012	0.365	4037
Janajati (%)	0.421	0.468	0.048**	0.024	4037
Muslim (%)	0.017	0.025	0.008	0.269	4037
Age	24.537	24.242	-0.294	0.249	4101
Currently Married (%)	0.580	0.594	0.014	0.463	4101
Any Children (%)	0.505	0.526	0.021	0.248	4101
Completed SLC (10th grade) (%)	0.163	0.105	-0.059***	0.000	4101
<i>Employment</i>					
Any IGA in past month (%)	0.594	0.619	0.025	0.182	4101
Any non-farm IGA in past month (%)	0.266	0.307	0.041**	0.012	4101
Earnings in past month (NRs)	1201.970	1295.522	93.552	0.285	4069
Earnings > 3000 in past month (%)	0.172	0.197	0.025*	0.094	4101
Trade-specific IGA in past month (%)	0.154	0.189	0.035**	0.014	4101
Hours worked past month	62.774	71.502	8.728***	0.008	4101
<i>Empowerment</i>					
Any savings (%)	0.585	0.604	0.019	0.311	4080
Total Cash Savings (NRs)	3114.676	3177.379	62.703	0.832	4080
Entrepreneurship Score (0-32)	15.151	14.865	-0.286	0.235	4101
Financial Literacy	0.636	0.609	-0.028	0.109	4101
Analytical ability (Raven's test, 0-5)	2.944	2.911	-0.033	0.516	4101
Any mentor related to work	0.597	0.603	0.006	0.761	4099
Self-Regulation Score (0-16)	10.500	10.370	-0.131	0.167	4101
Self-Confidence Score (0-27)	19.015	19.014	-0.001	0.990	4100
Any money of your own (%)	0.702	0.673	-0.029*	0.083	4094
Control over earnings, if has earnings (%)	0.670	0.662	-0.008	0.697	2511
Control over own savings, if has savings (%)	0.600	0.561	-0.040*	0.073	2466
Control over household spending (0-8)	3.401	3.351	-0.050	0.685	4088
Mobility (0-9)	4.773	4.692	-0.081	0.179	4101
<i>Family and Reproductive Health</i>					
Number of children	1.130	1.124	-0.006	0.911	4101
Currently Pregnant (self or partner, if applicable)	0.038	0.038	-0.000	0.971	3923
Desired number of children	2.220	2.180	-0.040	0.249	4099
Use any type of contraception, if applicable (%)	0.754	0.773	0.019	0.305	2613
HIV Knowledge score (0-3)	1.124	1.155	0.031	0.318	4046
Household food insecurity (0-4)	0.540	0.555	0.015	0.534	4007
HH consumes eggs, meat, fish ≥ 5 times per week	0.348	0.347	-0.001	0.958	3582

Notes: This table reports average values for treatment and control groups, with p-value of a Student's t-test for equality of means between the two groups. The tests are conducted on the panel sample (those interviewed at baseline and follow-up). Standard errors are clustered by training course. "ITT" indicates that treatment is defined as having a score that qualifies the respondent for an EF training course. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.



**Table 5. Survey response rates**

	Baseline	Follow-up	Follow-up rate
<i>2010 cohort</i>			
Treatment	1184	1047	88.43%
Control	372	330	88.71%
<b>Total</b>	<b>1556</b>	<b>1377</b>	<b>88.50%</b>
<i>2011 cohort</i>			
Treatment	1237	1113	89.98%
Control	349	306	87.68%
<b>Total</b>	<b>1586</b>	<b>1419</b>	<b>89.40%</b>
<i>2012 cohort</i>			
Treatment	1044	889	85.15%
Control	491	417	84.93%
<b>Total</b>	<b>1535</b>	<b>1306</b>	<b>89.40%</b>

Notes: This table uses “ITT” treatment status—those whose scores qualify them for training.

