Vocational Training Programs and Youth Labor Market Outcomes: Evidence from Nepal[*]

SHUBHA CHAKRAVARTY[1], MATTIAS LUNDBERG[1], PLAMEN NIKOLOV[2,3], and JULIANE ZENKER[4†]

Current Draft: April 14, 2018

Lack of skills is arguably one of the most important determinants of high levels of unemployment and poverty. In response, policymakers often initiate vocational training programs in effort to enhance skill formation among the youth. Using a regression-discontinuity design, we examine a large youth training intervention in Nepal. We find, twelve months after the start of the training program, that the intervention generated an increase in non-farm employment of 10 percentage points (ITT estimates) and up to 31 percentage points for program compliers (LATE estimates). We also detect sizeable gains in monthly earnings. Women who start self-employment activities inside their homes largely drive these impacts. We argue that low baseline educational levels and non-farm employment levels and Nepal's social and cultural norms towards women drive our large program impacts.

Keywords: Training, Employment, Labor, Economic Development, Nepal, Regression Discontinuity, Gender.

JEL Codes: I21, I28, I38, J08, J24, O15

[†] Corresponding Author: jzenker@uni-goettingen.de
[1] The World Bank
[2] State University of New York (at Binghamton), Department of Economics. Email: nikolov.pv@gmail.com
[3] Harvard University Institute for Quantitative Social Science
[4] University of Göttingen.

I. Introduction

In much of the developing world, unemployment among the youth is extremely high: youths (ages 16-24) compose 40 percent of the world's unemployed while accounting for only 17 percent of the world's population (United Nations, 2012; ILOSTAT, 2017). This unemployment is not only related to poverty, but also has a strong influence on other important social outcomes: it impacts crime rates (Blattman and Annan, 2016; Fella and Gallipoli, 2014), depression prevalence (Frese and Mohr, 1987), substance abuse rates (Linn et al., 1985), and rates of social exclusion (Goldsmith et al.,1997). Moreover, the low labor force participation of women is particularly pronounced in many regions of the world, resulting in female-specific consequences of un(der)employment, such as low decision-making power in the household, and domestic abuse (Majlesi, 2016; Lenze and Klasen, 2016). Therefore, targeting youth unemployment with effective interventions, specifically taking female needs into account, is one of the highest priorities for low-income countries (World Bank, 2013). Although there are many determinants for high levels of unemployment and poverty, lack of skills is arguably one of the most important (Heckman et al., 2004; Malamud and Pop-Eleches, 2010). One common policy response in effort to enhance skill formation among the youth is vocational training programs. To date, more than 700 youth employment programs from 100 countries have been implemented and more than 80 percent of them offer skills training.[5]

In this paper, we examine one of the largest youth training interventions in Nepal, which serves almost 15,000 poor and disadvantaged young men and women annually by subsidizing skills training and employment placement services. Youth employment rates (other than in subsistence agriculture and informal activities) are exceptionally low in Nepal. In particular, women face difficulties to enter the non-farm sector or find paid employment. In combination with a comprehensive advertisement strategy, called the Adolescent Girls Employment Initiative (AGEI), the program specifically targeted young Nepali women. Despite the rapid expansion of skill-enhancement employment programs across the world, this study is among the few to rigorously evaluate such a program in the context of a low-income country.

We examine the program's labor market outcomes based on quasi-experimental techniques. We use a large panel data set of three consecutive cohorts of applicants to the program. Program eligibility was based on individual scores determined by a standardized application procedure

---

[5] See Youth Employment Inventory (http://www.youth-employment-inventory.org/)

and a course-based threshold score. Therefore, we exploit the individual-assignment scores to implement a regression discontinuity design. Because we find some evidence that the actual individual assigned score was manipulated in practice, we instead use application-form data to reconstruct the underlying score components and generate our own individual-based score. We then use the reconstructed forcing variable as an instrument for training to estimate the program impacts on compliers (LATE estimates) and as treatment variable to estimate the programs intent-to treat (ITT) effects.

We report three major findings. First, approximately twelve months after the start of the training program, the intervention generated an increase in non-farm employment of at least 10 percentage points (based on ITT estimates) and up to 31 percentage points for compliers (based on LATE estimates), both of which are heavily driven by women starting self-employment activities inside (but not outside) their homes. The program also generated an average monthly earnings gain of at least 659 NRs (≈9 USD) (based on ITT estimates) and up to 2,113 NRs (≈28 USD) (based on LATE estimates) for women. Second, in contrast to women, men do not significantly gain from the program along the extensive margin of non-farm employment in the short run, but do show an increase in earnings conditional on any employment of at least 698 NRs, which suggests that they use the program to upgrade their skills. Third, using a small sub-sample of the initial study population, we find suggestive evidence that 24 months after the start of the program, men have gained along the extensive margin with an increased non-farm (self)-employment rate of at least 19 percentage points, while female medium-term employment and earnings effects stay in a similar range compared to the short-term effects but, possibly due to the lower statistical power, turn insignificant.[6]

The magnitude of the effect sizes we find requires some further discussion to strengthen the credibility of our results. Much of the vocational training literature on similar programs in high- and middle-income countries finds low or insignificant effects (Card et al., 2010; Kluve, 2010). Generally, programs from middle-income countries are found to be more effective than programs from the industrialized world (Betcherman et al., 2007), which suggests that effects may be even larger for programs in very poor countries. To the best of our knowledge, only three other studies exist that rigorously evaluate similar programs in low-income countries: the large impacts we

---

[6] The rise in effect sizes from the short- to the medium-term is in line with results of other studies: Lechner et al. (2011) find average treatment effects on the treated in employment rates of 10–20 percentage points in West Germany. Also see Kluve (2010), Card et al. (2010), and Ibarraran (2015) for evidence on larger medium-term effects.

find are consistent with the results of a related experiment in Liberia. The experiment generated an increase of 47 percent in non-farm employment and 80 percent in earnings among young women (Adoho et al., 2014). Furthermore, Bandiera et al. (2012) examines the impacts of a field experimental project in Uganda that implemented village-level girls' clubs to provide life skills, reproductive health, and livelihood skills to young women aged 14 to 20; the study detected substantial LATE estimates in self-employment (63 percentage points from a baseline of 7 percent). In contrast, Cho (2016) reports that a similar program in Malawi enhanced skills but did not change labor market outcomes (based on a four months follow-up survey). The authors document that drop out rates were high due to the costliness of the program for participants. Additionally, female participation rates were particularly low: women reported to face time constraints based on their household and family responsibilities and were less supported by training providers, i.e. men were more likely to receive support in terms of financial help or paid work after the training completion.

A number of differences in the baseline education, unemployment and underemployment levels, job sector structures of the underlying economy as well as social and cultural norms suggest that the magnitude of program effects and effect channels might be very different in poor countries from those found in more developed economies or even middle-income countries. High poverty rates often go hand-in-hand with low educational levels and, to the best of our knowledge, the baseline educational levels in our study sample are the lowest among the existing experimental or quasi-experimental studies.[7] Low baseline human capital will imply a relatively steeper increase in output (and earnings). This relationship between low human capital and a large marginal increase in earnings may be further pronounced if gender inequality in educational attainment is high (see e.g. Knowles et al., 2002; Klasen, 2002). Moreover, low-income countries without unemployment insurance are often characterized by a large agrarian and informal sector: out of the pure necessity to survive, most individuals are involved in either subsistence agriculture or informal small-scale business, while formal employment activities are unavailable (Nayar, 2011). Thus, unemployment, as it is known in more developed countries, is

---

[7] Based on program eligibility criteria, in our study sample in particular, very few of the participants have completed higher than a 10th grade education (approximately 12 percent) and a total of 38 percent of individuals in our sample have no or primary education. This is in contrast to other studies from middle- and low-income countries. For instance, Card et al. (2011) reports an average years of schooling of approximately nine years at baseline. Alzua et al. (2016) reports that only 11 percent of the sample have primary education or lower, while 57 percent of the sample has completed secondary schooling or more. Hirshleifer et al. (2015) report, on average, eleven years of schooling at baseline. Maitra and Mani (2017) report that 45 percent of their sample has completed secondary education at baseline. Adoho et al. (2014) reports 46 percent of their sample to have attained an education level of 10th grade or above.

relatively rare, while underemployment is widespread. Comparing effect sizes from low-income to middle- or high-income countries is further complicated by the fact that the impacts found in existing literature on more developed economies are gauged along the lines of any employment and formal versus non-formal employment, while both are usually not useful in poor, mostly rural samples from low-income countries.[8] Studying vocational training in a poor country will likely not yield large employment effects on the extensive margin or in formal employment rates, but can be expected to show large effects on the intensive margin (hours worked) and changes in sectorial composition of employment (farm versus non-farm employment), which is the case in our study. Our estimated ITT impacts on the probability of any employment (4 percent, insignificant) or wage employment (3 percent, insignificant) are in fact in line or smaller than the coefficients on wage employment found in middle-income countries, see effects, for example, in Attanasio et al. (2011), Card et al. (2011), and Reis (2015), where coefficients range between 0 and 8 percent.[9] In contrast, our large effects on the probability of self-employment are unique to the low-income context – studies from middle-income countries usually find no impact on probability of self-employment or on self-employment earnings (see e.g. Attanasio et al., 2011; Maitra and Mani, 2017). Finally, social and cultural norms surrounding work can be expected to vary effects across countries. In traditional societies, women may be more confined by gender roles that constrain their involvement in the labor market. Restrictive norms regarding female labor force participation exist in many regions of the world, but they seem to be particularly strong for women in South Asia (World Bank, 2011; Asian Development Bank, 2015; Booth, 2016).[10] Further, certain ethnicities or castes may face obstacles to utilize their labor market potential as a result of discrimination by employers or lower access to education (Banerjee and Knight, 1985; Mainali et al., 2017). If a vocational training program manages to unlock this

---

[8] In our study, individuals employed in the formal sector make up only 1.5 percent of our sample at baseline and 2.5 at follow-up. Also, in the Liberia study of Adoho et al. (2014) only five to six percent of the sample is wage employed at baseline, while formal employment is virtually non-existent. In contrast to that, Attanasio et al (2011) report, a share of 7 and 12 percent of women and men who are active in the formal sector in their sample at baseline, respectively, and 23 and 38 percent at follow-up. Hirshleifer et al. (2015) observe a share of 30 to 43 percent formally employed participants in the control group.

[9] For the most part, these studies provide skepticism regarding the cost-effectiveness of training programs (Almeida et al., 2012). In contrast, the Liberia experiment mentioned above cost between $1200 and $1650 per beneficiary; although this is relatively high, the returns to the self-employment training are also sufficiently large to recoup this cost within three years (Adoho et al., 2014).

[10] Only one other quasi-experimental study in South Asia examines the impact of training programs on employment and earnings. Maitra and Mani (2017) evaluate a training program in stitching and tailoring offered to young women in New Delhi. They find that program participation increased employment by more than 5 percentage points, self-employment by almost 4 percentage points, and any employment by 6 percentage points. The program increased number of hours worked by only approximately 2.5 hours. Yet, the smaller effect sizes are based on ITT estimates, which are based on much lower compliance rates (56 percent) compared to our study (70 percent). Also initial levels of wage employment and education are higher in India as Nepalese socio-cultural norms regarding gender roles may be less restrictive in New Delhi compared to rural Nepal where most of our courses take place.

4

latent potential by taking the unique needs of particularly disadvantaged populations into account—as has been done in the program studied here—it might unleash large employment and earnings potential, whereas, in the opposite case, a similar program might prove less effective (see, e.g., Cho, 2016).

Although our nonfarm wage employment effects are similar to those in recent experimental interventions in developing countries, which find either insignificant or little overall impact on formal employment, our analysis shows unambiguously positive and significant effects on the probability of female non-farm self-employment together with a large increase in monthly earnings. We argue, that our effects may be particularly pronounced due to the low initial educational and non-farm employment levels of women in Nepal and the special social and cultural norms women face in the labor market. The striking fact that the program impacts in our study operate mainly through female self-employment inside but not outside the house hints to the socio-cultural norms in Nepal, which shape gender roles that identify women with more restrictive characteristics and capabilities in the labor market compared to other country contexts in the training literature (e.g., Latin America). Our results suggest that the program enables otherwise underemployed women to earn an income while staying at home – close to household chores and in line with the socio-cultural norms that bind them to the house and prevent them from taking up employment outside.

The remainder of the paper is structured as follows. Section II provides background information on the Nepal labor market and details Nepal's Employment Fund training program and the intervention design. Section III describes the data, sampling strategy and sample characteristics. Section IV presents the study design and Section V describes the empirical strategy. Section VI presents the results and Section VII provides various robustness checks. Section VIII concludes.

II. Background

A. The Labor Market in Nepal

Nepal's economy ranks among the world's poorest. In 2010, a quarter of the country's population lived below the national poverty threshold. The Nepalese economy is characterized by lack of formal sector jobs, a large informal sector and wide-spread underemployment (affecting approximately half of the workers in the younger age groups), all of which contribute

to very high poverty rates (ILO, 2004; Central Bureau of Statistics, 2009).[11,12] Based on the Nepal Labour Force Survey 2008, 74 percent of laborers work in the agricultural sector while 64 percent work in subsistence agriculture. Women are particularly underrepresented in non-agricultural employment. Although the Nepal Labour Force Survey 2008 reports a labor force participation of women (80.1 percent) that is similar to that of men (87.5 percent), approximately three quarters of employed women work in subsistence agriculture (compared to 52.9 percent of men), whereas only 26.2 percent are engaged in paid work (compared to 73.9 percent of men). Moreover, monthly earnings of female paid employees are much lower on average (NRs 3,402 versus 5,721 for men) (Central Bureau of Statistics, 2009).

Low skills and high youth unemployment. The country's labor market outcomes can, to some degree, be attributed to low levels of human capital accumulation. According to the Nepal Labour Force Survey 2008, the total literacy rate for individuals of ages 15 years and above is 55.6 percent (70.7 percent for men and 43.3 for women). Only 22 percent of women and 29 percent of men have more than primary education and approximately half of the Nepalese have no formal schooling at all (Central Bureau of Statistics, 2009). These numbers are staggeringly low even in comparison to the human capital indicators of other South Asian countries and the rest of the developing world. For instance, in Latin America, where most of the available experimental evidence on the effectiveness of training programs in developing countries comes from, approximately 75 (50) percent of the population of age 15 (20) or above had completed primary (lower secondary) education in 2004 (UNESCO, 2007). Moreover, lack of vocational skills is predominant among young Nepalese. Around half of a million young people join the labor force each year, the vast majority of them being unskilled. Although the market demand for skills is high, access to vocational training is limited, particularly among the poor and disadvantaged groups.

---

[11] Although no labor survey specifically collects informal sector data, some economic measures point to the size of the informal sector constituting more than two-thirds of the economically active population in Nepal (Suwal and Pant, 2009). One of the reasons for Nepal's bad economy and an its underdeveloped formal sector is the political turmoil the country experienced in the last two decades: a peace agreement between the government and Maoist insurgency, an interim constitution promulgated in 2007, the 2008 declaration of a democratic republic, new and rising ethnic political movements, and a democratically elected Constituent Assembly in 2013.

[12] Fewer individuals work between 20 to 40 hours per week, as 32 percent worked less than 20 hours per week and 22 percent work less than 40 hours. Nepal's 2008 labor force survey (Central Bureau of Statistics, 2009) shows that 21 percent of the individuals working less than 40 hours per week in fact are able to work more and are therefore underemployed. Shortage of employment opportunities has generated a migration flow to urban areas and migration to other countries, especially to the Gulf States (Lokshin and Glinskaya, 2009). Nepal's net migration has registered an outflow of migrants near half a million for the period 2008 to 2012. During the same period, Nepal's net migration rate has exceeded the same number for the overall migration rate of South Asia.

Cultural and Social Norms Regarding Female Employment. A combination of low educational attainment, restrictive norms towards marriage, childbirth and household duties generate multiple constraints to young women who wish to enter the labor market. These constraints are reflected in the different reasons young men and women report for not participating in the labor force: although the majority of economically inactive men between ages 15 and 29 report that they are attending school (85.3 percent), only 43.9 percent of women of the same age group report school enrollment as their major reason. Instead, 41.6 percent of economically inactive women between ages 15 and 29 and even 80.1 percent of women between ages 30 and 44 report to be economically inactive because they have to engage in (unpaid) household duties, while virtually none of the men in these age groups state household duties as a reason for being absent in the labor market (Central Bureau of Statistics, 2009).

In addition, being bound to their households and families because of gender roles, women also face three broad gender-based employment barriers, which further complicate their economic well-being: restricted mobility, cultural norms and societal norms regarding gender expectations for certain occupations. Largely influenced by Hindu philosophy related to men and women's positions in society, Nepal's socio-cultural practices differ by caste and ethnicities. For example, in southern Nepal (Terai and Madhesh), women are frequently confined to the household and are unable to travel outside of their immediate community for any work. Various proscriptions based on the Madhesi culture restrict females from leaving their homes. The second gender-related barrier relates to cultural norms that prohibit women from interacting with men other than their family members. For example, in conservative parts of Terai districts, females who are trained as barbers are not employed due to cultural prohibitions against women touching men. The third gender-based labor market barrier relates to women who choose to train and work in traditionally male-dominated occupations. In various typically male-dominated occupations – for example, computer and television repair, auto body making and construction work – customers frequently express doubts regarding the quality of skills of women assigned to such repair jobs. As a result of the societal difficulties women face, they remain engaged in unpaid, home-based labor to a large extent. Approximately 80 percent of unpaid family labor force is female (Central Bureau of Statistics, 2009).

In summary, the low educational levels and the shortage of skill training opportunities suggest that the return to obtaining additional training –especially for women – may be

particularly high in Nepal. With that said, training for women may only be successful if the program takes female-specific needs into account so that it can enable them to expand their economic activities while adhering to their social roles and cultural norms of their communities.

B. The Nepal Employment Fund

Started in 2008, the Employment Fund (EF) is one of the largest skills training initiatives in Nepal. The program provides vocational training and placement services under a unique governance structure in cooperation with local training providers. The EF program subsidizes short-term market oriented skill training in combination with other services for disadvantaged young women and men. The fund's objective is to place trainees into gainful employment upon training completion. Each year, the EF sponsors about 600 to 700 training courses. The EF-subsidized courses are announced publicly in local communities with the intent to encourage potential candidates to apply. The applicants are then selected by a standardized procedure based on eligibility criteria. Available seats are restricted due to the limited capacity of training providers. As of 2010, the program operates at scale nationwide and covers 54 districts, providing training for over 65 occupations and has expanded in the consecutive years. Table 1 provides an overview on the total number of training events, trainees, and training providers for the time of our study period (2010-2012).