**Table 6. Uptake of EF-sponsored training in the year following baseline survey**

	Participated in an EF training course		Did not participate in an EF training course	
	Number	Percent	Number	Percent
<b>2010 Cohort (n=1372)</b>				
Assigned to Treatment (n=1040)	671	64.52%	369	35.48%
Assigned to Control (n=332)	86	25.90%	246	74.10%
<b>2011 Cohort (n=1415)</b>				
Assigned to Treatment (n=1110)	826	74.41%	284	25.59%
Assigned to Control (n=305)	110	36.07%	195	63.93%
<b>2012 Cohort (n=1306)</b>				
Assigned to Treatment (n=889)	597	67.15%	292	32.85%
Assigned to Control (n=417)	127	30.46%	290	69.54%

Notes: There are four individuals from the 2011 cohort and five individuals from 2010 whose status in the EF database is unknown. For these individuals, we rely on the respondent’s self-report of whether they took an EF training in the past year for the ATT results. The table only includes those individuals who were surveyed for the first follow-up.

**Table 7. First-stage Estimation of Propensity Scores**

Dependent variable:	Treat (ITT)
Age of applicant	0.000 (0.009)
Sex of applicant (1=Female)	0.056 (0.094)
Education level of applicant	-0.010 (0.010)
Education of hh head	-0.003 (0.006)
Household size	0.007 (0.007)
Married (1=Yes)	-0.038 (0.074)
Has child (1=Yes)	0.260** (0.103)
Number of children	-0.079** (0.039)
Any IGA (1= Yes)	0.057 (0.072)
Zero earnings (1=Yes)	-0.075 (0.069)
Janajati (1=Yes)	0.031 (0.443)
Dalit (1=Yes)	0.436 (0.690)
Muslim (1=Yes)	4.127*** (0.357)
Analytical Ability (0-5)	-0.011 (0.019)
Entrepreneurial score (0-32)	-0.003 (0.004)
Financial literacy (1=Yes)	-0.047 (0.050)
N	4449
Pseudo R2	0.050

Notes: Standard errors (reported in brackets) clustered by training course. "Treat (ITT)" equals 1 if individual qualified for a training course and 0 otherwise. Other independent variables (not shown): district and T&E provider fixed effects, training-type categories, quintiles of household wealth. All variables measured at baseline. Although baseline data were collected on 4,677 individuals, incomplete data on ethnicity reduces the number of observations to 4,449. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.

**Table 8. Employment outcomes (ITT), 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts			
	Baseline mean [Std Dev]	OLS (1)	IPSW (2)	NN (3)
Any IGA (1=Yes)	0.612 [0.487]	0.071*** (0.022)	0.093*** (0.022)	0.070*** (0.020)
Any non-farm IGA (1=Yes)	0.296 [0.457]	0.149*** (0.023)	0.160*** (0.024)	0.150*** (0.021)
Trade-specific IGA (1=Yes)	0.18 [0.384]	0.182*** (0.023)	0.187*** (0.025)	0.184*** (0.020)
Hours worked in past month	69.261 [87.273]	18.740*** (3.890)	21.130*** (4.148)	19.014*** (3.940)
Earnings	1271.542 [2197.669]	856.087*** (152.941)	921.323*** (159.517)	850.880*** (135.139)
Logged earnings	3.291 [3.817]	0.957*** (0.191)	1.209*** (0.203)	0.975*** (0.173)
Earnings > 3000 NRs. (1=Yes)	0.19 [0.393]	0.130*** (0.021)	0.140*** (0.022)	0.131*** (0.020)
Self-Employed (1=Yes), if any IGA	0.317 [0.465]	0.057* (0.029)	0.063** (0.029)	0.060** (0.027)
Works outside of home (1=Yes)	0.576 [0.494]	0.005 (0.042)	0.016 (0.046)	0.001 (0.038)
Any Savings (1=Yes)	0.599 [0.490]	0.024 (0.023)	0.043* (0.023)	0.032 (0.021)
Total Savings (NRs).	3161.273 [7916.744]	901.440** (433.792)	1171.483** (469.584)	929.787** (450.135)
Logged Savings	4.506 [3.868]	0.335* (0.179)	0.497*** (0.179)	0.387** (0.164)
Taken out loan (1=Yes)	0.343 [0.475]	0.008 (0.021)	0.005 (0.021)	0.009 (0.019)
Clustered Standard Errors		Yes	Yes	No

Notes: All columns report difference-in-difference estimates. "ITT" indicates that everyone whose score qualified them for a given training event is included in the "treatment" group. Standard errors (reported in brackets) clustered at the event level where possible. Self-employment and location of work were not asked in 2010. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.

**Table 9. Empowerment outcomes (ITT), 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts			
	Baseline mean [Std Dev]	OLS (1)	IPSW (2)	NN (3)
Any money of your own (1=Yes)	0.681 [0.466]	0.055** (0.022)	0.052** (0.024)	0.058*** (0.021)
Control over Earnings (1=Yes), if respondent has earnings	0.664 [0.472]	0.041 (0.029)	0.047 (0.033)	0.046 (0.030)
Control Over Savings (1=Yes), if respondent has savings	0.571 [0.495]	0.041 (0.030)	0.035 (0.033)	0.043 (0.033)
Control over household spending (# categories out of 8)	3.364 [3.304]	0.221** (0.109)	0.249** (0.113)	0.246** (0.114)
Mobility (# places visited in last month out of 9)	4.713 [1.674]	0.035 (0.081)	0.015 (0.082)	0.040 (0.076)
Any Mentor Related to Work (1=Yes)	0.602 [0.490]	0.067*** (0.024)	0.068*** (0.024)	0.073*** (0.023)
Entrepreneurship Score (0-32)	14.938 [6.029]	0.751*** (0.279)	0.535* (0.275)	0.758*** (0.265)
Financial Literacy (1=Yes)	0.616 [0.486]	0.020 (0.025)	0.021 (0.027)	0.020 (0.023)
Self-Regulation Score (0-16)	10.403 [2.355]	0.330** (0.162)	0.296* (0.159)	0.285* (0.164)
Self-Confidence Score (0-27)	19.014 [3.312]	0.346*** (0.117)	0.375*** (0.114)	0.383*** (0.110)
Clustered Standard Errors		Yes	Yes	No

Notes: All columns report difference-in-difference estimates. "ITT" indicates that everyone whose score qualified them for a given training event is included in the "treatment" group. Standard errors (reported in brackets) clustered at the event level where possible. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.

**Table 10. Family and reproductive health outcomes (ITT), 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts			
	Baseline mean [Std Dev]	OLS (1)	IPSW (2)	NN (3)
Desired number of children	2.19 [0.767]	-0.001 (0.021)	-0.010 (0.024)	0.001 (0.020)
Pregnant (1= Yes), if married	0.038 [0.192]	0.005 (0.010)	0.007 (0.011)	0.006 (0.011)
Number of children, if respondent has children	1.126 [1.322]	-0.014 (0.012)	-0.013 (0.013)	-0.015 (0.012)
Ever argue with spouse/partner, if applicable (1=Yes)	0.799 [0.401]	-0.006 (0.023)	-0.002 (0.023)	-0.009 (0.024)
Use any type of contraception (1=Yes), if sexually active	0.768 [0.422]	0.022 (0.022)	0.019 (0.023)	0.018 (0.023)
HIV Knowledge score (0-3)	1.147 [0.764]	0.030 (0.040)	0.021 (0.041)	0.031 (0.039)
Household food insecurity (0-4)	0.552 [0.772]	-0.030 (0.029)	-0.061** (0.030)	-0.033 (0.032)
Weekly protein consumption (1= more than 5x per week)	0.348 [0.476]	-0.003 (0.029)	0.011 (0.028)	0.001 (0.025)
Household receives remittances (1=Yes)	0.241 [0.428]	0.016 (0.022)	0.008 (0.025)	0.016 (0.020)
Household received more remittances than last year (1=Yes)	0.104 [0.306]	0.019 (0.018)	0.014 (0.018)	0.019 (0.017)
Clustered Standard Errors		Yes	Yes	No

Notes: All columns report difference-in-difference estimates. "ITT" indicates that everyone whose score qualified them for a given training event is included in the "treatment" group. Standard errors (reported in brackets) clustered at the event level where possible. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.