Program components. Admission to the program offers the trainees a bundle of services, where the core components are technical training (including certification) and job-search-assistance. Training courses in technical skills vary across a wide range of trades (*e.g.,* incense stick rolling, carpentry, tailoring, welding and masonry) and last from four weeks to three months. Each trainee is encouraged to complete a skills certification test offered by the National Skills Testing Board (NSTB). Upon completion of the classroom-based training, the EF emphasizes job placement services. Once the first training phase is completed, training providers are required to link trainees to trade-specific employers for six months of paid on-the-job-training. Providers often use their trade-specific networks of trainers and employers to find a suitable internship for their graduates. Through the internship, trainees obtain immediate work experience as they apply their learned skills in the market and subsequently strengthen their social capital and contacts with potential employers. In addition to the core components of the program, all females receive

40 hours of life-skills training (started in 2010 and fully implemented in 2011). The forty-hour curriculum covered topics such as negotiation skills, workers' rights, sexual and reproductive health and discrimination response. A subset of trainees received a short course in basic business skills.

Outcome-based payments. Training providers are rewarded for their services in three installments based on a set of pre-determined outcomes. A provider qualifies to receive full payment – *i.e.,* the full price of the training and services provided plus a bonus based on whether the trainee belongs to a vulnerable group –by a set contract between the EF and the training provider. The contract stipulates that all accepted trainees must successfully complete their skills training and sit for a skills test given by the Nepal National Skills Testing Board (NSTB). When trainees complete the test, providers obtain their first installment (40 percent of total payment). Upon the exam completion, providers are expected to ensure that graduates remain continuously employed for the next six months and that they earn above 3000 NRs per month ("gainful employment"). The EF verifies three months and six months after training completion whether trainees are gainfully employed. If the verification process is successful, the training provider obtains the second and the third installments (25 percent and 35 percent of total payment, respectively).[13] The cost for training and employment services is pre-financed by the providers and is reimbursed to them only when they accomplish the outcomes in their contract. Therefore, training providers bear the risk of losing their investment if they are unsuccessful in training and placing their trainees accordingly.

Targeting vulnerable groups. Generally, three factors comprise the eligibility criteria for all EF-sponsored training programs: age (from 16 to 35), education (below SLC[14], or less than 10 years of formal education) and low self-reported economic status. Only applicants who meet all three criteria were viable for being short-listed in the admission procedure. Furthermore, and as mentioned above, providers receive a bonus payment for successfully training and placing candidates who belong to particular disadvantaged groups.[15] The bonus payment is calculated as

---

[13] Employment status of a randomly selected sample of graduates is verified by EF field monitors.
[14] The School Leaving Certificate (SLC) is obtained after successfully passing examinations after the 10th 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.
[15] Disadvantaged groups are defined by EF as people belonging to the Dalit community, ex-combatants, internally displaced, widows (only women), disabled, HIV/AIDS infected, and formerly bonded laborers.

a percentage of the full cost of training and services provided and is issued proportionally together with the three installments described above as long as the particular requirement for each installment is fulfilled. A provider receives a bonus payment of 40 percent of the base cost of training for a man who is poor and 50 percent for a man who is poor *and* belongs to a disadvantaged group. A provider further receives a bonus payment of 70 percent of the base cost of training for a woman who is poor and 80 percent for a women who is poor *and* belongs to a disadvantaged group.[16]

In 2010, the EF, partnering with the World Bank, made additional efforts to specifically target 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. In addition to regular training course advertisement, the EF sponsored radio and newspaper ads specifically geared towards young women. Many of these ads specifically encouraged women to sign up for traditionally male trades, such as mobile phone repair, electronics, or construction.

III. Data

A. Sampling

We used two primary sources of data. First, we used data from training application forms and the selection procedure of EF-sponsored trainings that covered three consecutive cohorts of applicants (from 2010 to 2012). Second, we conducted individual (applicant) and household surveys with two rounds of data collection for each cohort. For the 2010 cohort, a second follow-up was conducted on half of the cohort. Figure 1 shows the survey timeline.

[Figure 1 about here]

Sampling into this study included a combination of stratified, random and convenience sampling and was done in two consecutive steps. For each cohort, the first step consisted of a selection of training events and the second step consisted of selecting individuals according to standardized procedures. The event sampling-frame for this study consisted of all training events

---

[16] Poverty is defined as less than 6 months of food sufficiency for farm households or less than 3,000 per capita family income, from non-farm based income.

from the universe of EF funded trainings that occurred between January and April of each year. Events were grouped into clusters of close-by districts before sampling for survey administrative reasons. We then randomly sampled up to 15 district clusters in each of the three years, respectively. Further, from the list of training events that took place in these district clusters, we randomly selected 20 percent. Because of the focus on young women in this study, events that were likely to include more young women (identified by training providers) were purposely oversampled in 2011 and 2012. In 2010, because a complete event listing was not available in advance, the events were not chosen randomly but by convenience, based on scheduling and accessibility. Table 1 details the resultant sample of events for the three cohorts.

[Table 1 about here]

The 2010 sample comprised 65 events across 12 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 (usually due to weather conditions) or because the event was not "oversubscribed".[17] The remaining 69 events in 28 districts were included in the 2011 baseline sample. In 2012, 85 out of 112 sampled events were included in the study sample, covering 26 districts. Figure 2 depicts the study areas by survey.

[Figure 2 about here]

To sample applicants, a survey team visited each sampled training event on the day when applicant selection happened. Sampling of applicants was based on the standardized interview procedure that was used to determine assignment to training. During the assignment procedure (which we will describe in more detail below) applicants received scores in five different categories that were added up to form a total score. A ranking sheet was then used to list applicants from the top scorer to the bottom scorer and indicated the threshold (*i.e.,* minimum score) for admission to the course. The individuals we study comprise a subset of the ranked individuals -- those who fell in the range of 20 percent below or above the threshold.

---

[17] Oversubscription was necessary to obtain a sufficiently large "quasi"-experimental control group as detailed in the description of the applicant sampling below. The survey team was instructed to drop the event from the sample if there were not at least 3 rejected candidates who could be sampled for the control group (i.e., at least 3 candidates that fell within 20 percent of the threshold score).

Immediately following the sampling of applicants and before the results of the selection process were announced, a baseline survey was administered.

[Table 2 about here]

The sampling procedures resulted in a study sample of 4,677 individuals across all three cohorts at baseline, see Table 2. For the first follow-up surveys, we were able to track and successfully interview 88 percent of the baseline survey respondents, yielding a panel for analysis of 4,101 individuals.[18] Because training courses vary in length from one to three months, the follow-up survey examines outcomes nine to eleven months after the end of the training. The EF itself conducts follow-ups with a sample of participants up to six months after the training to verify employment and earnings. This is also the time point when providers receive their last payment installment. Hence, the first follow-up survey occurs three to five months after the training evaluation and the treatment group's last contact with the program.

B. Sample Characteristics

We present an overview of baseline characteristics in Table 3. Eligibility to apply for EF-sponsored training courses was restricted by age, education level, and poverty status. Therefore, individuals in our sample are on average young, low educated and relatively poor. For the pooled sample (i.e., 2010-2012 cohorts), the study population has an average age of 24 years and is 63 percent female. A total of 38 percent of individuals have either have no education or only primary education. Approximately 61 percent of the sample engaged in some income-generating activity in the month prior to the survey. When we restrict to non-farm income-generating activities, the employment rate falls to 30 percent. The average number of hours worked in the month previous to the survey was 69. At baseline, the average monthly earnings were 1,269 NRs in the month prior to the survey (equivalent to about 17 USD). This figure appears low as it represents the average earnings over the entire study population of 4,677 individuals, including those with zero earnings. Earnings conditional on any income generating activity were 2,082 NRs. Only 18 percent earned more than 3,000 NRs per month, a level deemed to represent

---

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

"gainful" employment. Further, 18 percent of the sample was already engaged in the same trade as the training to which they applied (denoted as "trade-specific IGA"), indicating that a significant minority of applicants had been looking to upgrade existing skills.

[Table 3 about here]

Generally, women have lower *paid* employment levels and earnings at baseline. Forty-seven percent of women engage in activities *inside* the house that yield some income (e.g., self-employment activities), while only 36 percent of the women engage in *paid* activities *outside* the house. In contrast, 59 (69) percent of men engage in *paid* activities *inside* (*outside*) the house. Also men (69 percent) are more likely than women (56 percent) to carry out *unpaid* work *outside* the house (*e.g.,* helping relatives). Yet, much more women carry out *unpaid* work *inside* the house (*e.g.,* household chores, child care). Almost all women (94 percent) work in the household without pay for at least five hours a week, while this is only true for 61 percent of men. Furthermore, 55 percent of the women work even more than 20 hours per week in the house without pay, which is only true for 12 percent of the men in our sample.

IV. Study Design

Admission to the program was based on a calculated score for each applicant and coupled with a course-specific threshold score. For each course, applicants with scores above the threshold were assigned to the training program, whereas applicants whose scores fell below the threshold were not assigned to the program. To form a sufficiently large quasi-experimental control group, training providers were advised to shortlist at least 50 percent more candidates than the number of spaces available in the training event. The assignment procedure followed streamlined guidelines, including a detailed scoring rubric, instructions for ranking the shortlisted candidates by score, and selecting the top-scoring candidates for admission to the program.

The individual score used in ranking candidates consisted of five sub-scores based on: (a) applicants' trade-specific education (prerequisite, 15 points)[19], (b) applicants' economic status

---

[19]Applicants had to fulfill course specific prerequisites (e.g. literacy, certain trade-specific knowledge or experience) to be eligible for the ranking procedure. If an applicant did not fulfill these prerequisites, he was then removed from the selection procedure. If applicants fulfilled the

(up to 20 points), (c) applicants' social caste, gender, and special circumstances (up to 25 points), (d) development status of applicants' district of origin (up to 10 points) as well as (e) a score determined by a selection committee during an interview procedure (up to 30 points). Sub-scores for the first four components were determined based on the information each applicant provided in his or her application form (shown in Figure A1 in Appendix A). The application form lists the exact questions on which the distribution of scores was based upon. Tables A1 and A2 show how these criteria were converted into numeric scores assigned within each of the four categories. Based on the aggregated four-component score, candidates were short-listed and invited for an interview. The fifth sub-score was determined based on an interview with a three- to five-member committee comprised of representatives from the training institution and potential employers. Moreover, representatives of the survey firm and/or the donor institutions (e.g. EF, World Bank, etc.) were usually present to observe the procedure. The selection committee jointly decided on the fifth sub-score by assessing the candidates commitment, motivation, attitude, aptitude, and clear vision for employment and enterprising. Eventually, the selection procedure yielded a total score for each individual by summing across the five components. Possible total scores ranged from 0 to 100.

Admission in each course was then based on candidates' decreasing rank; available seats were assigned starting with the top-scorer. Therefore, the threshold in each course was based both on the distribution of candidates' scores and on the number of pre-determined available seats. Because the distribution of scores and the number of seats determine the selection process, the threshold score varies for each course, something we take into account in the empirical strategy described in Section V. Figure A2 displays a sample ranking form used by training providers.[20] Although eligibility for training, based on the actual score, influenced the likelihood of training course enrollment, individual assignment to training was not automatic as it was originally envisioned for two main reasons: non-compliance and manipulation of the assignment procedure. We discuss the implications of these two factors in the remainder of this section.

As we examined the compliance with the selection process, we found that approximately 30 percent of the group assigned to the program did not take-up the training opportunity

---

prerequisites, they then received 15 points as their first score-component. In exceptional cases (approx. 9 percent of the sample) this rule was not adhered to and candidates received 0, 5, or 10 points.

[20] Data in each column approximately represents the distribution of the respective component in the full sample. In this example, 15 seats were available and the score of the 15th ranked person in the list was 73. Hence, in this example, 73 would be the threshold score for this particular course.

(possibly due to taking up other training or employment opportunities), whereas 32 percent of the non-assigned applicants did. These take-up numbers can be explained by a simple process in which seats not taken by individuals assigned to training were then given to individuals who were originally not assigned to training but rather next in line based on the ranking form. In the regression discontinuity estimation setup we use, an important step is to examine whether and how the probability of treatment jumped at the threshold. To account for the course-specific threshold, we subtract each individual's assignment score by the course-specific threshold score, whereby we obtain a standardized relative score around the cut-off of zero. We then plot the probability of treatment against the relative assignment score. Figure 3's left graph shows the results of this procedure. The plot reveals a clear jump in the probability of program participation at the cut-off. As expected, the jump is less than one, a fact that we incorporate in the estimation strategy we follow.

[Figure 3 about here]

To assess possible manipulation of the admission procedure we plot the density of the relative score (shown in the left graph of Figure 4). The plot reveals discontinuities in the distribution of the score around zero, which suggests that candidates' scores may have been manipulated to shift certain individuals across the assignment cut-off. In our scenario, such *precise* manipulation of the score was virtually impossible for candidates themselves, as it would have required access to the ranking form after the official selection procedure was completed and the course-specific threshold score was determined.[21] In contrast, providers may have had the opportunity to precisely alter the scores of those candidates who seemed favorable to them.[22] Although the selection committee included several persons from different interest groups and was designed to avoid this type of manipulation at a large scale, we cannot rule out the possibility that providers were able to manipulate the ranking sheet after the official selection procedure was completed. Because possible manipulation of the score can bias our estimates, we

---

[21] Lee and Lemieux (2010) distinguish between *precise* and *imprecise* manipulation. While applicants were certainly able to manipulate the information they gave in the application form aiming to raise their score, the forms were filled out long before the course threshold was determined, which only happened once all candidates were interviewed after the application forms where submitted to the providers. Applicants control over their score was therefore *imprecise*, which is actually what sorts them randomly around the cut-off. It is, therefore, not a threat to internal validity and not likely to have caused the discontinuity in our graph (Lee and Lemieux, 2010).
[22] Based on the above described payment scheme, providers had a motive to select candidates for the program who seem most employable or most disadvantaged. Also, favoritism or bribes from otherwise rejected candidates may have played a role for manipulation.

specifically address this manipulation issue by reconstructing the assignment score, described in more detail below.

[Figure 4 about here]

V. Empirical Strategy

To estimate program effectiveness on labor market outcomes, we use a non-parametric regression discontinuity strategy by running local linear regressions. Specifically, we estimate local average treatment effects (LATE) for the people who comply with assignment status, intent-to-treat (ITT) effects and heterogeneous effects by gender and trade within these two frameworks. In the following section, we describe the estimation strategy, the bandwidth selection and how we reconstruct the assignment score.[23]

A. Treatment Effect Estimators

Local Average Treatment Effects (LATE). To address imperfect compliance to treatment, we employ a fuzzy regression discontinuity set-up similar to the one proposed by Hahn, Todd, and Van der Klaauw (2001). We run the following first-stage equation:

$$D_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 (X_i - t_c) + \beta_4 T_i (X_i - t_c) + \epsilon_i , (1)$$

where the treatment dummy D_i indicates whether or not an applicant i has received training and T_i is the excluded instrument, specifying whether an applicant has been assigned to training (i.e., whether the absolute assignment score X_i of the applicant is greater or equal to the threshold score t of the respective course c he or she applied to). Furthermore, the forcing variable $(X_i - t_c)$ is the applicant's relative assignment score (i.e., it is the difference between an applicant's absolute assignment score and the threshold score t_c of the course). The predicted values of D_i are then used to run a second stage equation:

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 X_i + \gamma_3 (X_i - t_c) + \gamma_4 D_i (X_i - t_c) + \mu_i , (2)$$

---

[23] We follow the practical guidelines for regression discontinuity designs in Imbens and Lemieux (2008) and Lee and Lemieux (2010).

where γ_1 captures the local average treatment effect, or the treatment effect for the compliers, and coefficients γ_3 and γ_4 represent the different slopes of the linear regression line left and right of the cut-off, respectively. In both equations, we have added the absolute assignment score as control variable to account for the heterogeneity in the cut-off values across courses.[24]

It is important to note that the LATE estimate is not necessarily equal to the population average treatment effect as it is based only on those candidates who comply with program assignment. In our sample, it is likely that compliers actually have higher returns to additional education compared to the average individual, an assumption that we further explain in the results section below. We interpret the LATE estimates as the upper bound estimates of program impacts.

Intent-to-treat effects (ITT). The overall program effect (regardless of compliance) is of policy relevance. In an attempt to deal with the issue that the complier population might differ from the full sample, we employ an intent-to-treat regression discontinuity set-up in which we treat assignment to training as the treatment variable. We estimate the reduced form equation:

$$Y_i = \delta_0 + \delta_1 T_i + \delta_2 X_i + \delta_3 (X_i - t_c) + \delta_4 T_i (X_i - t_c) + \eta_i \,, (3)$$

where T_i is again an indicator that represents whether an applicant has been assigned to training or not. The coefficient δ_1 can be interpreted as the non-parametric local intention-to-treat effect (Lee and Lemieux, 2010) or the effect of *training assignment* on outcomes. This effect is likely lower than the population average treatment effect in our scenario because several candidates in the assigned group have not been trained, while several people in the non-assigned group have received training—likely biasing the estimate, δ_1, towards zero. We, therefore, interpret the ITT effects as the lower bound estimates of program impacts.

Heterogeneous local average treatment effects (HLATE). Because treatment heterogeneity has important implications for eliciting the mechanisms through which the program operates, we

---

[24] Based on suggestions in Cattaneo et al. (2016), we ran a graphical analysis that aggregates courses by the absolute threshold value to exploit fully all the information available by our multi-cutoff setup and explore how treatment effects may vary based on this, see Figure A3 in Appendix A. Unfortunately, our sample size is not large enough to draw really robust conclusions from this analysis as confidence intervals in the graphs are relatively large. However, some of the graphs do suggest that the program may have been less effective in courses with relatively higher thresholds (i.e., courses that are likely to be located in particularly poor areas or frequented by particularly poor or disadvantaged applicants).

estimate heterogeneous local average treatment effects (HLATE) based on the framework proposed by Becker et al. (2013). In particular, we estimate a two-stage procedure similar to the one described above with a second stage represented by the equation:

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 X_i + \gamma_3 (X_i - t_c) + \gamma_4 D_i (X_i - t_c) + \gamma_5 H_i + \gamma_6 H_i D_i + \mu_i , (4)$$

where H_i is an indicator for subgroup. In the first stage, we use the predicted probability of training and its interaction with the subgroup indicator as instruments for D_i and $H_i D_i$.

Furthermore, we estimate heterogeneous ITT effects by adjusting Equation (3) to include the subgroup dummy H_i and its interaction with the assignment indicator.

B. Determining Bandwidths

For choosing the optimal bandwidth we follow Ludwig and Miller (2007), who suggest a cross-validation procedure to find the optimal balance between precision and bias.[25] The cross-validation procedure chooses relatively large bandwidths in our sample, which in some cases include almost all candidates. We provide plots of the cross-validation functions for all our main outcomes in Figures A4 and A5 of Appendix A. In a robustness section below, we further investigate the sensitivity of our estimates to the choice of bandwidth.