**Table 11. Tabulation of training types, 2010, 2011 and 2012 cohorts**

Panel 1: Event-wise tabulation	2010		2011		2012	
	Number	%	Number	%	Number	%
Farming	0	0	0	0	5	6
Poultry	2	3	0	0	0	0
Food Prep/ Hospitality	11	17	3	4	2	2
Electrical/ Electronics/Computer	9	14	14	20	14	16
Handicraft & Incense	3	4	4	6	5	6
Construction/Mechanical/Automobile	20	31	13	19	30	35
Beautician /Barber	2	3	5	7	4	5
Tailoring/ Garment/Textile	18	28	30	44	24	28
<b>TOTAL</b>	<b>65</b>	<b>100</b>	<b>69</b>	<b>100</b>	<b>85</b>	<b>99</b>
Panel 2: Applicant-wise tabulation	2010		2011		2012	
	Number	%	Number	%	Number	%
Farming	0	0	0	0	92	7
Poultry	41	3	0	0	0	0
Food Prep/ Hospitality	195	14	38	3	32	2
Electrical/ Electronics/Computer	178	13	277	19	186	14
Handicraft & Incense	87	6	79	6	69	5
Construction/Mechanical/Automobile	413	30	258	18	457	35
Beautician /Barber	61	4	117	8	61	5
Tailoring/ Garment/Textile	415	30	650	46	396	30
<b>TOTAL</b>	<b>1390</b>	<b>100</b>	<b>1419</b>	<b>100</b>	<b>1306</b>	<b>99</b>

Notes: This table only includes panel observations (those who were interviewed at baseline and midline).

**Table 12. Employment outcomes disaggregated by trade (ITT), 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts IPSW Model ITT Effects			
	any nonfarm IGA (1)	trade- specific IGA (2)	monthly earnings (NRs) (3)	earnings > 3000 (4)
Full Sample (pooled across training types)	0.160*** (0.024)	0.187*** (0.025)	921.323*** (159.517)	0.140*** (0.022)
training: Farming n=92	0.155* (0.081)	-0.059 (0.104)	1167.151 (1000.983)	0.081 (0.169)
training: Poultry Technician n=41	0.226 (0.173)	0.342*** (0.099)	1139.704 (969.082)	0.189 (0.145)
training: Food prep/Hospitality n=265	-0.057 (0.096)	0.007 (0.064)	-965.418 (1048.109)	-0.146 (0.095)
training: Electrician & Electronics n=641	0.187*** (0.044)	0.258*** (0.058)	1282.843*** (359.255)	0.160*** (0.054)
training: Handicraft & Incense stick making n=235	0.107 (0.082)	0.207*** (0.075)	967.311* (524.717)	0.129 (0.094)
training: Construction n=1128	0.067 (0.054)	0.100* (0.058)	509.836 (322.866)	0.089** (0.038)
training: Beautician/Barber n=239	0.247*** (0.094)	0.402*** (0.089)	1529.259*** (533.151)	0.241*** (0.078)
training: Weaving/Tailoring/Garment Making n=1461	0.249*** (0.038)	0.222*** (0.037)	1185.755*** (233.399)	0.196*** (0.035)
Clustered standard errors (by event)	Yes	Yes	Yes	Yes

Notes: No poultry technician trainings were included in the 2011 sample. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.

**Table 13. Employment outcomes (ITT), disaggregated by gender, 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts IPSW model				
	Baseline mean for men [Std Dev]	Baseline mean for women [Std Dev]	Men (1)	Women (2)	Difference between men and women (3)
Any IGA (1=Yes)	0.774 [0.418]	0.518 [0.500]	0.025 (0.035)	0.130*** (0.028)	-0.106** (0.045)
Any non-farm IGA (1=Yes)	0.471 [0.499]	0.195 [0.396]	0.105** (0.044)	0.192*** (0.028)	-0.087* (0.051)
Trade-specific IGA (1=Yes)	0.295 [0.499]	0.113 [0.396]	0.147*** (0.046)	0.209*** (0.028)	-0.062 (0.053)
Hours worked in past month	107.772 [99.126]	46.887 [70.525]	11.564 (8.796)	26.287*** (4.242)	-14.723 (9.795)
Total monthly earnings (NRs)	2137.947 [2539.479]	774.683 [1796.025]	681.698** (300.488)	1036.088*** (173.214)	-354.390 (341.802)
Logged earnings	4.796 [3.917]	2.428 [3.476]	0.281 (0.341)	1.688*** (0.237)	-1.407*** (0.414)
Earnings > 3000 NRs. (1=Yes)	0.350 [0.477]	0.098 [0.297]	0.091** (0.039)	0.166*** (0.027)	-0.075 (0.047)
Self-Employed (1=Yes), if any IGA	0.324 [0.468]	0.313 [0.464]	0.020 (0.048)	0.086** (0.034)	-0.066 (0.058)
Works outside of home (1=Yes), if any IGA	0.736 [0.441]	0.430 [0.495]	-0.005 (0.054)	0.038 (0.074)	-0.043 (0.091)
Any Savings (1=Yes)	0.484 [0.500]	0.666 [0.472]	0.047 (0.037)	0.039 (0.029)	0.008 (0.047)
Total Savings (NRs).	2328.389 [6704.381]	3644.997 [8506.123]	1823.869** (808.152)	812.143 (583.697)	1011.726 (1004.908)
Logged Savings	3.619 [3.874]	5.021 [3.770]	0.537* (0.301)	0.463** (0.219)	0.074 (0.369)
Taken out loan (1=Yes)	0.379 [0.485]	0.322 [0.467]	-0.008 (0.036)	0.013 (0.025)	-0.021 (0.043)
Clustered Standard Errors			Yes	Yes	Yes

Notes: Standard errors (reported in brackets) clustered at the event level where possible. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.



**Table 14. Employment outcomes (ITT), disaggregated for women, 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts IPSW model				
	Baseline mean for young women [Std Dev]	Baseline mean for women [Std Dev]	Young women (1)	Older women (2)	Difference (3)
Any IGA (1=Yes)	0.5 [0.500]	0.543 [0.498]	0.140*** (0.040)	0.127*** -0.041	0.013 (0.058)
Any non-farm IGA (1=Yes)	0.168 [0.374]	0.225 [0.418]	0.196*** (0.041)	0.187*** -0.042	0.009 (0.062)
Trade-specific IGA (1=Yes)	0.096 [0.295]	0.131 [0.338]	0.213*** (0.036)	0.204*** -0.039	0.010 (0.048)
Hours worked in past month	39.569 [62.475]	55.560 [78.058]	25.348*** (5.746)	27.881*** -6.83	-2.533 (9.259)
Total monthly earnings (NRs)	560.537 [1438.980]	1026.533 [2113.341]	834.168*** (183.918)	1283.426*** -283.993	-449.259 (320.888)
Logged earnings	2.063 [3.264]	2.857 [3.665]	1.633*** (0.329)	1.791*** -0.366	-0.158 (0.505)
Earnings > 3000 NRs. (1=Yes)	0.071 [0.256]	0.131 [0.337]	0.144*** (0.030)	0.192*** -0.043	-0.048 (0.051)
Self-Employed (1=Yes), if any IGA	0.323 [0.468]	0.303 [0.460]	0.071 (0.043)	0.101* -0.054	-0.030 (0.067)
Works outside of home (1=Yes), if any IGA	0.434 [0.496]	0.424 [0.495]	0.193* (0.107)	-0.121 -0.093	0.314** (0.140)
Any Savings (1=Yes)	0.558 [0.497]	0.791 [0.407]	0.016 (0.045)	0.058 -0.039	-0.041 (0.062)
Total Savings (NRs).	1991.965 [6304.061]	5577.033 [10185.000]	-153.140 (809.785)	2026.673** -968.191	-2179.813 (1326.634)
Logged Savings	3.953 [3.685]	6.264 [3.483]	0.316 (0.332)	0.596* -0.322	-0.280 (0.489)
Taken out loan (1=Yes)	0.189 [0.392]	0.478 [0.500]	0.011 (0.032)	0.015 -0.039	-0.004 (0.050)
Clustered Standard Errors			Yes	Yes	Yes

Notes: Standard errors (reported in brackets) clustered at the event level where possible. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level. Younger women (aged 16 to 24) compared to older women (age 25 to 35).