C. Reconstructing the Assignment Score

In an ideal case, we can examine the effect of training provision on outcomes by using the individual scores assigned by the providers during the interview procedure. The discontinuity in training assignment induced by the threshold score theoretically should generate an exogenous change in the probability of training, holding individual characteristics constant. However, and as mentioned previously, we have a reason to believe that training providers were influencing the assigned scores – possibly in response to the payment structure, which rewarded training completion and trainee placement over drop-outs and non-placed trainees. Therefore, manipulating the individual scores is likely to be related to unobserved individual characteristics, and therefore likely to bias the estimates of interest (McCrary, 2008). Which direction this bias takes is not obvious. It is likely that applicants who seemed particularly employable were

---

[25] We apply the user-written program *rdbwselect* provided by Calonico, Cattaneo, and Titiunik (2014) to estimate the cross-validation functions that determine our bandwidths.

favored by providers, in which case our estimates would be upward-biased. On the other hand, it is also conceivable that providers favored vulnerable groups, as successfully training those groups was also incentivized with higher final payments. Finally, we cannot rule out the possibility that manipulation may have been the result of bribery or favoritism toward friends or relatives, in which case the direction of possible bias could go in either direction. To determine the size and the direction of the potential bias, we run balancing tests on relevant characteristics at the baseline using the LATE and ITT specifications described in Section V. Specifically, we examine the balancing of the outcome variable of interest at baseline in response to potentially influential characteristics like age, education, gender, and ethnicity, which are likely to determine labor market outcomes. We report these balance tests in Table 3. The difference tests reveal that the initial assignment, based on the original scores, does not perform very well in balancing relevant covariates as well as labor market potential at baseline. Assigned individuals are more likely to be male, less likely to be of Dalit ethnicity, more likely to have engaged in a non-farm wage employment activity in the past month, more likely to have worked more monthly hours, and more likely to exhibit higher initial earnings. In accordance with the incentivized payment structure, providers seem to have shifted those candidates across the threshold who appear to have been more employable. Moreover, given the imbalances of gender and Dalit ethnicity, which are in contradiction with EF incentives to focus training on vulnerable groups, it seems possible that providers may have used score manipulation as a risk reduction or risk diversification strategy. Overall, we conclude that using the original score will most likely bias our results upwards—for the estimation of treatment effects—as a result of its manipulation.

In order to overcome this challenge, we follow the approach by Miller et al. (2013), who reconstruct the 'actual' individual-specific score from survey data. Currie and Gruber (1996a, b), Cutler and Gruber (1996), and Hoxby (2001) also follow this approach. As we cannot exactly be sure which of the sub-scores have been subject to manipulation by the training providers, we reconstruct all five of them and later aggregate them to obtain a new total score. We use data from candidates' original application forms to assign three out of five sub-scores. As described above, this data is necessarily free of *precise* manipulation, as the forms were filled out long before the ranking sheet and the course thresholds were determined. Figure A1 in Appendix A shows the section of the application form that contains the relevant applicant information. We

assign points based on this information as well as the exact scoring rubric used for the original score (see Table A1 and A2).

For the remaining two score components, we follow two different strategies. The first sub-score refers to the applicants' trade-specific education and was initially meant to be 15 for all short-listed candidates. Usually, if candidates did not fulfill the course specific education prerequisites, they were not eligible for short-listing and immediately rejected. However, in exceptional cases (approximately nine percent of the sample) this criterion was not adhered to and instead applicants received 0, 5, or 10 points. We, therefore, reconstruct the first component based on an OLS specification that regresses the original first sub-score on candidates' general and course-specific education attainments, each interacted with a set of course dummies. To avoid the estimation bias induced by the aforementioned manipulation, we remove all candidates from the model who fall within five index scores of the cut-off (where most of the manipulation took place).[26] We then predict the outcomes (including all candidates) and round them to values that are factors of five in order to reflect the original distribution of the first sub-score.

For the fifth score component, the selection committee was supposed to assess employability and non-cognitive qualities of the applicants to rate their overall probability to successfully complete the program. If *precise* manipulation was applied to an applicant's fifth sub-score, it was likely carried out vis-a-vis the sum of the other four sub-scores, which were available at the time of the interview. Given the scoring rubric in Table A1, candidates who are better educated, less poor, less disadvantaged, and from a more developed district are ranked relatively lower, but may have higher potential to be successful in the labor market after training completion and, therefore, might be the more interesting candidates for the providers. If this is indeed the case, the incentive for manipulation would be positively correlated to the first score component and negatively correlated to the following three sub-scores. In other words, the higher (lower) the first (second, third, and forth) component-score, the higher the incentive is for the provider to secretly add points onto the applicant's fifth score in order to shift him or her over the threshold. In order to substantiate these considerations, we regress the manipulated fifth sub-score on the first four sub-scores and a set of course dummies. Results are presented in Table A3 of Appendix A. We find that the first four score components predict the interview score as

---

[26] The treatment effects presented below are qualitatively and quantitatively robust to a number of alternative specifications, including the use of a different range of index scores to drop individuals, or adding other baseline characteristics to the model.

expected (Column 1 and 2). In Column 3 we also add all possible interaction terms created from the four sub-scores of the model, which slightly improve its predictive power. Assuming that candidates' commitment and motivation, etc. is not (perfectly) correlated with education, being poor or being disadvantaged, the residual of this regression should now contain some relevant information on the selection committees assessment of the candidates' aptitudes. We therefore use the predicted residual of the model in Column 3 to create the fifth component in the reconstructed assignment score.

Because the points in the first four components were originally distributed by factors of five, we divide all score components by five to smooth the score and to minimize the heaping at multiples of five found in the original score. The final reconstructed score, therefore, ranges from 0 to 20.

D. Simulating the Assignment Threshold

The assignment indicator we need for our analysis is not only determined by the individual scores, but also by the course-specific threshold score, which is likely to be affected by the manipulation of providers as well.[27] To construct a valid instrument, we therefore re-estimate the threshold scores for each course following the approach proposed by Miller et al. (2013).[28] The authors' proposed approach rests on finding the optimal assignment variable based on a simulation exercise that maximizes the number of compliers given the reconstructed individual scores in a given course. Specifically, for each course, we run a set of simplified first stages similar to the one in Equation (1). We subsequently alter the threshold score, used to create the assignment dummy variable, T_i, from the lowest to the highest possible value. We then keep the threshold rule out of all possible assignment thresholds based on the specification that yields the largest R^2 for the respective course. Based on this optimal threshold, we proceed with calculating a reconstructed relative score, which serves as our new forcing variable.

E. Balancing Performance of the Reconstructed Score

Our empirical approach assumes that no individual characteristics (other than vocational training enrollment) that could influence the outcomes of interest vary discontinuously across the

---

[27] This is the case because the assignment threshold automatically moves with the distribution of the individual scores in each course.
[28] The authors follow Chay et al. (2005).

estimated eligibility thresholds. As a first cut, and to assess whether reconstruction of the score improves the differential sorting around the cut-off, we provide graphical evidence with respect to the density of the new forcing variable in Figure 4 (right graph). The density plot is significantly improved and does not appear to be discontinuous around the cut-off. We also employ the same balancing tests as before, now using the reconstructed score variable. Table 3's last two columns report the results. The new score successfully removes the imbalances we previously detected in all outcome variables as well as in the demographic characteristics at the baseline. Consistent with our assumption, estimates are not generally distinguishable from zero, except for the variables of age and Dalit ethnicity. In Figures 5 and 6, we present additional graphical evidence that outcomes and demographic characteristics are continuous around the cut-off of the running variable at baseline, with the exception of age and primary education. Age, being Dalit, and primary education are (practically) time-invariant characteristics in our sample. We, therefore, address the remaining unbalancing by following an estimation strategy on differenced outcomes.

[Figure 5 and 6 about here]

Additionally, we show evidence in Figure 5 and 7 that the subgroup indicators we use to estimate heterogeneous treatment effects (i.e., an applicant's gender and the trade of training) are continuous across the threshold. This continuity across the threshold confirms that assignment status is not correlated with interaction variables around the cut-off, which is an important condition necessary for the estimation of unbiased heterogeneous treatment effects in the regression discontinuity setup (Becker et al., 2013).

[Figure 7 about here]

VI. Results

A. Program Effects on Employment and Earnings

We now turn our attention to the impact on the combined 2010, 2011 and 2012 samples based on the specifications described in Section V. Table 4, Panel A shows the local average treatment effects (LATE) estimated from Equations (1) and (2) – *i.e.,* the effects the program had considering those individuals who complied with their assignment status – for employment and

earnings on differenced outcomes using the *original* score. We find relatively large estimates across all outcomes, including significant (conditional) employment effects of 22 to 61 percentage points, as well as average earning gains of more than 3,000 NRs. The F-statistics (which range from 32 to 67) and the highly significant coefficients of the assignment variable (β_1) presented in the same panel suggest a strong first stage. However, as pointed out earlier, the estimates are likely to be biased upwards because the possible manipulation of the original score variable led to unbalanced individual characteristics and outcome variables at baseline.

[Table 4 about here]

Therefore, in Panel B, we use the *reconstructed* score, which led to a significantly improved covariate balancing. The first stage is very strong with F-statistics ranging from 38 to 88 and coefficients of the assignment variable (β_1) that are statistically significant at the 1-percent level. Comparing the effect sizes of Panels A and B (first rows, respectively) reveals that manipulation resulted in a strong upward bias in almost all outcomes. Using the corrected score, we no longer find evidence of a statistically significant impact on the employment rate, measured by whether individuals self-report any income-generating activities in the past month (Column 1, Panel B). When we restrict attention to employment in all non-farm activities, we find a lower but still statistically significant increase (Column 2, Panel B). The rate of participation in non-farm income-generating activities increases by 31 percentage points (from a base of 29 percent) as compared to 52 percentage points from using the manipulated score (Column 2, Panel A). Converting the results in percentage change terms, we find that the program increased non-farm employment by almost 94 percent. These impacts are not only statistically significant, but they are also economically meaningful. Disentangling these impacts into wage- and self-employment activities suggests that the effect is strongly driven by self-employment activities. The program increased non-farm self-employment by 30 percentage points, whereas we do not detect statistically significant impacts on non-farm wage employment rates (Columns 3 and 4, Panel B). 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 of 40 percentage points (Column 5, Panel B). The increase in the employment rate within the same occupational fields for which the individuals trained implies that the skills

trainees acquired in the vocational training part of the program were useful in starting employment activities within their respective trades. The EF program also led to improvements in the underemployment rate (*i.e.,* cases in which people are working fewer hours than they wish). Column 6 of Panel B shows that EF-sponsored training courses increased hours worked in IGAs by 49 hours per month (*i.e.,* 71 percent).

When we examine program impacts on monthly earnings using the corrected score, we find that effect sizes are about two-thirds of the estimates in Panel A. The reported results still show large program impacts (Columns 7 to 10, Panel B). We measure earnings as an individual's total earnings in the past month, including income from all IGAs, but not including unearned income.[29] We observe a statistically significant increase in monthly earnings for the treatment group by 1,754 NRs (≈ 23 USD), from a baseline average of 1,260 NRs. Once we restrict earnings only to those individuals who engage in an income generating activity, we detect an increase in earnings by 2,025 NRs (≈ 27 USD), from a baseline average of 2,075 NRs. In percentage terms, the increases in earnings translate to 140 and 98 percent, respectively. With alternative measurements of earnings, we also detect large program impacts. To account for the highly skewed nature of earnings distributions, we examine for impacts on logged earnings and find very sizable increases. In a third approach, we consider the proportion of participants who earned a "decent living." The EF considers 3,000 NRs per month (≈ 40 USD) as "gainful employment" or "being productively employed." At baseline, only 18 percent of the sample was gainfully employed. The EF training program increased this rate by 31 percentage points.

In Panels A and B, we use differenced outcomes to address the (remaining) disparities in participant characteristics we observed at baseline. In addition, the results reported in the discussion above are based on individuals who complied with their assignment status and, as a result, may differ from individuals in the full sample. In Panel C and D, we examine how altering these two features—that of differenced outcomes and the existence of compliers—affect our results. When comparing the estimates on differenced outcomes and on outcomes in levels, (Panel B versus C) using the reconstructed score, the difference in effect sizes is minor across most results. The most important difference we detect is on the variable that captures conditional earnings (Column 10, Panel C), which increases by approximately one-third compared to the estimate from the differenced outcome in Panel B. While the assigned group is slightly younger

---

[29] If an individual did not work for pay in the past month, his/her earnings are recorded as zero.

on average, they are also less likely to be Dalit than the rejected group. Older individuals may have been employed in a particular IGA for a longer time or, in general, may have more work experience, which is likely to lead to a higher level of earnings. On the other hand, individuals of Dalit ethnicity face substantial discrimination in the labor market, which likely makes them earn less compared to non-Dalits. The reduced effect size when controlling for time-invariant factors suggests that ethnicity outweighs the differences in age.

Turning to Panel D, we examine the intent-to-treat (ITT) effects, based on Equation (3) in Section V. We also provide graphical evidence of these effects in Figure 8.

[Figure 8 about here]

As expected, we see sizeable differences: all ITT coefficients are substantially lower as compared to the LATE estimates. Although the ITT effects are smaller in size, they still indicate that training led to a significant increase in non-farm self-employment and trade-specific employment of 9 and 13 percentage points, an average increase of 15 hours per month of working time as well as an average rise in overall earnings of 572 NRs. The smaller effect sizes are to some extent due to the partial crossover between assignment groups (as we document above). Therefore, these coefficients can be interpreted as lower-bound estimates of the program effects.

Moreover, compliers in our sample may differ from non-compliers due to Nepal's labor market context, along with the program-specific targeting of particularly disadvantaged groups. Although the eligibility criteria of the program automatically exclude better-educated and non-poor individuals, the sample of applicants is still quite diverse along various characteristics that might affect the magnitude of the return an applicant can expect from participation in the program: in caste, in gender, in educational attainment, and in baseline intensity of poverty. It is entirely possible that the complier-population may be a subgroup of individuals for whom the returns to participation are larger than for the overall sample. For example, individuals who comply with program assignment may be otherwise unable to secure a self-paid seat in a similar program due to extreme poverty or caste- and gender-based discrimination. At the same time, these particular groups of applicants likely also have lower baseline educational levels and higher returns to additional education as compared to the rest of the sample. Therefore, we

interpret the complier-based LATE coefficient estimates as upper-bound estimates of the program's effects.

B. Gender-Disaggregated Program Effects

We also examine program impacts disaggregated for men and women (shown in Tables 5 and 6). The results reveal striking differences in the program's effectiveness by gender. For all outcomes, except for the conditional and unconditional earnings variables, the LATE estimates are larger and significantly different for women as compared to men. Except for the earnings conditional on employment outcome, none of the effects for males are statistically significant. The coefficients are, in fact, negative for most of the outcomes. In contrast, most effects for females are statistically significant, except for non-farm wage employment, and earnings conditional on employment. The ITT results paint a similar picture and, additionally, show weak evidence of non-farm wage employment gains for women. In summary, women seem to gain on the extensive margin across most outcomes (*i.e.,* employment rates, hours worked, and earnings) without significant gains on the intensive margin (*i.e.,* conditional earnings). In comparison, men exhibit gains exclusively on the intensive margin.

[Table 5 about here]

Several factors could explain the differential gender impacts on employment outcomes. First, the EF introduced life-skills training for women, in 2010, in all of its training courses. 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. Because all women received life-skills training, we ultimately cannot disentangle the influence of this particular factor from other program elements. Second, men exhibit higher non-farm baseline employment than women (49 percent compared to 18 percent for women) and therefore it is easier for women to make larger gains on the extensive margin. Third, women face higher labor market barriers as compared to men. We investigate this third potential explanation in Table 6,

26

where we examine time allocation decisions by men and women for paid versus unpaid work inside and outside the household more closely. The results in Table 6 present a compelling story: program impacts on employment seem to be strongly driven by women who start self-employment activities *inside* the house, whereas unpaid work inside the house, and activities outside the house, remain unaffected by the program. These results suggest that the program is particularly effective at placing women into income generating activities while they remain at home – in this way, they conform to Nepal's social norms that restrict female mobility and bind them to household responsibilities.

[Table 6 about here]

C. Trade-Wise Program Effects

We examine program effects by type of trade and we classify training courses into nine categories. The most common training categories are: Tailoring/Garment/Textile (*e.g., galaicha* weaving, garment fabrication, hand embroidery, tailoring and dressmaking), Construction/Mechanical/Automobile (*e.g.,* arc welding, brick molding, furniture making, motor bike service), and Electrical/Electronics/Computer (*e.g.,* electric wiring, computer hardware technician, and mobile phone repair). These are followed by trainings related to Food preparation/Hospitality skills, Beautician/Barber skills, and Handicraft/Incense making skills. Finally, a few events in our sample are related to Farming, Poultry, and Security Guard skills training. Table 7 shows the breakdown of courses by trade. Although the EF specifically tried to encourage female participation in non-traditionally female trades, most of the training courses tended to be heavily gender-segregated. For example, men tended to dominate electronics and construction courses, whereas the tailoring, handicraft, and beautician training sessions were almost exclusively comprised of women.

[Table 7 about here]

Table 8 examines program impacts for the six most common training categories. Due to the low number of courses and participants, we dropped the remaining three categories from the analysis. Table 8 shows that the impacts of the skills training program differed markedly by type of trade. Trainings in beautician and tailoring consistently show strong impacts on

27

employment—graduates of these training programs are more likely to have any (non-farm) employment and are also more likely to be working within the trade for which they were trained. Both trades also show large impacts on monthly hours worked and some of the earnings indicators.

[Table 8 about here]

For the remaining four trades, we do not detect conclusive, and significant, positive impacts. The coefficients of the trainings related to food preparation and hospitality even seem to reflect some negative influence of training on labor market outcomes (although these coefficients are mostly insignificant). Overall, the results in Tables 8 reveal substantial heterogeneity in employment and earning outcomes across the various types of training. The positive, and significant, impacts previously discussed are driven almost entirely by two categories of trades: beautician training and tailoring, both of which are almost entirely occupied by female trainees, which is in line with the large gender-differences in program impacts presented earlier.

VII. Robustness Checks

Next, we present a number of consistency checks to investigate whether the magnitude of our estimates are within a reasonable range. We also provide additional robustness checks to address possible identification threats related to attrition issues or specification choices.

A. Magnitude of Program Impacts

Second Follow-up. Most of the employment results we find are driven by self-employment activities and, hence, they do not seem to be a direct effect of the job-search-assistance component of the program. Still, providers could eagerly support trainees even in their entrepreneurial efforts in order to secure the last two installments of EF payments. The EF verified placement 3 and 6 months upon training completion and released the final payment installment to providers, if and only if, they verified that trainees were in fact engaged in gainful employment. Therefore, it is possible that the employment effects we detect 9 to 11 months upon completion of training might be a mechanical effect of the program design. To examine this hypothesis, we investigate data on a sub-sample of training applicants, for whom a second follow-up survey (21 to 23 months after end of the training) was collected. Approximately half

of the applicants from the 2010 cohort were randomly selected to participate in this second follow-up survey. The survey team tracked and re-interviewed 634 individuals (79.5 percent) who were also interviewed at baseline. Using data from the follow up survey, we re-run the specifications from Table 4, Panel B and D, as well as the ones presented Tables 5 and 6 on medium-term outcomes. Tables 9 and 10 present the results.