**Table 15. Empowerment outcomes (ITT), disaggregated by gender, 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts IPSW model				
	Baseline mean for men	Baseline mean for women	Men	Women	Difference between men and women
	[Std Dev]	[Std Dev]	(1)	(2)	(3)
Any money of your own (1=Yes)	0.720 [0.449]	0.658 [0.475]	0.076** (0.036)	0.039 (0.031)	0.037 (0.047)
Control over Earnings (1=Yes), if respondent has earnings	0.729 [0.445]	0.608 [0.488]	-0.020 (0.042)	0.112** (0.052)	-0.132* (0.068)
Control Over Savings (1=Yes), if respondent has savings	0.557 [0.497]	0.576 [0.494]	0.087 (0.064)	0.016 (0.038)	0.071 (0.075)
Control over household spending (# categories)	3.535 [3.048]	3.264 [3.441]	0.519*** (0.167)	0.103 (0.152)	0.415* (0.232)
Mobility (# places visited in last month out of 9)	5.315 [1.512]	4.363 [1.665]	-0.175 (0.131)	0.119 (0.103)	-0.294* (0.165)
Any Mentor Related to Work (1=Yes)	0.730 [0.444]	0.527 [0.499]	0.005 (0.037)	0.102*** (0.031)	-0.097** (0.047)
Entrepreneurship Score (0-32)	15.644 [5.472]	14.529 [6.294]	0.551 (0.400)	0.518 (0.372)	0.033 (0.559)
Financial Literacy (1=Yes)	0.670 [0.470]	0.584 [0.493]	-0.003 (0.043)	0.034 (0.035)	-0.037 (0.055)
Self-Regulation Score (0-16)	10.638 [2.306]	10.267 [2.372]	0.380** (0.182)	0.373*** (0.144)	0.008 (0.229)
Self-Confidence Score (0-27)	19.165 [3.232]	18.927 [3.356]	0.025 (0.264)	0.434** (0.201)	-0.410 (0.339)
Clustered Standard Errors			Yes	Yes	Yes

Notes: Standard errors (reported in brackets) clustered at the event level where possible. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.

**Table 16. Empowerment outcomes (ITT), disaggregated for women, 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts IPSW model				
	Baseline mean for young women	Baseline mean for women	Young women	Older women	Difference
	[Std Dev]	[Std Dev]	(1)	(2)	(3)
Any money of your own (1=Yes)	0.621 [0.485]	0.704 [0.457]	0.058 (0.042)	0.021 (0.042)	0.037 (0.058)
Control over Earnings (1=Yes), if respondent has earnings	0.507 [0.500]	0.712 [0.453]	0.099 (0.082)	0.110* (0.064)	-0.011 (0.104)
Control Over Savings (1=Yes), if respondent has savings	0.626 [0.484]	0.536 [0.499]	-0.050 (0.069)	0.060 (0.050)	-0.110 (0.089)
Control over household spending (# categories)	1.822 [2.695]	4.940 [3.446]	0.053 (0.201)	0.158 (0.275)	-0.104 (0.369)
Mobility (# places visited in last month out of 9)	4.321 [1.597]	4.418 [1.733]	-0.012 (0.129)	0.283 (0.172)	-0.295 (0.222)
Any Mentor Related to Work (1=Yes)	0.511 [0.500]	0.546 [0.498]	0.119*** (0.045)	0.078 (0.051)	0.041 (0.072)
Entrepreneurship Score (0-32)	14.790 [6.262]	14.251 [6.317]	0.951** (0.468)	0.137 (0.596)	0.814 (0.773)
Financial Literacy (1=Yes)	0.591 [0.492]	0.577 [0.494]	0.053 (0.048)	0.012 (0.053)	0.042 (0.072)
Self-Regulation Score (0-16)	10.346 [2.418]	10.195 [2.318]	0.266 (0.207)	0.484** (0.196)	-0.218 (0.283)
Self-Confidence Score (0-27)	18.885 [3.337]	18.967 [3.364]	0.386 (0.279)	0.479 (0.327)	-0.093 (0.459)
Clustered Standard Errors			Yes	Yes	Yes

Notes: Standard errors (reported in brackets) clustered at the event level where possible. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level. Younger women (aged 16 to 24) compared to older women (age 25 to 35).

**Table 17. Family and repro health outcomes (ITT), disaggregated by gender, 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts IPSW model				
	Baseline mean for men [Std Dev]	Baseline mean for women [Std Dev]	Men (1)	Women (2)	Difference between men and women (3)
Desired number of children	2.141 [0.704]	2.219 [0.800]	0.098** (0.043)	-0.068** (0.027)	0.167*** (0.051)
Pregnant (1= Yes), if married	0.044 [0.205]	0.035 [0.184]	-0.012 (0.019)	0.017 (0.013)	-0.029 (0.023)
Number of children	0.857 [1.197]	1.281 [1.366]	0.011 (0.018)	-0.027 (0.017)	0.039 (0.025)
Ever argue with spouse/ partner, if applicable (1=Yes)	0.772 [0.420]	0.812 [0.390]	0.001 (0.050)	-0.002 (0.025)	0.003 (0.056)
Use any type of contraception (1=Yes), if sexually active	0.818 [0.386]	0.742 [0.437]	0.031 (0.037)	0.013 (0.028)	0.018 (0.046)
HIV Knowledge score (0-3)	1.244 [0.739]	1.090 [0.772]	-0.044 (0.057)	0.057 (0.055)	-0.101 (0.079)
Household food insecurity (0- 4)	0.607 [0.787]	0.521 [0.762]	-0.058 (0.052)	-0.062* (0.036)	0.003 (0.063)
Weekly protein consumption	0.273 [0.446]	0.388 [0.487]	0.080* (0.046)	-0.025 (0.035)	0.105* (0.058)
Household receives remittances (1=Yes)	0.202 [0.402]	0.263 [0.440]	-0.004 (0.038)	0.014 (0.031)	-0.019 (0.048)
Household received more remittances than last year (1=Yes)	0.078 [0.268]	0.119 [0.324]	-0.013 (0.027)	0.029 (0.025)	-0.042 (0.036)
Clustered Standard Errors			Yes	Yes	Yes

Notes: Standard errors (reported in brackets) clustered at the event level where possible. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.

**Table 89. Family and repro health outcomes (ITT), disaggregated for women, 2010, 2011 and 2012 cohorts**

	Pooled 2010-2012 Cohorts IPSW model				
	Baseline mean for young women [Std Dev]	Baseline mean for women [Std Dev]	Young women (1)	Older women (2)	Difference (3)
Desired number of children	1.968 [0.504]	2.501 [0.965]	-0.094*** (0.035)	-0.038 (0.043)	-0.056 (0.055)
Pregnant (1= Yes), if married	0.044 [0.205]	0.025 [0.157]	0.030 (0.021)	0.001 (0.014)	0.029 (0.025)
Number of children	0.468 [0.812]	2.221 [1.265]	-0.014 (0.026)	-0.040* (0.023)	0.026 (0.035)
Ever argue with spouse/ partner, if applicable (1=Yes)	0.751 [0.433]	0.849 [0.358]	-0.022 (0.050)	0.013 (0.031)	-0.035 (0.063)
Use any type of contraception (1=Yes), if sexually active	0.645 [0.479]	0.795 [0.404]	-0.008 (0.063)	0.023 (0.027)	-0.031 (0.068)
HIV Knowledge score (0-3)	1.173 [0.761]	1.000 [0.773]	0.128* (0.072)	-0.029 (0.094)	0.157 (0.124)
Household food insecurity (0-4)	0.514 [0.767]	0.528 [0.761]	-0.084 (0.057)	-0.034 (0.055)	-0.050 (0.087)
Weekly protein consumption	0.389 [0.488]	0.381 [0.486]	-0.075 (0.047)	0.027 (0.051)	-0.102 (0.069)
Household receives remittances (1=Yes)	0.231 [0.422]	0.304 [0.460]	-0.026 (0.045)	0.073* (0.042)	-0.098 (0.061)
Household received more remittances than last year (1=Yes)	0.118 [0.323]	0.121 [0.327]	0.006 (0.033)	0.060* (0.035)	-0.054 (0.047)
Clustered Standard Errors			Yes	Yes	Yes

Notes: Standard errors (reported in brackets) clustered at the event level where possible. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level. Younger women (aged 16 to 24) compared to older women (age 25 to 35).