[Table 9 about here]

We find that most of the main effect estimates, presented in Table 9's Panel A and B, are in the range of the estimates we found using the first follow-up data for the pooled sample. Strikingly, some of the estimates are slightly larger. In particular, we find statistically significant estimates for non-farm employment, trade-specific employment, and gainful employment rates of 40, 33, and 40 percentage points, respectively, for the compliers (LATE) as well as 16, 14, and 16 percentage points, respectively, for the individuals assigned to training based on the intent-to-treat approach (ITT). Because the 2010 data do not include information on whether non-farm employment is based on wage- or self-employment, we cannot make inferences on what drives the results in that dimension. However, when we investigate the medium-term disaggregated effects for the 2010 cohort, we find a slightly different pattern in comparison to the pattern based on the short-term impacts for the pooled sample. Although the coefficients for females are still positive, slightly lower in magnitude, and insignificant (possibly a result of reduced statistical power), we now find positive employment effects for men. Specifically, the non-farm and trade-specific employment rates for male compliers rise by 50 and 48 percentage points, respectively, while the medium-term male ITT effects show an increase of 19 and 18 percentage points, respectively.

[Table 10 about here]

Program impacts on time-use (reported in Table 10) show negative but insignificant coefficient estimates for female paid work outside the house and positive but insignificant coefficients for women who perform paid work inside the house. For men, the coefficient estimates of both paid work inside and outside the house are positive but statistically

29

insignificant. Although we do not detect statistically significant results, one could cautiously interpret them as somewhat suggestive evidence that medium-run employment rates for women, and to some extent for men, are also driven by self-employment activities.

EF Verification Data. For verification and monitoring purposes of program outcomes, the EF collected information on trainees' labor market outcomes. We retrieved information on these outcomes in aggregated form from four reports published by the EF from 2011 to 2013.[30] As another robustness check of our results, we compare these aggregated employment and earnings data from EF with the data from our surveys (reported in Table 11).

[Table 11 about here]

Based on the aggregated data, the share of trainees employed was 91 percent, six months after training completion, for the 2012 graduates. The share of trainees in gainful employment (*i.e.,* monthly income > 3,000 NRs) three and six months after completion of training was between 79 and 81 percent for the three cohorts. The employment rates, on the other hand, based on the self-reported data from our survey, range from 57 to 62 percent (9 to 11 months and 21 to 23 months after training completion). This comparison suggests that the numbers, based on the self-reported survey, are generally lower than those reported in the EF-verified data, which is consistent with the fact that providers receive their full payment after the six-month verification and that they have no additional incentives to support the job search of trainees. The rates for gainful employment are still relatively high among successful trainees at the time of our first and second follow-ups. Given that the employment rates at the three- and six-month mark are much higher, our estimates seem within a reasonable range.

In Figure 9, we also compare the earnings numbers based on the EF-verified data with the information from our first and second follow-up surveys. Given the growth trajectory of gainfully employed trainees' earnings over the first six months, post-training, the data based on the self-reported surveys seems reasonable and consistent with the overall pattern of the growth in earnings. Based both on this robustness check and also on the results from the 2010 cohort's

---

[30] Data is available from the Employment Fund Annual Reports 2010 to 2012 as well as a tracer study of 2011 EF graduates, retrieved from http://www.employmentfund.org.np/category/resources/reports/ on Feb 24, 2018.

second follow-up, we provide additional evidence consistent with the notion that our estimates are plausible.

[Figure 9 about here]

B. Differential Attrition

As in all panel studies, our estimates could be affected by a group-wise differential attrition. For instance, if trainees who are not admitted to the course choose to migrate for work abroad and if these are the most able and employable candidates of the control group, we would have an upward bias in our estimates. Therefore, in Table A4 of Appendix A, we explore the possibility of "*differential attrition*" and show no evidence to support it. Table A4 shows the results of a panel-based regression with attrition status as a dependent variable on a set of covariates for both the first follow-up (Columns 1 to 4) and the second-follow-up (Columns 5 to 8). All models are estimated based on the reconstructed score. In order to avoid the dropping of observations due to missing values in the control variables, Columns (3), (4), (7), and (8) include indicators that respectively flag missing values for each of the control variables. The regression results indicate that attrition is not correlated with assignment status in either wave.

C. Alternative Bandwidths

To investigate the robustness of our results relative to bandwidth choice, we re-estimate our main specifications using bandwidths within 2, 3, 4, 5 and 10 index scores of the threshold based on the reconstructed score. Tables A5 presents the results and it shows that overall our estimates are relatively stable both in statistical significance and in coefficient magnitude. For most outcomes, the magnitudes of the effect sizes increase with lower bandwidth choice. Only for non-farm self-employment the effect sizes are reduced by a half in the full sample. However, the effects for the female sub-sample are still large, positive, and significant using a bandwidth of 3 (ITT: 11, LATE: 36 percentage points, not shown in table) and, although still similar in size, they become insignificant at a bandwidth of 2 (ITT: 9, LATE: 32 percentage points, not shown in table).

D. Propensity Score Approach

As a robustness check for estimates of program impact, we also employ a combination of differences-in-differences and a propensity score matching technique (Meyer, 1995). This approach, in the context of training programs, has the potential to purge possible differences between observable characteristics for trainees and non-trainees following Dehejia and Wahba (2002). The results based on this approach are presented in Appendix B, and they essentially confirm the pattern of the analysis using the regression discontinuity setup in the main paper.

VIII. Discussion and Conclusion

Training interventions have been hailed as one potential solution to facilitate youth transition to productive employment and higher earnings worldwide. Although previous evaluations of training programs, based on observational designs, typically show positive and statistically significant impacts of training on the probability of having a job and on labor market earnings, recent experimental interventions from middle- and high-income countries find little or no impact on employment and modest gains in earnings. Using a regression discontinuity design in the context of a large vocational training program in Nepal, we find very large positive and statistically significant effects from the training program on female employment, hours worked, and earnings. These effects, in particular, are driven by women, who engage in non-farm self-employment activities carried out inside (but not outside) the house.

In line with the few other existing studies on similar programs in low-income countries, our estimates of the employment effects of this training intervention are among the largest for training programs around the world. Features of the low-income background, the South Asian context, and the specific training intervention likely account for the large program impacts that we find. First, our program impacts are likely driven by a lack of alternative employment, skill training opportunities, and by extremely low education levels. Both of these phoneme are much more pronounced in context of Nepal when compared with the context of the most recent experimental interventions from middle-income countries (i.e., in Latin America), especially so for women.

A second explanation behind the large program impacts relates to the extremely restrictive social norms regarding female labor force participation in Nepal. We find large program impacts on self-employment, especially inside the home, and these program effects are likely driven by

social norms that prevent women from otherwise being active in the labor market. Restrictive norms regarding women entering the labor force exist in many regions of the world, but they seem to be particularly strong and restrictive for women in South Asia (World Bank, 2011; Asian Development Bank, 2015; Booth, 2016). Largely influenced by the Hindu philosophy on women's positions in society, Nepalese women face gender-based employment barriers due to restricted mobility, cultural norms and societal norms as a result of gender expectations for certain occupations.

Finally, the EF training program was designed around employment outcomes such that training providers had to complete market assessments to ensure future employability with respect to the individually assigned trades. The training program was also bundled along with services, such as job placement, life-skill training and business training, all of which likely additionally contributed to its effectiveness. Perhaps, most importantly, the program was directly tailored to the needs of the target population and training providers were incentivized and closely monitored to accomplish their output.

Our results have important implications for the design and implementation of future training interventions in low-income countries. The empirical analysis presented here suggests important lessons for the successful modeling of effective labor market interventions where youth and female unemployment is a challenge.

References

Adoho, Franck, Shubha Chakravarty, Dala T. Korkoyah, Jr., Mattias Lundberg, and Afia Tasneem. 2014. "The Impact of an Adolescent Girls Employment Program: The EPAG project in Liberia." *World Bank Policy Research Working Paper*, 6832.

Almeida, Rita, Jere Behrman, and David Robalino. 2012. *The Right Skills for the Job? Rethinking Training Policies for Workers*. World Bank Publications.

Alzúa, María L., Guillermo Cruces, and Carolina Lopez. 2016. "Youth Training Programs Beyond Employment. Evidence from a Randomized Controlled Trial." *Economic Inquiry,* 54(4): 1839-1859.

Asian Development Bank. 2015. *Women in the Workforce: An Unmet Potential in Asia and Pacific*. Asian Development Bank.

Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2011. "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics*, 3(3): 188-220.

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." *Mimeo*.

Banerjee, Biswajit and J. B. Knight. 1985. "Caste discrimination in the Indian urban labour market." *Journal of Development Economics*, 17(3): 277-307.

Becker, Sascha O., Peter H. Egger, and Maximilian von Ehrlich. 2013. "Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects." *American Economic Journal: Economic Policy,* 5(4): 29–77.

Betcherman, Gordon, Martin Godfrey, Susana Puerto, Friederike Rother, and Antoneta Stavreska. 2007. *Global Inventory of Interventions to Support Young Workers. Synthesis Report.* The World Bank.

Blattman, Christopher and Jeannie Annan. 2016. "Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State." *American Political Science Review*, 110(1): 1-17.

Booth, Anne. 2016. "Women, Work and the Family: Is Southeast Asia Different?" *Economic History of Developing Regions*, 31(1): 167-197.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs." *Econometrica*, 82(6): 2295-2326.

Card, David, Jochen Kluve, and Andrea Weber. 2010. "Active Labour Market Policy Evaluations: A Meta-analysis." *The Economic Journal*, 120: F451-F477.

Card, David, Pablo Ibarrarán, Ferdinando Regalia, David Rosas, and Yuri Soares. 2011. "The Labor Market Impacts of Youth Training in the Dominican Republic." *Journal of Labor Economics*, 29(2): 267-300.

Cattaneo, Matias D., Luke Keele, Rocío Titiunik, Gonzalo Vazquez-Bare. 2016. "Interpreting Regression Discontinuity Designs with Multiple Cutoffs." *The Journal of Politics*, 78(4): 1229-1248.

Central Bureau of Statistics. 2009. *Report on the Nepal Labour Force Survey 2008*, Central Bureau of Statistics, National Planning Commission Secretariat, Government of Nepal.

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urquiola. 2005. "The Central Role of Noise in Evaluating Interventions that use Test Scores to Rank Schools." *American Economic Review*, 95(4): 1237-1258.

Cho, Yoonyoung , Davie Kalomba, Ahmed Mushfiq Mobarak, Victor Orozco, and Derek Wolfson. 2016. "Gender Differences In The Effects Of Vocational Training: Constraints On Women And Dropout Behavior." *Mimeo*

Currie, Janet and Jonathan Gruber. 1996a. "Health Insurance Eligibility, Utilization of Medical Care, and Child Health." *The Quarterly Journal of Economics*, 111(2): 431-466.

------. 1996b. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." *Journal of Political Economy*, 104(6): 1263-1296.

Cutler, David M. and Jonathan Gruber. 1996. "Does Public Insurance Crowd Out Private Insurance?" *The Quarterly Journal of Economics*, 111(2): 391-430.

Dehejia, Rajeev H. and Sadek Wahba. 2002. "Propensity Score-Matching Methods for Non-experimental Causal Studies." *The Review of Economics and Statistics*, 84(1): 151-161.

Fella, G. and G. Gallipoli. 2014. "Education and Crime Over the Life Cycle." *The Review of Economic Studies*, 81(4): 1484–1517.

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.

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.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design." *Econometrica*, 69(1): 201-209.

Heckman, James, Alan B. Krueger, and B. M. Friedman. 2004. *Inequality in America.* Cambridge.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score." *Econometrica*, 71(4): 1161-1189.

Hoxby, Caroline M. 2001. "All School Finance Equalizations are Not Created Equal." *The Quarterly Journal of Economics*, 116(4): 1189-1231.

Ibarraran, Pablo, Jochen Kluve, Laura Ripani, and David Rosas. 2015. "Experimental evidence on the long term impacts of a youth training program." *IDB Working Paper Series*, 657.

International Labour Organization (ILO). 2004. *The Informal Economy and Workers in Nepal.* International Labour Organization.

ILOSTAT. 2017. http://data.worldbank.org/indicator/SL.UEM.1524.ZS

Imbens, Guido and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics,* 142: 615–635.

Klasen, Stephan. 2002. "Low Schooling for Girls, Slower Growth for All? Cross-Country Evidence on the Effect of Gender Inequality in Education on Economic Development." *The World Bank Economic Review*, 16(3): 345–373.

Kluve, Jochen. 2010. "The effectiveness of European active labor market programs." *Labour Economics*, 17: 904-918.

Knowles, Stephen, Paula K. Lorgelly, P. Dorian Owen. 2002. "Are educational gender gaps a brake on economic development? Some cross-country empirical evidence." *Oxford Economic Papers*, 54(1): 118–149.

Lechner, Michael, Ruth Miquel, Conny Wunsch. 2011. "Long-run effects of public sector sponsored training in West Germany." *Journal of the European Economic Association,* 9(4): 742–784.

Lee, David and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature, 48: 281–355.*

Lenze, Jana and Stephan Klasen. 2016. "Does Women's Labor Force Participation Reduce Domestic Violence? Evidence from Jordan." *Feminist Economics*, 23(1): 1-29.

Linn, Margaret, Richard Sandifer, and Shayna Stein. 1985. "Effects of Unemployment on Mental and Physical Health." *American Journal of Public Health*, 75(5): 502-506.

Lokshin, Michael and Elena Glinskaya. 2009. "The Effect of Male Migration on Employment Patterns of Women in Nepal." *The World Bank Economic Review*, 23(3): 481-507.

Ludwig, Jens and Douglas Miller. 2007. "Does head start improve children's life chances?

Evidence from a regression discontinuity design." *Quarterly Journal of Economics,* 122(1): 159–208.

Mainali, Ram, Saqib Jafarey and Gabriel Montes-Rojas. 2017. "Earnings and Caste: An Evaluation of Caste Wage Differentials in the Nepalese Labour Market." *The Journal of Development Studie*s, 53(3): 396-421.

Maitra, Pushkar and Subha Mani. 2017. "Learning and Earning: Evidence from a Randomized Evaluation in India." *Labour Economics*, 45: 116-130.

Majlesi, Kaveh. 2016. "Labor market opportunities and women's decision making power within households." *Journal of Development Economics*, 119: 34-47.

Malamud, Ofer and Cristian Pop-Eleches. 2010. "General Education Versus Vocational Training: Evidence from an Economy in Transition." *The Review of Economics and Statistics*, 92(1): 43-60.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142(2): 698-714.

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

Miller, Grant, Diana Pinto, and Marcos Vera-Hernandez. 2013. "Risk Protection, Service use, and Health Outcomes Under Colombia's Health Insurance Program for the Poor." *American Economic Journal: Applied Economics*, 5(4): 61-91.

Nayar, Reema. 2011. *More and Better Jobs in South Asia*. World Bank Publications.

Reis, Mauricio. 2015. "Vocational Training and Labor Market Outcomes in Brazil." *The BE Journal of Economic Analysis & Policy*, 15(1): 377-405.

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

Suwal, Rudra and Bishnu Pant. 2009. "Measuring Informal Sector Economic Activities in Nepal." Paper Prepared for the Special IARIW-SAIM Conference on *Measuring the Informal Economy in Developing Countries*, Kathmandu, Nepal.

The World Bank. 2011. *World Development Report 2012: Gender Equality and Development.* The World Bank.

------. 2013. *World Development Report 2013: Jobs.* The World Bank.

UNESCO. 2007. *The state of education in Latin America and the Caribbean: Guaranteeing quality education for all.* OREALC/UNESCO Santiago.

United Nations. 2012. *World Population Monitoring: Adolescents and Youth.* United Nations.

Figures

FIGURE 1: TIMELINE

	2010				2011			2012			2013
	Jan-Apr				Jan-Apr			Jan-Apr			Jan-Apr
2010 Cohort	Baseline Survey /Start of Training	1st EF Verif.	2nd EF Verif.		1st Follow-up Survey			2nd Follow-up Survey*			
2011 Cohort					Baseline Survey /Start of Training	1st EF Verif	2nd EF Verif.	1st Follow-up Survey			
2012 Cohort								Baseline Survey /Start of Training	1st EF Verif.	2nd EF Verif.	1st Follow-up Survey

Notes: Baseline interviews were usually conducted a few days or weeks before start of training. Follow-up interviews were conducted approximately 12 and 24 months after baseline interviews, respectively. Training duration lasts from 1 to 3 months. EF verifies employment status and earnings of applicants 3 and 6 months after end of training, i.e. approx. 8 to 6 and 5 to 3 months before the first follow-up survey. *Roughly half of the initial sample of the 2010 cohort was randomly selected for the second follow-up survey.