**Table 19. Correlates of survey attrition**

Dependent variable: inpanel (=1 if the individual is observed at baseline and followup,= 0 if baseline only)

	2010-2012 Pooled Cohorts		
	ITT		
	(1)	(2)	(3)
treated	0.052 (0.056)	0.042 (0.056)	0.090 (0.090)
Female		0.472*** (0.116)	0.485*** (0.168)
AGEI		0.065 (0.105)	0.018 (0.109)
parent		0.062 (0.104)	0.006 (0.107)
married		0.136 (0.101)	0.084 (0.101)
dalit		0.259*** (0.097)	-0.250** (0.112)
janjati		-0.048 (0.062)	-0.039 (0.066)
any IGA at baseline		0.204*** (0.061)	0.150** (0.062)
age: under 25		-0.217** (0.088)	-0.242*** (0.093)
N	4677	4487	4487
District, TE dummies	No	No	Yes

Notes: All regressions use probit models. "District, TE dummies" indicates that the regression controls for district and training provider effects. Columns 2 and 3 also include training category dummies (not shown). All standard errors are clustered by event. \*, \*\*, and \*\*\* denote significance at the 10% level, 5% level, and 1% level.

## Annex 2. Example Ranking Form

Note: Red line indicates cut-off between accepted and rejected candidates

#	Name and Surname	Immediate contact telephone	Entry Requirement (Y/N)			Selection Criteria (Individual Scores)					Final Marks	Rank
			Age 16-35 (Write age)	Education < SLC	<6 mon. food sufficiency / < Rs. 3,000 mthly income	1 - 4 . Short-listing ( 70 % )			5. Interview (30%)			
						1. Trade-specific education (15)	2. Economic status (20)	3. Social caste ( 25)		4. Geographical rep (10)		
1	Jane Doe 1	12345678	21	Y	Y	15	20	20	5	26	86	1
2	John Doe 1	12345678	35	Y	Y	15	20	20	5	26	86	2
3	Jane Doe 2	12345678	23	Y	Y	15	20	20	5	25	85	3
4	John Doe 2	12345678	16	Y	Y	15	20	20	5	25	85	4
5	Jane Doe 3	12345678	27	Y	Y	15	20	20	5	23	83	5
6	John Doe 3	12345678	19	Y	Y	15	15	20	5	25	80	6
7	Jane Doe 4	12345678	37	Y	Y	15	15	20	5	25	80	7
8	John Doe 4	12345678	35	Y	Y	15	15	20	5	23	78	8
9	Jane Doe 5	12345678	22	Y	Y	15	15	20	5	23	78	9
10	John Doe 5	12345678	23	Y	Y	15	15	20	5	23	78	10
11	Jane Doe 6	12345678	25	Y	Y	15	15	20	5	23	78	11
12	John Doe 6	12345678	18	Y	Y	15	15	20	5	23	78	12
13	Jane Doe 7	12345678	20	Y	Y	15	15	20	5	23	78	13
14	John Doe 7	12345678	16	Y	Y	15	15	20	5	22	77	14
15	Jane Doe 8	12345678	18	Y	Y	15	15	20	5	22	77	15
16	John Doe 8	12345678	24	Y	Y	15	15	20	5	21	76	16
17	Jane Doe 9	12345678	25	Y	Y	15	15	20	5	21	76	17
18	John Doe 9	12345678	32	Y	Y	15	15	20	5	21	76	18
19	Jane Doe 10	12345678	20	Y	Y	15	15	20	5	18	73	19
20	John Doe 10	12345678	30	Y	Y	15	15	20	5	8	63	20