FIGURE 2: DISTRICTS COVERED BY COHORT
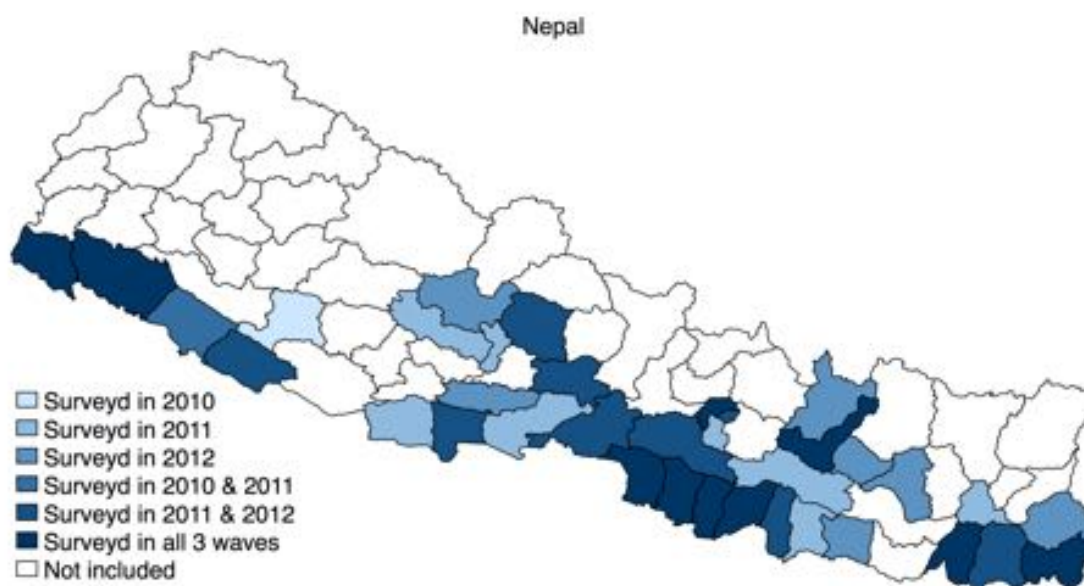
FIGURE 3: PROBABILITY OF PROGRAM PARTICIPATION
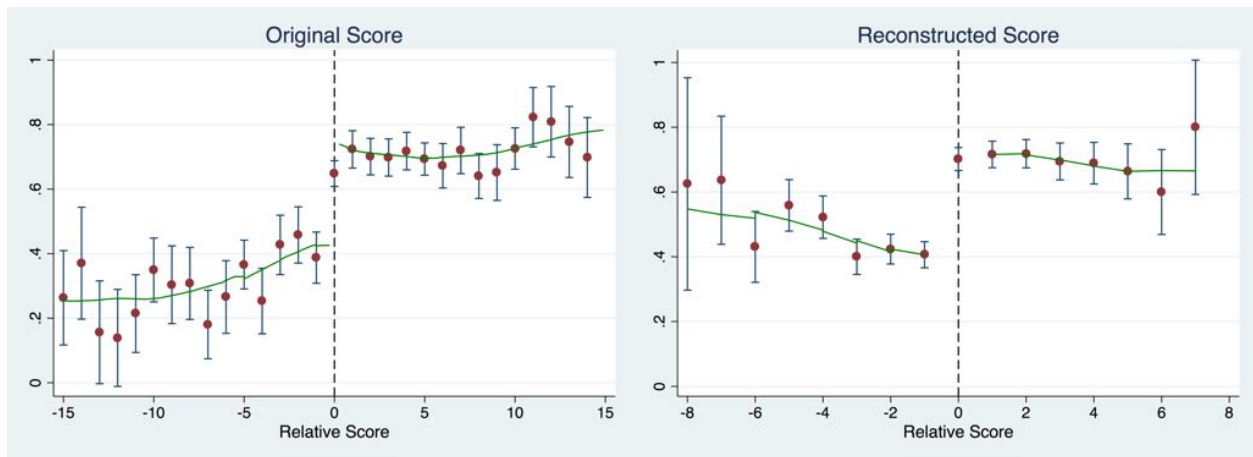
FIGURE 4: DENSITY OF FORCING VARIABLE
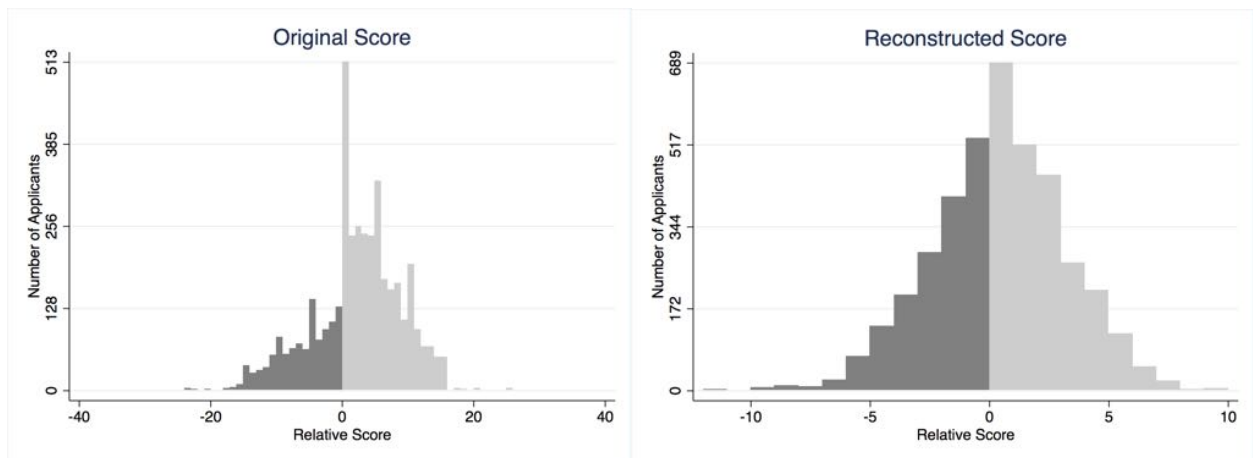
FIGURE 5: CONTINUITY OF COVARIATES AROUND THE CUT-OFF AT BASELINE
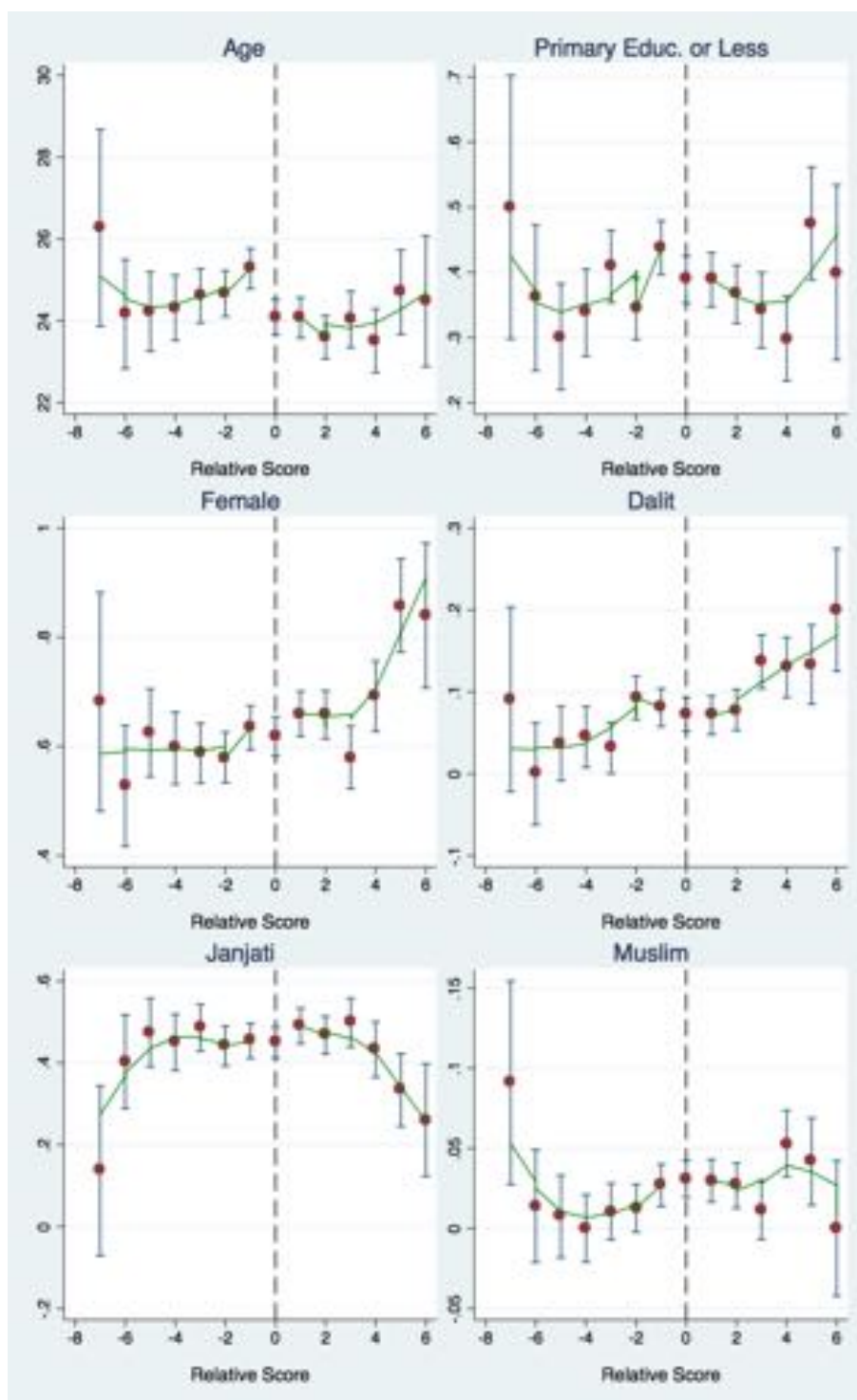
FIGURE 6: CONTINUITY OF OUTCOMES AROUND THE CUT-OFF AT BASELINE
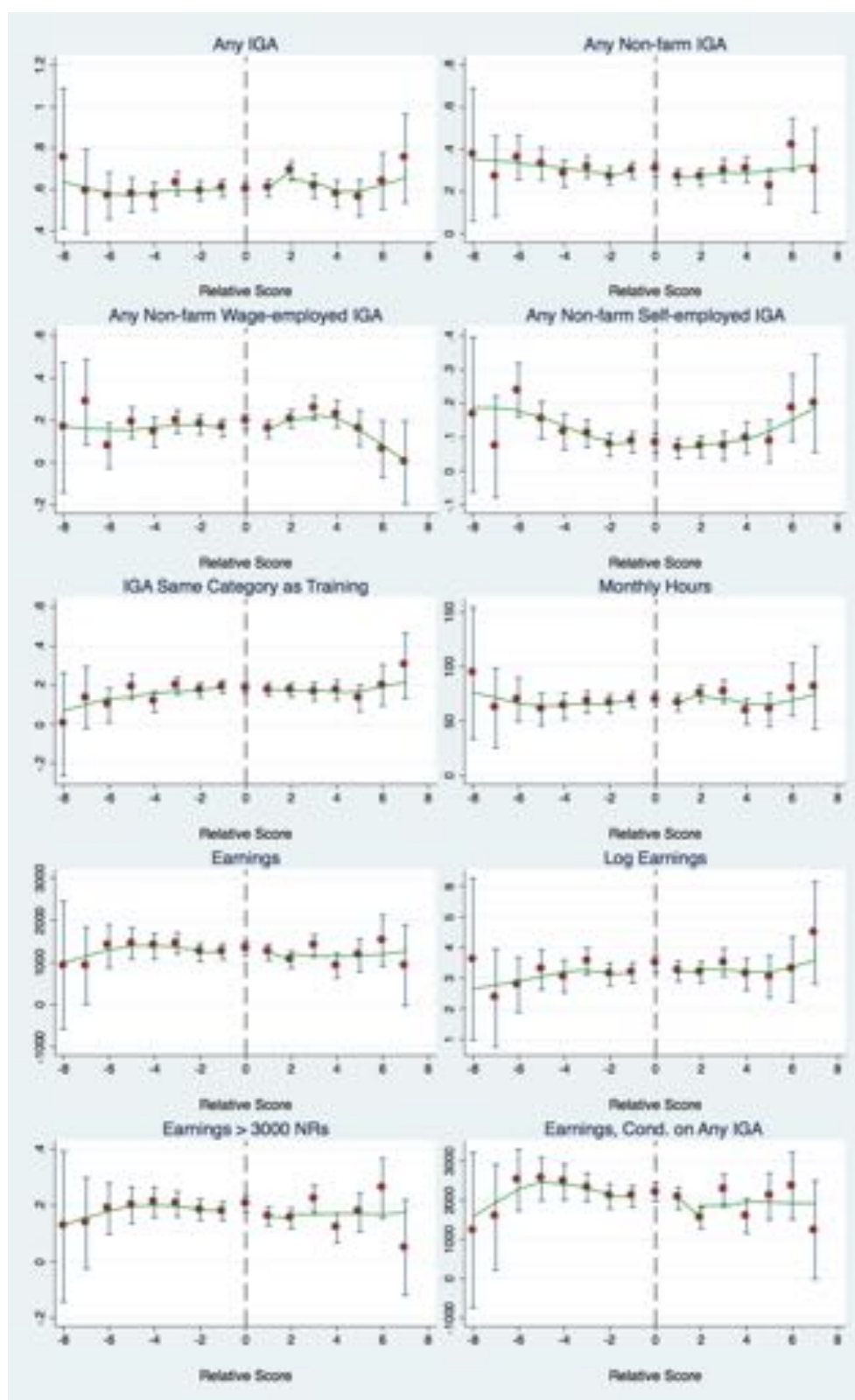
FIGURE 7: CONTINUITY OF TRADE CHOICES AROUND THE CUT-OFF
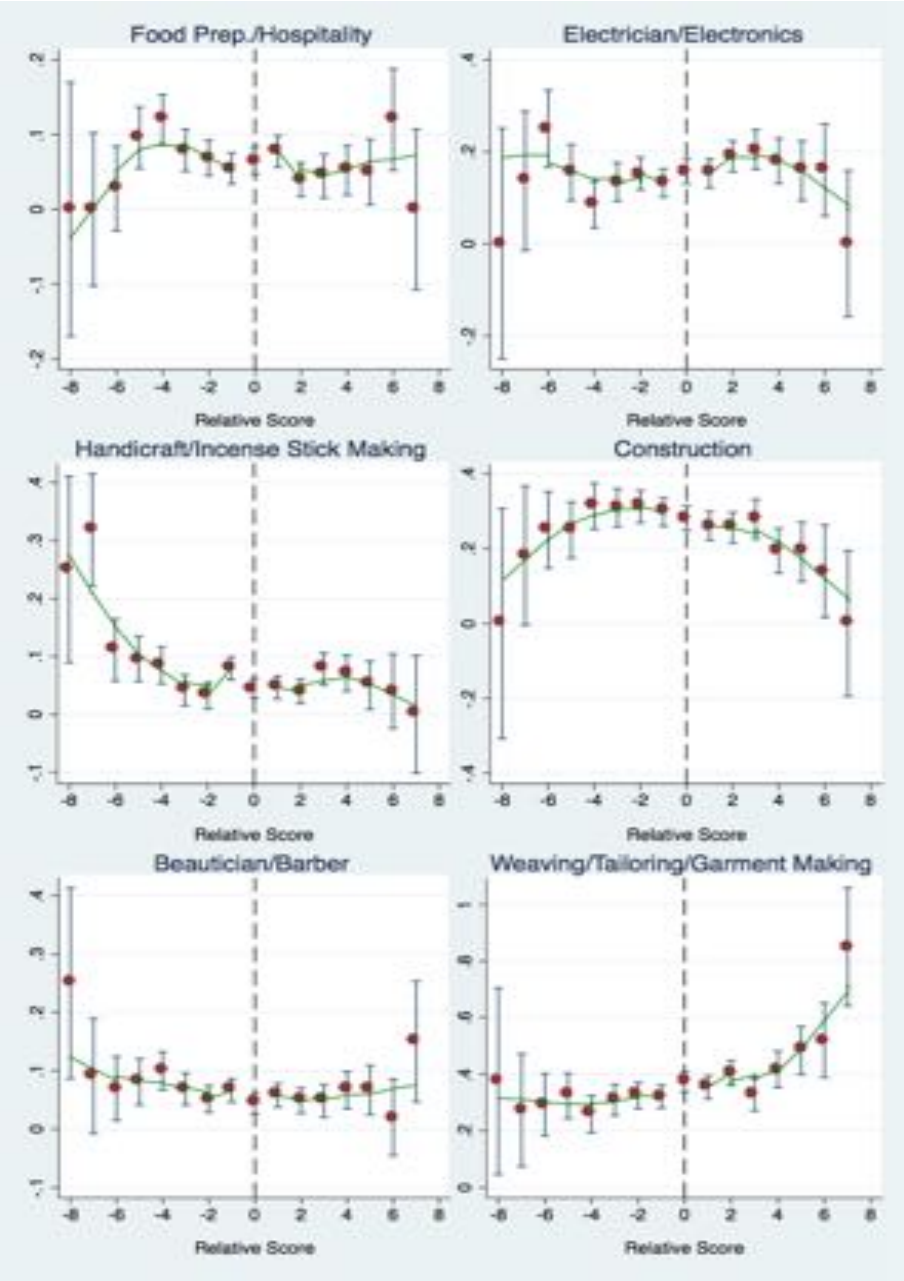
FIGURE 8: GRAPHICAL REPRESENTATION OF INTENT-TO-TREAT RESULTS
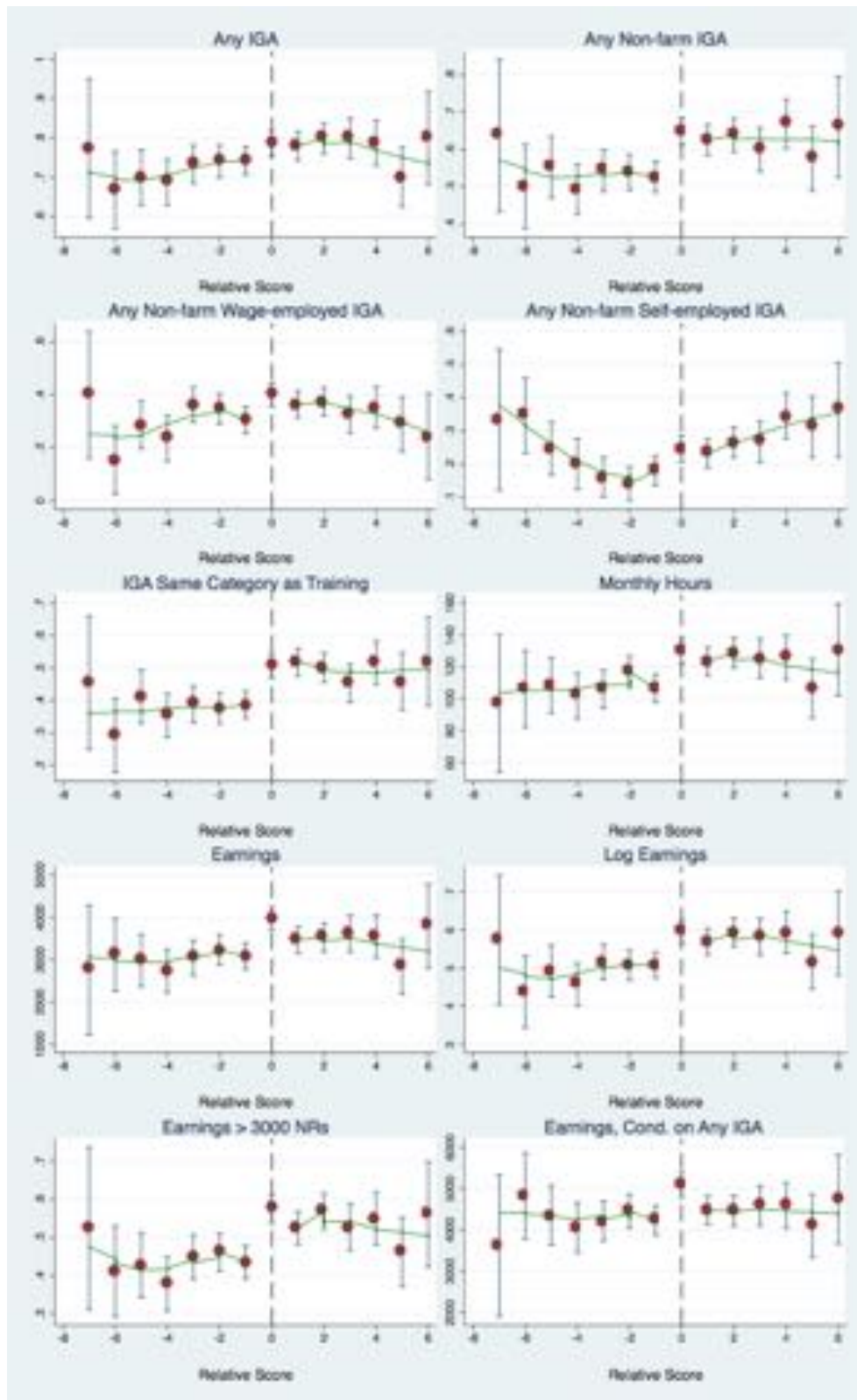
FIGURE 9: EARNINGS BASED ON SURVEY AND ADMINISTRATIVE DATA COMPARISION

Note: Solid lines are based on EF verification data, retrieved from EF Annual Report 2012. Dashed lines are based on self-reported survey data approx. 9 to 11 and 21 to 23 month after training. Numbers represent average earnings of those trainees who are gainfully employed at a given month.
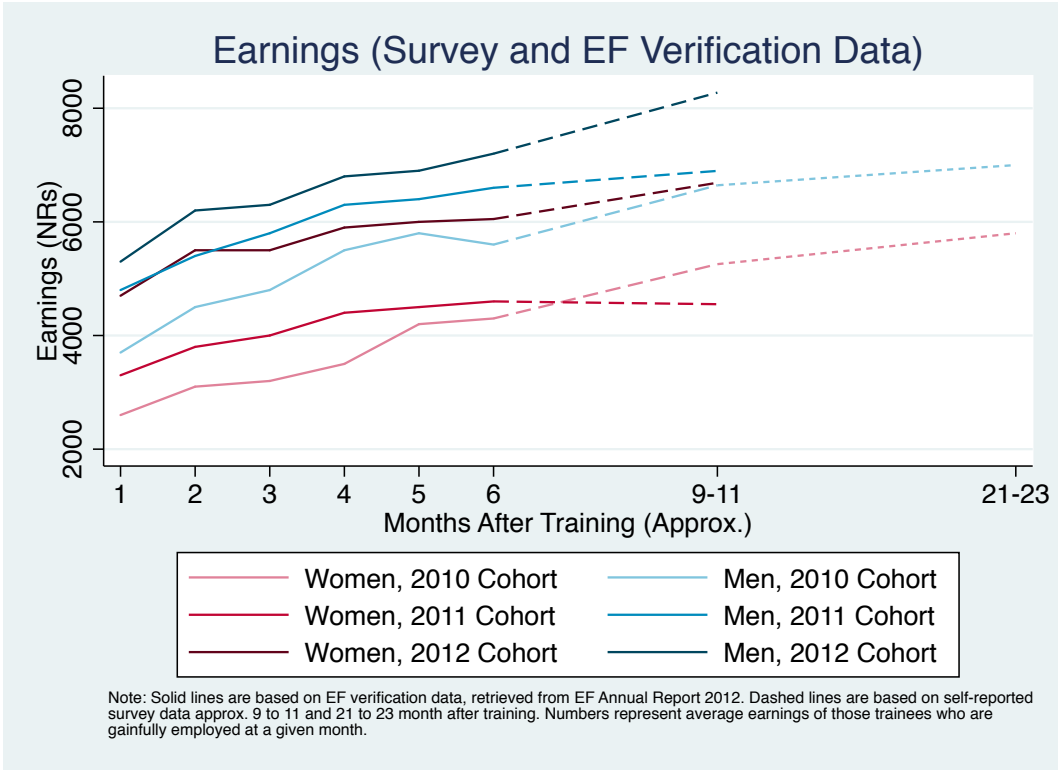
Tables

TABLE 1: TOTAL AND SAMPLED EF EVENTS

	2010	2011	2012	Pooled
All EF events				
# Events conducted by EF (all year)	598	645	711	1954
# Events conducted by EF (Jan-Apr)	110	142	143	395
# Training providers	21	32	35	-
# Trainees	11750	12869	14255	38874
Sampled EF events				
# Events randomly sampled for baseline survey	N/A	182	112	-
# Events included in baseline survey	65	69	85	219
# Districts covered	12	28	26	-
# Training 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 2: SURVEY RESPONSE RATES

	Baseline	Follow-up	Follow-up rate
2010 cohort			
Above Threshold	1184	1047	88.4%
Below Threshold	372	330	88.7%
Total	1556	1377	88.5%
2011 cohort			
Above Threshold	1237	1113	90.0%
Below Threshold	349	306	87.7%
Total	1586	1419	89.4%
2012 cohort			
Above Threshold	1044	888	85.1%
Below Threshold	491	417	84.9%
Total	1535	1305	85.0%
Pooled cohorts			
Above Threshold	3465	3049	88.0%
Below Threshold	1212	1053	86.9%
Total	4677	4101	87.7%

*Notes***:** The second follow-up survey for the 2010 cohort was conducted on a reduced (randomly selected) sample of 800 individuals. Out of those, the survey team was able to interview 634 individuals (79.5 percent) who were also interviewed at baseline.

Estimates by Absolute Threshold Score
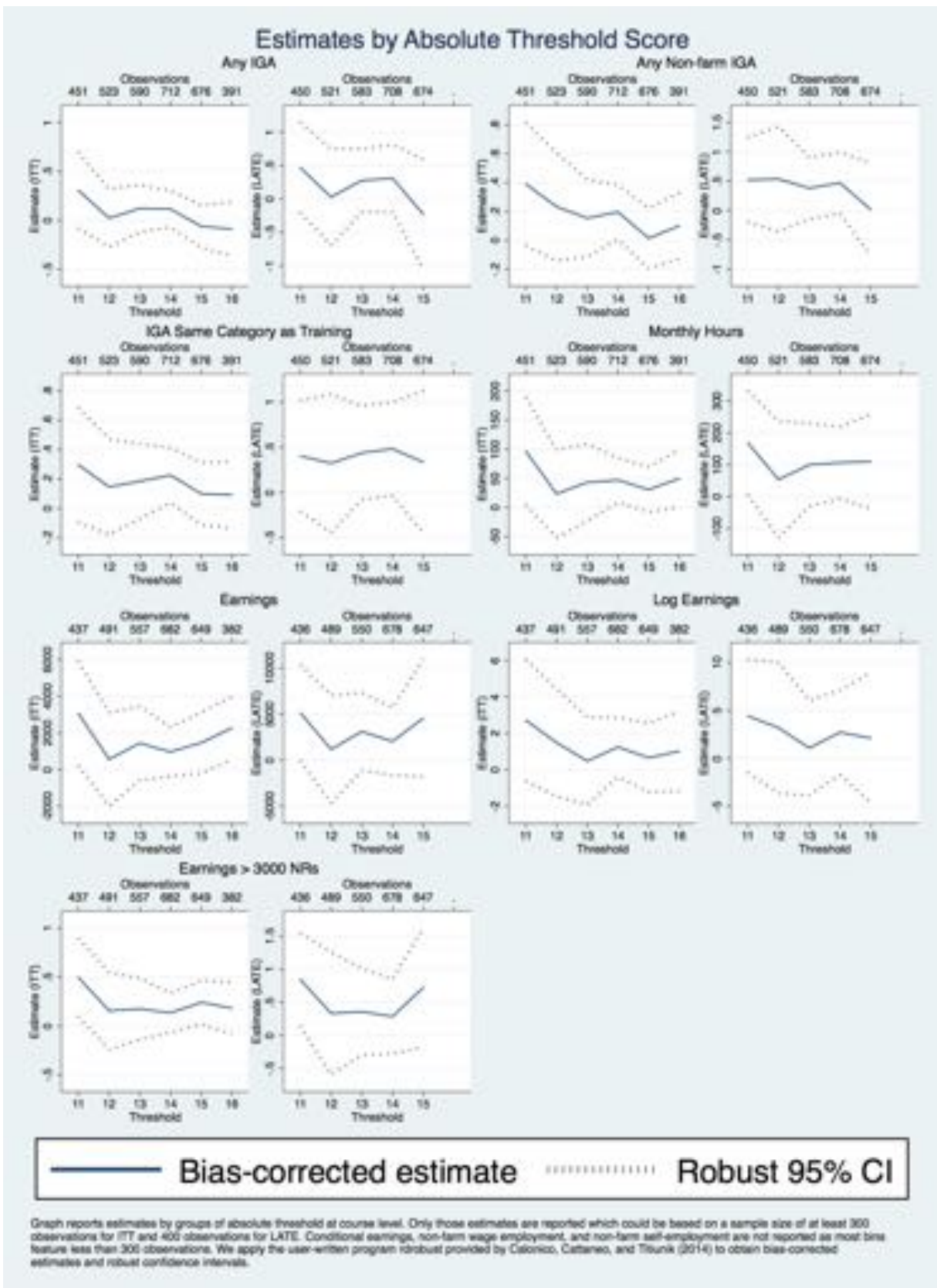
FIGURE A4: CROSS-VALIDATION FUNCTION FOR OUTCOMES IN LEVELS
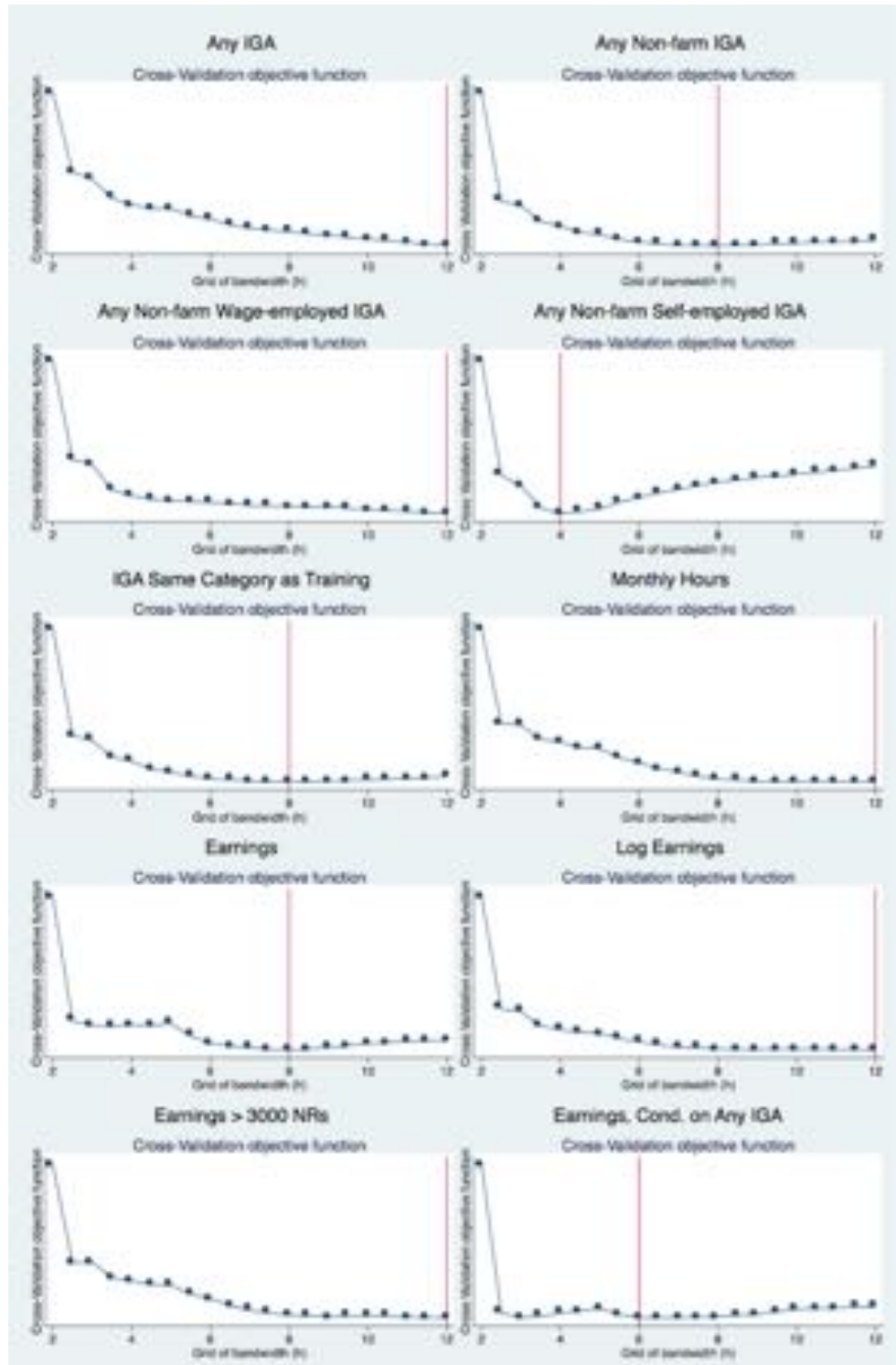
FIGURE A5: CROSS-VALIDATION FUNCTION FOR DIFFERENCED OUTCOMES
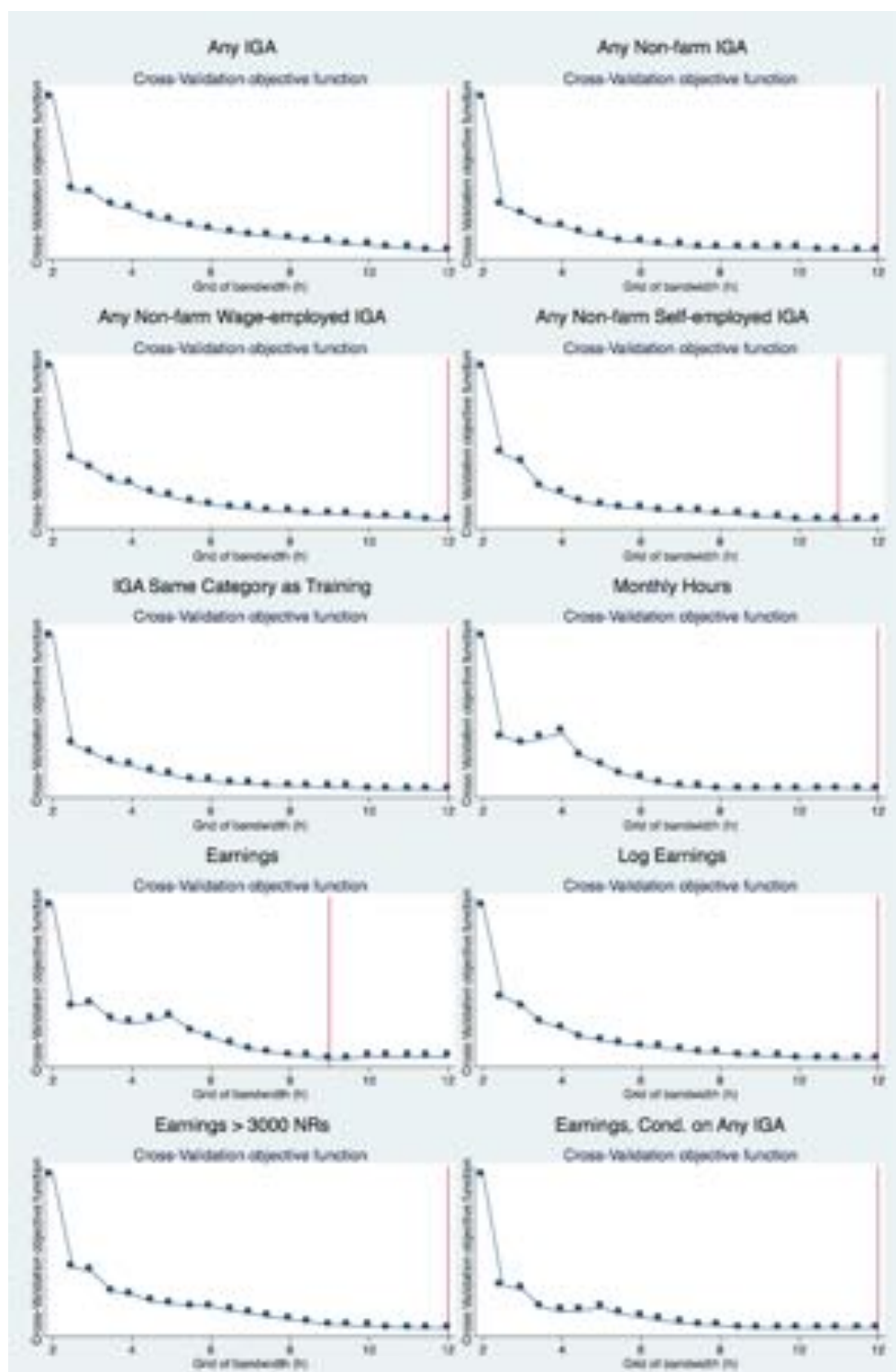
TABLE A1: ORIGINAL AND RECONSTRUCTED SCORE COMPONENTS

Sub-Score	Total weight	Basis for evaluation	Indicators	Original Score	Reconstructed Score	
				Assigned Sub-Score	Assigned Sub-Score	Source
1	15%	Trade-specific education requirement	Compulsory prerequisite: All candidates must meet the minimum requirement for the training they applied to*	15 (0, 5 or 10 in exceptional cases)	Integers of 0 to 3	Predicted
2	20%	Economic poverty	Less than 3 months of food sufficiency	20	4	Application form data
			Less than 6 months of food sufficiency or less than 3000 per capita family income from non-farm based income	15	3	
			More than 6 months of food sufficiency and per capita family income from non-farm based income equal or more than 3000	0	0	
3	25%	Social caste	Women: Dalit women or women from the following special groups: widows; internally displaced; ex-combatants; physically disabled; HIV-infected infected	25	5	Application form data
			Women: Economically poor women not referred to above	20	4	
			Men: Dalit, Janjati, Madhesi men or men from the following special groups: internally displaced; ex-combatants; physically disabled; HIV-infected infected	15	3	
			Men: Economically poor men not referred to above	10	2	
			Neither of the above	0	0	
4	10%	Development status of district of origin	Least developed districts	10	2	Application form data
			Moderately developed districts	5	1	
			Developed districts	0	0	
Preliminary score for short-listing (Sub – total)				**70**	**14**	
5	30%	Interview	Commitment, Motivation, Attitude, Aptitude, Clear Vision for Employment and Enterprising	0-30	0-6	Predicted
Total score after interview				**100**	**20**	

Notes: *If candidates did not fulfill the course specific education prerequisites they were not eligible for short-listing and immediately rejected. Yet, in exceptional cases (approx. 9 percent of the sample) this criterion was not adhered to and instead applicants received 0, 5, or 10 points. When reconstructing this component, we therefore allow for integer values between 0 and 3.

TABLE A2: DEVELEOPMENT STATUS OF NEPAL DISTRICTS

Developed District	Moderately Developed District	Least Developed District
Kathmandu	Makawanpur	Ramechhap
Chitwan	Gulmi	Parsa
Jhapa	Surkhet	Rasuwa
Bhaktapur	Solukhumbu	Kapilbastu
Lalitpur	Banke	Bara
Kaski	Bhojpur	Dadeldhura
Dhankuta	Gorkha	Darchula
Palpa	Taplejung	Siraha
Syangja	Bardiya	Jajarkot
Manang	Kanchanpur	Rukum
Morang	Nuwakot	Sarlahi
Illam	Nawalparasi	Baitadi
Rupandehi	Khotang	Dailekh
Sunsari	Okhaldhunga	Rolpa
Kabhreplanchok	Kailali	Mahotari
Tanahu	Dolakha	Doti
Terhathum	Arghakhanchi	Dolpa
Sankhuwasabha	Udayapur	Rautahat
Mustang	Dhading	Jumla
Parbat	Salyan	Kalikot
Dang	Dhanusa	Bajura
Lamjung	Saptari	Achham
Panchthar	Sindhipalchok	Bajhang
Baglung	Sundhuli	Humla
Myagdi	Pyuthan	Mugu

Source: Districts of Nepal, Indicators of Development. Updated 2003. CBS/Nepal, ICIMOD. December, 2003

TABLE A3: CORRELATION OF 5TH SUB-SCORE WITH OTHER SCORE COMPONENTS

	Sub-Score 5 (1)	Sub-Score 5 (2)	Sub-Score 5 (3)
Aggregated Sub-Scores 1 to 4	-0.220*** (0.021)		
Sub-Score 1 (Trade-specif. Edu.)		0.117* (0.062)	2.752*** (0.431)
Sub-Score 2 (Econ. Poverty)		-0.182*** (0.029)	1.046* (0.551)
Sub-Score 3 (Social Caste)		-0.348*** (0.031)	1.013*** (0.350)
Sub-Score 4 (Development Status of Dist. Of Origin)		-0.473*** (0.095)	3.560*** (1.354)
SS 1 x SS 2			-0.120*** (0.039)
SS 1 x SS 3			-0.136*** (0.027)
SS 1 x SS 4			-0.379*** (0.101)
SS 2 x SS 3			-0.049 (0.030)
SS 2 x SS 4			-0.194** (0.096)
SS 3 x SS 4			-0.278*** (0.087)
SS 1 x SS 2 x SS 3			0.006*** (0.002)
SS 1 x SS 2 x SS 4			0.019*** (0.007)
SS 1 x SS 3 x SS 4			0.026*** (0.006)
SS 2 x SS 3 x SS 4			0.013** (0.006)
SS 1 x SS 2 x SS 3 x SS 4			-0.001*** (0.000)
N	4090	4090	4090
Adj. R2	0.38	0.39	0.40

Notes: All models include event dummies. ***, **, and * denote significant at the 1, 5, and 10 percent level, respectively.

TABLE A4: ATTRITION

	Pooled Cohorts, 1st Follow-Up				2010 Cohort, 2nd Follow-Up			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
"Above Threshold"	0.008	0.017	0.007	0.016	-0.011	-0.047	-0.004	-0.052
	(0.010)	(0.019)	(0.010)	(0.019)	(0.027)	(0.045)	(0.029)	(0.047)
Female X "Above Threshold"		-0.020		-0.014		0.059		0.080
		(0.021)		(0.021)		(0.066)		(0.066)
Female		0.077***	0.066***	0.074***		-0.026	0.026	-0.019
		(0.021)	(0.014)	(0.020)		(0.060)	(0.035)	(0.056)
Age			0.004***	0.004***			0.007**	0.007**
			(0.001)	(0.001)			(0.003)	(0.003)
Parent			-0.003	-0.003			-0.006	-0.009
			(0.019)	(0.019)			(0.045)	(0.045)
Married			0.011	0.011			0.006	0.010
			(0.019)	(0.019)			(0.043)	(0.043)
Dalit			-0.052**	-0.052**			-0.052	-0.053
			(0.023)	(0.023)			(0.062)	(0.062)
Janjati			-0.014	-0.014			0.109***	0.110***
			(0.012)	(0.012)			(0.032)	(0.032)
Any IGA at baseline			0.017	0.017			0.024	0.023
			(0.011)	(0.011)			(0.027)	(0.027)
N	4585	4585	4585	4585	1547	1547	1547	1547
Training provider dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All models include district and training provider dummies. In Columns (3), (4), (7), and (8) indicators are included which flag missing values for each of the control variables, respectively. All models are estimated based on the reconstructed score. The difference in sample size between the initial baseline sample and the sample we use in this analysis arises due to missing values in the variables that were necessary to reconstruct the score variable, which determines assignment. All standard errors are clustered at event level. ***, **, and * denote significant at the 1, 5, and 10 percent level, respectively.

TABLE A5: BANDWIDTH SENSITIVITY

	Any IGA (1=Yes) (1)	Any non-farm IGA (1=Yes) (2)	Non-farm wage IGA (1=Yes) (3)	Non-farm self IGA (1=Yes) (4)	Trade-specific IGA (1=Yes) (5)	Hours worked in past month (6)	Earnings (7)	Logged earnings (8)	Earnings > 3000 NRs. (1=Yes) (9)	Earnings, conditional on any IGA (10)
Panel A: LATE, DD (Reconstructed Score)										
2 Index Scores	0.19	0.50**	0.35	0.14	0.43**	121.8***	3387**	2.33	0.43**	3058
	(0.21)	(0.20)	(0.31)	(0.27)	(0.20)	(42.98)	(1341)	(1.63)	(0.21)	(2254)
Observations	2133	2133	1364	1364	2133	2133	2049	2049	2049	1012
F-statistic	27.50	27.50	12.22	12.22	27.50	27.50	27.83	27.83	27.83	12.85
3 Index Scores	0.13	0.38**	0.28	0.18	0.45***	77.66**	2319**	1.52	0.34**	1722
	(0.16)	(0.15)	(0.19)	(0.17)	(0.16)	(31.35)	(1083)	(1.27)	(0.16)	(1645)
Observations	2874	2874	1862	1862	2874	2874	2759	2759	2759	1411
F-statistic	36.18	36.18	20.31	20.31	36.18	36.18	37.95	37.95	37.95	17.43
4 Index Scores	0.09	0.31**	0.15	0.25*	0.45***	58.86**	1755**	1.26	0.24**	1258
	(0.12)	(0.12)	(0.13)	(0.14)	(0.13)	(24.97)	(872)	(1.06)	(0.12)	(1097)
Observations	3340	3340	2145	2145	3340	3340	3202	3202	3202	1622
F-statistic	62.42	62.42	33.90	33.90	62.42	62.42	65.38	65.38	65.38	37.29
5 Index Scores	0.07	0.29***	0.10	0.25**	0.41***	52.67**	1666**	1.29	0.24**	1393
	(0.11)	(0.10)	(0.12)	(0.12)	(0.12)	(20.54)	(744)	(0.92)	(0.10)	(994)
Observations	3685	3685	2391	2391	3685	3685	3530	3530	3530	1778
F-statistic	77.89	77.89	42.88	42.88	77.89	77.89	79.88	79.88	79.88	42.02
10 Index Scores	0.11	0.29***	0.09	0.28**	0.41***	45.13**	1685**	1.53*	0.28***	1662
	(0.10)	(0.10)	(0.11)	(0.12)	(0.12)	(20.99)	(705)	(0.89)	(0.11)	(1017)
Observations	4000	4000	2604	2604	4000	4000	3834	3834	3834	1929
F-statistic	83.98	83.98	51.84	51.84	83.98	83.98	82.98	82.98	82.98	38.61
Panel B: ITT, DD (Reconstructed Score)										
2 Index Scores	0.06	0.15**	0.08	0.03	0.13**	36.23***	1033***	0.71	0.13**	885
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(12.15)	(392)	(0.51)	(0.06)	(602)
3 Index Scores	0.03	0.11**	0.07	0.05	0.13***	21.29**	662**	0.42	0.10**	455
	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(8.66)	(309)	(0.38)	(0.04)	(437)
4 Index Scores	0.03	0.10**	0.05	0.07*	0.15***	18.97**	587**	0.42	0.08**	435
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(8.26)	(295)	(0.37)	(0.04)	(386)
5 Index Scores	0.02	0.10***	0.03	0.08**	0.14***	17.88**	579**	0.45	0.09**	473
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(7.20)	(264)	(0.33)	(0.04)	(344)
10 Index Scores	0.03	0.09***	0.03	0.08**	0.13***	14.11**	532**	0.49*	0.09***	494
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(6.60)	(223)	(0.29)	(0.03)	(305)

Notes: Standard errors adjusted for clustering at event level reported in parentheses. Panel A presents second stage results. In both panels, each cell represents an estimate from a separate regression, which includes the total score, the relative score (forcing variable), and an interaction of the relative score with the assignment variable as counterfactuals, as well as a constant. *Non-farm wage IGA* and *Non-farm self IGA* are only available for 2011 and 2012 cohorts. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 3: BALANCE OF COVARIATES AND OUTCOME VARIABLES AT BASELINE (FULL SAMPLE)

Variable	Baseline Means							Differences (Original Score)				Differences (Reconstructed Score)			
	All	Women	Men	Non-treated	Treated	Non-assigned	Assigned	LATE		ITT		LATE		ITT	
Demographics															
Age	24.32	24.66	23.72	24.39	24.27	24.34	24.31	-1.21	(1.53)	-0.32	(0.42)	-4.37***	(1.35)	-1.39***	(0.41)
Primary Education or Less (1=Yes)	0.38	0.43	0.29	0.39	0.37	0.35	0.39	-0.16	(0.12)	-0.04	(0.03)	-0.12	(0.09)	-0.04	(0.03)
Men (1=Yes)	0.37	0	1	0.37	0.37	0.36	0.37	0.38***	(0.14)	0.10***	(0.04)	0.00	(0.10)	0.00	(0.03)
Dalit (1=Yes)	0.08	0.08	0.08	0.10	0.07	0.09	0.08	-0.38***	(0.08)	-0.10***	(0.02)	-0.11**	(0.06)	-0.03*	(0.02)
Janjati (1=Yes)	0.46	0.42	0.51	0.41	0.49	0.42	0.47	0.09	(0.14)	0.02	(0.04)	0.04	(0.11)	0.02	(0.03)
Muslim (1=Yes)	0.03	0.03	0.02	0.02	0.03	0.02	0.03	0.01	(0.04)	0.00	(0.01)	0.04	(0.04)	0.01	(0.01)
Outcome Variables															
Any IGA (1=Yes)	0.61	0.52	0.77	0.59	0.63	0.59	0.62	0.11	(0.12)	0.03	(0.03)	-0.04	(0.10)	-0.01	(0.03)
Any non-farm IGA (1=Yes)	0.30	0.19	0.47	0.28	0.31	0.27	0.31	0.20*	(0.10)	0.05*	(0.03)	-0.03	(0.09)	-0.01	(0.03)
Non-farm wage IGA (1=Yes)	0.19	0.08	0.39	0.18	0.20	0.16	0.20	0.25**	(0.12)	0.07**	(0.03)	0.07	(0.09)	0.02	(0.03)
Non-farm self IGA (1=Yes)	0.09	0.10	0.09	0.09	0.09	0.09	0.10	0.07	(0.09)	0.02	(0.02)	-0.02	(0.07)	-0.01	(0.02)
Trade-specific IGA (1=Yes)	0.18	0.11	0.29	0.17	0.18	0.15	0.19	0.11	(0.09)	0.03	(0.03)	-0.11	(0.08)	-0.04	(0.02)
Hours worked in past month	69.26	47	108	66	71	63	72	40**	(19)	11**	(5)	-1	(18)	-0	(6)
Earnings	1269	770	2138	1254	1280	1180	1303	851*	(512)	221	(140)	-44	(407)	-13	(128)
Logged earnings	3.29	2.43	4.80	3.15	3.39	3.03	3.39	1.37	(0.96)	0.36	(0.26)	0.15	(0.73)	0.05	(0.23)
Earnings > 3000 NRs. (1=Yes)	0.18	0.10	0.34	0.18	0.19	0.16	0.19	0.16*	(0.09)	0.04*	(0.02)	0.01	(0.07)	0.00	(0.02)
Earnings, conditional on any IGA	2082	1491	2774	2127	2054	2009	2109	943	(735)	256	(209)	201	(608)	57	(190)
Hours/week working ...															
Unpaid, inside house > 5 (1=Yes)	0.82	0.94	0.61	0.81	0.82	0.79	0.83	0.04	(0.10)	0.01	(0.03)	-0.02	(0.06)	-0.01	(0.02)
Unpaid, inside house > 10 (1=Yes)	0.63	0.82	0.31	0.62	0.64	0.60	0.65	-0.00	(0.13)	-0.00	(0.04)	-0.10	(0.09)	-0.03	(0.03)
Unpaid, inside house > 20 (1=Yes)	0.39	0.55	0.12	0.39	0.40	0.38	0.4	-0.13	(0.11)	-0.03	(0.03)	-0.09	(0.09)	-0.03	(0.03)
Unpaid, outside house > 0 (1=Yes)	0.61	0.56	0.69	0.61	0.61	0.60	0.61	0.18	(0.12)	0.05*	(0.03)	0.07	(0.09)	0.02	(0.03)
Paid, inside house > 0 (1=Yes)	0.52	0.47	0.59	0.52	0.52	0.51	0.52	0.04	(0.12)	0.01	(0.03)	-0.15	(0.10)	-0.05	(0.03)
Paid, outside house > 0 (1=Yes)	0.48	0.36	0.69	0.45	0.49	0.45	0.49	0.04	(0.12)	0.01	(0.03)	0.00	(0.10)	0.00	(0.03)

Notes: *Non-farm wage IGA* and *Non-farm self IGA* are only available for 2011 and 2012 cohorts.

TABLE 4: EMPLOYMENT

	Any IGA (1=Yes) (1)	Any non-farm IGA (1=Yes) (2)	Non-farm wage IGA (1=Yes) (3)	Non-farm self IGA (1=Yes) (4)	Trade-specific IGA (1=Yes) (5)	Hours worked in past month (6)	Earnings (7)	Logged earnings (8)	Earnings > 3000 NRs. (1=Yes) (9)	Earnings, conditional on any IGA (10)
Panel A										
LATE, DD	0.27*	0.52***	0.38***	0.22*	0.61***	75.92***	3074***	2.85*	0.43***	3179**
(Original Score)	(0.15)	(0.15)	(0.15)	(0.12)	(0.14)	(25.57)	(987)	(1.55)	(0.14)	(1295)
First Stage	0.27***	0.26***	0.29***	0.29***	0.27***	0.27***	0.26***	0.23***	0.26***	0.28***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
F-statistic	67.30	54.90	47.53	47.53	58.42	67.30	53.73	36.04	52.31	32.37
Baseline mean	0.61	0.30	0.19	0.09	0.18	69.26	1271	3.27	0.18	2084
Bandwidth	37	15	37	37	19	37	23	12	19	26
Observations	4080	3976	2677	2677	4066	4080	3899	3544	3891	1979
Panel B										
LATE, DD	0.13	0.31***	0.11	0.30**	0.40***	48.85**	1754**	1.75**	0.31***	2025*
(Reconstr. Sc.)	(0.10)	(0.10)	(0.11)	(0.12)	(0.12)	(20.50)	(696)	(0.87)	(0.11)	(1115)
First Stage	0.32***	0.32***	0.31***	0.30***	0.32***	0.32***	0.33***	0.32***	0.32***	0.30***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.05)
F-statistic	87.47	87.47	54.84	51.12	87.47	87.47	85.92	86.33	86.33	38.23
Baseline mean	0.61	0.29	0.18	0.09	0.18	68.62	1260	3.27	0.18	2075
Bandwidth	12	12	12	11	12	12	9	12	12	12
Observations	4004	4004	2608	2606	4004	4004	3822	3838	3838	1931
Panel C										
LATE, Level	0.09	0.33***	0.17	0.26**	0.33***	48.08**	1934***	1.80**	0.32***	2686**
(Reconstr. Sc.)	(0.08)	(0.09)	(0.11)	(0.12)	(0.10)	(21.01)	(745)	(0.81)	(0.10)	(1355)
First Stage	0.32***	0.34***	0.31***	0.31***	0.34***	0.32***	0.34***	0.32***	0.32***	0.29***
	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.05)
F-statistic	87.47	88.14	57.61	35.82	88.14	87.47	89.99	88.26	88.26	29.78
Baseline mean	0.61	0.29	0.18	0.08	0.18	68.62	1261	3.27	0.18	2043
Bandwidth	12	8	12	4	8	12	8	12	12	3
Observations	4004	3976	2633	2164	3976	4004	3836	3864	3864	2110
Panel D										
ITT, DD	0.04	0.10***	0.03	0.09**	0.13***	15.41**	572**	0.57**	0.10***	598*
(Reconst. Sc.)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(6.45)	(227)	(0.28)	(0.03)	(324)
Group means at follow-up (Within given Bandwidth of Panel B)										
Non-treated	0.69	0.46	0.28	0.16	0.30	98.36	2820	4.54	0.40	4157
Treated	0.81	0.68	0.38	0.29	0.55	132.79	3780	6.13	0.57	4717
Non-assigned	0.72	0.53	0.31	0.19	0.37	107.78	3035	4.96	0.43	4262
Assigned	0.79	0.64	0.36	0.27	0.50	126.56	3634	5.84	0.55	4665

Notes: Standard errors adjusted for clustering at event level reported in parentheses. Each cell represents an estimate from a separate regression, which includes the total score, the relative score (forcing variable), and an interaction of the relative score with the assignment variable as counterfactuals, as well as a constant. For comparability, Panel D estimates and follow-up means are conducted within the same bandwidths as in Panel B. Assigned/Non-assigned group means are based on the reconstructed score. *Non-farm wage IGA* and *Non-farm self IGA* are only available for 2011 and 2012 cohorts. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 5: EMPLOYMENT BY GENDER

	Any IGA (1=Yes)	Any non-farm IGA (1=Yes)	Non-farm wage IGA (1=Yes)	Non-farm self IGA (1=Yes)	Trade-specific IGA (1=Yes)	Hours worked in past month	Earnings	Logged earnings	Earnings > 3000 NRs. (1=Yes)	Earnings, conditional on any IGA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: HLATE, DD (Reconstructed Score)										
Women	0.26*	0.53***	0.21	0.46***	0.54***	90.43***	2113***	3.51***	0.52***	1590
	(0.14)	(0.14)	(0.13)	(0.15)	(0.15)	(24.67)	(784)	(1.11)	(0.13)	(1256)
Men	-0.09	-0.04	-0.11	0.03	0.18	-18.63	977	-1.20	-0.04	2461*
	(0.12)	(0.15)	(0.17)	(0.13)	(0.16)	(27.33)	(1028)	(1.10)	(0.14)	(1470)
Difference	0.35**	0.56***	0.32*	0.43**	0.36*	109.1***	1136	4.71***	0.56***	-871
	(0.16)	(0.19)	(0.18)	(0.17)	(0.20)	(30.65)	(1143)	(1.40)	(0.17)	(1560)
Bandwidth	12	12	12	11	12	12	9	12	12	12
Observations	4004	4004	2608	2606	4004	4004	3822	3838	3838	1931
Group means at Baseline (Within a given Bandwidth)										
Women	0.52	0.19	0.07	0.10	0.11	46.94	770	2.43	0.10	1492
Men	0.77	0.47	0.38	0.09	0.29	106.90	2130	4.77	0.34	2778
Panel B: Heterogeneous ITT, DD (Reconstructed Score)										
Women	0.08*	0.16***	0.07*	0.14***	0.16***	27.15***	659***	1.06***	0.16***	488
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(7.01)	(242)	(0.33)	(0.04)	(381)
Men	-0.01	0.01	-0.02	0.02	0.07	-1.88	386	-0.20	0.01	698*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(8.10)	(310)	(0.32)	(0.04)	(392)
Difference	0.09**	0.15***	0.09*	0.12**	0.09	29.03***	273	1.25***	0.15***	-210
	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(8.01)	(313)	(0.37)	(0.04)	(420)

Notes: Standard errors adjusted for clustering at event level reported in parentheses. Panel A presents second stage results. In both panels, each cell represents a treatment effect retrieved from a regression, which includes the group variable, an interaction term of the group variable with the training or assignment indicator, the total score, the relative score (forcing variable), and an interaction of the relative score with the assignment variable as counterfactuals, as well as a constant. *Difference* is the coefficient of the interaction term with the group variable from the respective regression. For comparability, all estimates are conducted within the same bandwidths as in Panel B, Table 4. *Non-farm wage IGA* and *Non-farm self IGA* are only available for 2011 and 2012 cohorts. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 6: TIME USE BY GENDER

	Hours/week working:					
	Unpaid, inside house > 5 (1=Yes)	Unpaid, inside house > 10 (1=Yes)	Unpaid, inside house > 20 (1=Yes)	Unpaid, outside house > 0 (1=Yes)	Paid, inside house > 0 (1=Yes)	Paid, outside house > 0 (1=Yes)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: HLATE, DD (Reconstructed Score)						
Women	-0.13	-0.12	0.10	0.03	0.42***	0.07
	(0.09)	(0.12)	(0.22)	(0.16)	(0.15)	(0.12)
Men	0.04	0.02	0.15	-0.11	0.08	-0.08
	(0.16)	(0.15)	(0.20)	(0.17)	(0.17)	(0.13)
Difference	-0.17	-0.14	-0.06	0.15	0.34	0.16
	(0.17)	(0.17)	(0.17)	(0.22)	(0.21)	(0.16)
Bandwidth	12	12	2	12	12	12
Observations	4003	4003	2133	4000	4004	4002
Group means at Baseline (Within a Given Bandwidth)						
Women	0.94	0.82	0.60	0.56	0.47	0.36
Men	0.61	0.32	0.11	0.70	0.59	0.69
Panel B: Heterogeneous ITT, DD (Reconstructed Score)						
Women	-0.04	-0.04	0.03	0.00	0.13***	0.02
	(0.03)	(0.03)	(0.06)	(0.05)	(0.04)	(0.03)
Men	0.01	-0.00	0.05	-0.03	0.04	-0.02
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)
Difference	-0.05	-0.04	-0.02	0.03	0.09	0.04
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.04)

Notes: Standard errors adjusted for clustering at event level reported in parentheses. Panel A presents second stage results. In both panels, each cell represents a treatment effect retrieved from a regression, which includes the group variable, an interaction term of the group variable with the training or assignment indicator, the total score, the relative score (forcing variable), and an interaction of the relative score with the assignment variable as counterfactuals, as well as a constant. *Difference* is the coefficient of the interaction term with the group variable from the respective regression. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 7: TYPES OF TRAINING

	By Event		By Applicant		Female Applicants
	Frequency	%	Frequency	%	%
Farming	5	2	92	2	88
Poultry	2	1	41	1	100
Food Prep/ Hospitality	16	7	260	6	54
Electrical/ Electronics/Computer	37	17	639	16	18
Handicraft & Incense	12	5	233	6	89
Construction/Mechanical/Automobile	63	29	1127	27	30
Beautician /Barber	11	5	239	6	100
Tailoring/ Garment/Textile	72	33	1457	36	98
Security Guard	1	0	13	0	100
Total	**219**	**100**	**4101**	**100**	

TABLE 8: EMPLOYMENT BY TRADE

	Any IGA (1=Yes)	Any non-farm IGA (1=Yes)	Non-farm wage IGA (1=Yes)	Non-farm self IGA (1=Yes)	Trade-specific IGA (1=Yes)	Hours worked in past month	Earnings	Logged earnings	Earnings > 3000 NRs. (1=Yes)	Earnings, conditional on any IGA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: HLATE, DD (Reconstructed Score)										
Food prep. & Hospitality	-0.34	-0.62	-1.73	1.64***	-0.73	-101.23	-5608	-4.98	-0.43	-156
	(0.43)	(0.64)	(1.14)	(0.55)	(0.69)	(110.95)	(4626)	(5.22)	(0.56)	(6068)
Electrician & Electronics	-0.25	0.04	0.02	0.03	0.42***	-48.10	-172	-1.53	-0.06	2028
	(0.19)	(0.17)	(0.19)	(0.17)	(0.16)	(35.92)	(1280)	(1.39)	(0.18)	(1644)
Handicraft & Incense stick making	0.11	-0.58	-0.22	0.25	-1.02	51.40	1150	-2.05	0.18	1715
	(0.47)	(0.57)	(0.14)	(0.24)	(0.80)	(56.27)	(2666)	(3.55)	(0.30)	(2900)
Construction & Mechanics	-0.08	-0.03	-0.01	-0.07	0.12	11.47	139	-1.49	0.00	1716
	(0.11)	(0.17)	(0.21)	(0.13)	(0.20)	(29.32)	(1139)	(1.27)	(0.16)	(1407)
Beautician & Barber	1.08*	1.06*	0.67**	0.34	0.99***	205.57**	4379	6.63*	0.56**	-1391
	(0.65)	(0.54)	(0.29)	(0.29)	(0.37)	(89.18)	(2721)	(3.62)	(0.27)	(4295)
Weaving, Tailoring & Garment Making	0.38**	0.82***	0.30*	0.57***	0.86***	112.5***	3572***	5.70***	0.75***	2985*
	(0.18)	(0.18)	(0.18)	(0.20)	(0.20)	(32.10)	(1127)	(1.59)	(0.19)	(1621)
Panel B: Heterogeneous ITT, DD (Reconstructed Score)										
Food prep. & Hospitality	-0.05	-0.09	-0.31***	0.31	-0.10	-14.68	-924*	-0.83	-0.06	-10
	(0.08)	(0.10)	(0.09)	(0.20)	(0.09)	(18.02)	(545)	(0.82)	(0.09)	(836)
Electrician & Electronics	-0.08	0.02	0.01	0.01	0.15**	-13.52	28	-0.42	-0.01	639
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(11.24)	(433)	(0.44)	(0.06)	(524)
Handicraft & Incense stick making	0.04	-0.12	-0.09	0.10	-0.22*	16.38	424	-0.34	0.07	675
	(0.12)	(0.11)	(0.07)	(0.10)	(0.12)	(14.99)	(629)	(0.77)	(0.07)	(1018)
Construction & Mechanics	-0.02	0.01	0.00	0.00	0.06	6.56	526	-0.29	0.02	552
	(0.04)	(0.05)	(0.05)	(0.03)	(0.06)	(9.25)	(347)	(0.38)	(0.05)	(383)
Beautician & Barber	0.25**	0.26***	0.21***	0.11	0.25***	49.38***	1096*	1.62**	0.15**	-366
	(0.11)	(0.08)	(0.07)	(0.09)	(0.07)	(18.88)	(600)	(0.77)	(0.06)	(1165)
Weaving, Tailoring & Garment Making	0.11**	0.24***	0.09*	0.16***	0.25***	32.32***	1043***	1.64***	0.22***	985*
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(8.09)	(304)	(0.39)	(0.05)	(509)

Notes: Standard errors adjusted for clustering at event level reported in parentheses. Panel A presents second stage results. In both panels, each cell represents an estimate from a separate regression, which includes the group variable, an interaction term of the group variable with the training or assignment indicator, the total score, the relative score (forcing variable), and an interaction of the relative score with the assignment variable as counterfactuals, as well as a constant. For comparability, all estimates are conducted within the same bandwidths as in Panel B, Table 4. *Non-farm wage IGA* and *Non-farm self IGA* are only available for 2011 and 2012 cohorts. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 9: EMPLOYMENT, SECOND FOLLOW-UP (2010 COHORT)

	Any IGA (1=Yes)	Any non-farm IGA (1=Yes)	Trade-specific IGA (1=Yes)	Hours worked in past month	Earnings	Logged earnings	Earnings > 3000 NRs. (1=Yes)	Earnings, conditional on any IGA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
LATE, DD	0.06	0.40**	0.33**	53.16	1277	0.65	0.40*	903
(Reconstr. Sc.)	(0.17)	(0.19)	(0.16)	(32.90)	(1362)	(1.52)	(0.23)	(1510)
First Stage	0.41***	0.41***	0.41***	0.41***	0.41***	0.41***	0.41***	0.47***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
F-statistic	48.33	48.33	48.33	48.33	42.58	42.58	42.58	36.07
Baseline mean	0.61	0.31	0.18	81.37	1388	3.63	0.20	2304
Bandwidth	10	10	10	10	10	10	10	10
Observations	621	621	621	621	590	590	590	336
Panel B								
ITT, DD	0.01	0.16*	0.14**	20.60	459	0.19	0.16*	349
(Reconst. Sc.)	(0.07)	(0.08)	(0.07)	(14.33)	(570)	(0.65)	(0.09)	(726)
Panel C: HLATE, DD (Reconstructed Score)								
Women	-0.01	0.30	0.18	32.57	1630	1.12	0.38	1220
	(0.26)	(0.25)	(0.23)	(38.49)	(1604)	(2.05)	(0.27)	(1680)
Men	0.13	0.50*	0.48*	76.71	1087	0.13	0.43	425
	(0.21)	(0.28)	(0.26)	(58.26)	(1957)	(2.48)	(0.35)	(2318)
Difference	-0.14	-0.20	-0.29	-44.15	542	0.99	-0.05	795
	(0.32)	(0.36)	(0.37)	(73.76)	(2408)	(3.38)	(0.42)	(2578)
Bandwidth	10	10	10	10	10	10	10	10
Observations	621	621	621	621	590	590	590	336
Group means at Baseline (Within a Given Bandwidth)								
Women	0.54	0.22	0.13	60.18	943	2.92	0.12	1736
Men	0.70	0.45	0.26	113.65	2074	4.73	0.33	2987
Panel D: Heterogeneous ITT, DD (Reconstructed Score)								
Women	-0.01	0.13	0.10	15.51	624	0.33	0.16	593
	(0.09)	(0.09)	(0.08)	(14.54)	(623)	(0.77)	(0.10)	(871)
Men	0.04	0.19*	0.18**	28.63	350	0.03	0.15	157
	(0.07)	(0.10)	(0.09)	(21.19)	(658)	(0.83)	(0.12)	(781)
Difference	-0.05	-0.06	-0.08	-13.12	274	0.30	0.01	436
	(0.09)	(0.11)	(0.11)	(20.81)	(672)	(0.96)	(0.12)	(854)

Notes: Standard errors adjusted for clustering at event level reported in parentheses. Each cell in Panel A and B represents an estimate from a separate regression, which includes the total score, the relative score (forcing variable), and an interaction of the relative score with the assignment variable as counterfactuals, as well as a constant. Panel C presents second stage results. For comparability, Panel B, C, and D estimates are conducted within the same bandwidths as in Panel A. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 10: TIME USE BY GENDER, SECOND FOLLOW-UP (2010 COHORT)

	Hours/week working ...					
	Unpaid, inside house > 5 (1=Yes)	Unpaid, inside house > 10 (1=Yes)	Unpaid, inside house > 20 (1=Yes)	Unpaid, outside house > 0 (1=Yes)	Paid, inside house > 0 (1=Yes)	Paid, outside house > 0 (1=Yes)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: HLATE, DD (Reconstructed Score)						
Women	0.20	0.32	0.30	0.43	0.26	-0.09
	(0.23)	(0.25)	(0.32)	(0.32)	(0.26)	(0.35)
Men	-0.24	0.23	0.26	-0.09	0.13	0.22
	(0.24)	(0.26)	(0.21)	(0.28)	(0.27)	(0.28)
Difference	0.44	0.10	0.04	0.53	0.13	-0.31
	(0.37)	(0.34)	(0.35)	(0.43)	(0.33)	(0.41)
Bandwidth	10	10	10	10	10	9
Observations	620	620	620	620	621	621
Group means at Baseline (Within a Given Bandwidth)						
Women	0.95	0.82	0.54	0.55	0.41	0.43
Men	0.65	0.36	0.14	0.65	0.48	0.67
Panel B: Heterogeneous ITT, DD (Reconstructed Score)						
Women	0.06	0.13	0.12	0.15	0.09	-0.02
	(0.08)	(0.09)	(0.12)	(0.11)	(0.11)	(0.13)
Men	-0.09	0.09	0.10	-0.01	0.07	0.07
	(0.09)	(0.09)	(0.08)	(0.10)	(0.10)	(0.10)
Difference	0.15	0.04	0.03	0.15	0.03	-0.09
	(0.11)	(0.10)	(0.11)	(0.12)	(0.10)	(0.12)

Notes: Standard errors adjusted for clustering at event level reported in parentheses. Panel A presents second stage results. In both panels, each cell represents a treatment effect retrieved from a regression, which includes the group variable, an interaction term of the group variable with the training or assignment indicator, the total score, the relative score (forcing variable), and an interaction of the relative score with the assignment variable as counterfactuals, as well as a constant. *Difference* is the coefficient of the interaction term with the group variable from the respective regression. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 11: SURVEY AND ADMINISTRATIVE DATA COMPARISION

	EF Verification Data	EF Verification Data	Survey Data	EF Verification Data	Survey Data
Months After Training	3	6	9-11	15	21-23
2010					
IGA with Earnings > 3000 NRs. (1=Yes)	79%[1]	-	62%	-	57%
2011					
Any IGA (1=Yes)	-	-	83%	72%[4]	-
IGA with Earnings > 3000 NRs. (1=Yes)	-	80%[2]	57%	-	-
2012					
Any IGA (1=Yes)	-	91%[3]	73%	-	-
IGA with Earnings > 3000 NRs. (1=Yes)	-	81%[3]	60%	-	-
IGA with Earnings > 4600 NRs. (1=Yes)	-	67%[3]	46%	-	-

Sources:

[1] EF Annual Report 2010, p2.

[2] EF Annual Report 2011, p5.

[3] EF Annual Report 2012, p4.

[4] EF Tracer Study 2013, p23.

APPENDIX

A. Reconstruction of Assignment Score and Threshold

FIGURE A1: APPLICATION FORM

Application Form

Registration #: _____

	Stamp of the Training Institution

Personal Details

Name and Surname: _____

Sex: ☐ Female ☐ Male ☐ Other

Marital Status: ☐ Single ☐ Married

Date of Birth (Day/Month/Year): _____/_____/_____   Age: _____

Caste/Ethnicity: ☐ Dalit ☐ Janajati ☐ Others (please specify)_____

Special circumstances: ☐ HIV-infected ☐ Disabled ☐ Widow
☐ Ex-combatants ☐ Internally Displaced People

Permanent Address	Current Address
District:	District:
Municipality/ VDC:	Municipality/ VDC:
Tole: Ward No:	Tole: Ward No:
Phone Number (Land line):	Phone Number (Land line):
Phone Number (Mobile):	Phone Number (Mobile):
Father's Name:	
Mother's Name:	**In case of getting information of you:**
Citizenship No.:	Reference person:
Issued District:	Mobile Number:

Education Details

Highest completed level of education:

☐ Illiterate ☐ Below class 5 ☐ Class 5-8 ☐ Class 9-10 ☐ SLC Pass ☐ +2
☐ Bachelor's degree and above
If you have an academic certificate, please attach a copy.

Employment and Income Information

What is your employment status and monthly earnings (own earnings only):

☐ Self-employed   Monthly earnings:_____ ☐ Wage earner   Monthly earnings:_____
☐ Agriculture   Monthly earnings:_____ ☐ Unemployed   Monthly earnings:_____
☐ Other _____   Monthly earnings:_____

Estimated total monthly income Rs. _____

What is your family's average annual earnings in the following areas (excluding your own)?

☐ Labour Wages _____ ☐ Salary _____ ☐ Business _____
☐ Animal Husbandry _____ ☐ Foreign income _____ ☐ Others _____

Number of family members _____

Estimated total monthly income per member Rs. _____
(Please divide the sum of total annual income by total no of family members and 12 months)

FIGURE A1: APPLICATION FORM (CONTINUED)

How many months can you feed your family through agricultural income from your own land?

☐ Less than 3 months   ☐ Less than 6 months   ☐ More than 6 months   ☐ Do not have land

<u>Training Information</u>

Name of training:_____

District: _____V.D.C: _____ Ward No.: _____ Start date:_____

Reason for interest in training:

☐ To start own work   ☐ To work for wages
☐ To upgrade skills   ☐ To go for foreign employment

How did you come to know about this training?

☐ Newspaper   ☐ Relatives/Friends   ☐ Muslim Women Society
☐ Poster   ☐ Training Centre   ☐ Dalit Women Association
☐ Pamphlets   ☐ Local Development Agencies ☐ Indigenous People Association
☐ FM radio   ☐ Single Women Group   ☐ Other _____

Have you had any previous training?   ☐ Yes   ☐ No

If yes, please provide the following information:

Name of Training:_____

Hours of Training:_____

Date completed:_____

I state that the above-mentioned details are true.

_____   _____

Signature of Applicant   Date (Day/Month/Year)

FIGURE A2: SAMPLE RANKING FORM

#	Name and Surname	Immediate contact telephone	Age 16-35 for men, 16-40 for women (Write age)	Education < SLC (completion of 10th grade)	Less than 6 months food sufficiency (rural areas) or income < NRs. 3,000	1. Trade-specific education (15)	2. Economic status (20)	3. Social caste (25)	4. Development status of district of origin (10)	5. Interview (30%)	TOTAL SCORE (100) List candidates from highest to lowest	Rank
			Entry Requirement (Y/N)			1-4. Short-listing (70%)				5. Interview (30%)	Final Score	Rank
1	Jane Doe 1	12345678	28	Y	Y	15	20	20	10	27	92	1
2	Jane Doe 2	12345678	29	Y	Y	15	20	20	10	25	90	2
3	John Doe 1	12345678	26	Y	Y	15	20	20	10	24	89	3
4	Jane Doe 3	12345678	20	Y	Y	15	20	20	0	30	85	4
5	Jane Doe 4	12345678	21	Y	Y	15	20	20	5	20	80	5
6	John Doe 2	12345678	24	Y	Y	15	15	20	5	25	80	6
7	Jane Doe 5	12345678	19	Y	Y	15	20	15	0	29	79	7
8	Jane Doe 6	12345678	33	Y	Y	15	15	25	10	13	78	8
9	John Doe 3	12345678	17	Y	Y	10	15	20	5	28	78	9
10	Jane Doe 7	12345678	21	Y	Y	15	20	15	5	22	77	10
11	Jane Doe 8	12345678	27	Y	Y	15	15	10	10	26	76	11
12	John Doe 4	12345678	23	Y	Y	15	20	10	10	20	75	12
13	Jane Doe 9	12345678	18	Y	Y	15	15	20	0	25	75	13
14	Jane Doe 10	12345678	35	Y	Y	15	15	20	0	23	73	14
15	John Doe 5	12345678	19	Y	Y	15	15	20	5	18	73	15
16	Jane Doe 11	12345678	22	Y	Y	15	0	20	10	27	72	16
17	Jane Doe 12	12345678	30	Y	Y	5	20	25	0	16	66	17
18	John Doe 6	12345678	25	Y	Y	15	10	20	5	15	65	18
19	Jane Doe 13	12345678	24	Y	Y	15	15	10	5	10	55	19
20	John Doe 7	12345678	32	Y	Y	15	15	10	5	6	51	20

Notes: Red line indicates cut-off between accepted and rejected candidates. Candidates are sampled for the survey if their score is within a range of the cut-off score plus/minus 20%. The shaded area represents candidates who would have been sampled for the baseline survey based on this example.

FIGURE A3: EFFECT SIZES BY DIFFERENT ABSOLUTE THRESHOLD VALUES

Estimates by Absolute Threshold Score

Graph reports estimates by groups of absolute threshold at course level. Only those estimates are reported which could be based on a sample size of at least 300 observations for ITT and 400 observations for LATE. Conditional earnings, non-farm wage employment, and non-farm self-employment are not reported as most bins feature less than 300 observations. We apply the user-written program rdrobust provided by Calonico, Cattaneo, and Titiunik (2014) to obtain bias-corrected estimates and robust confidence intervals.

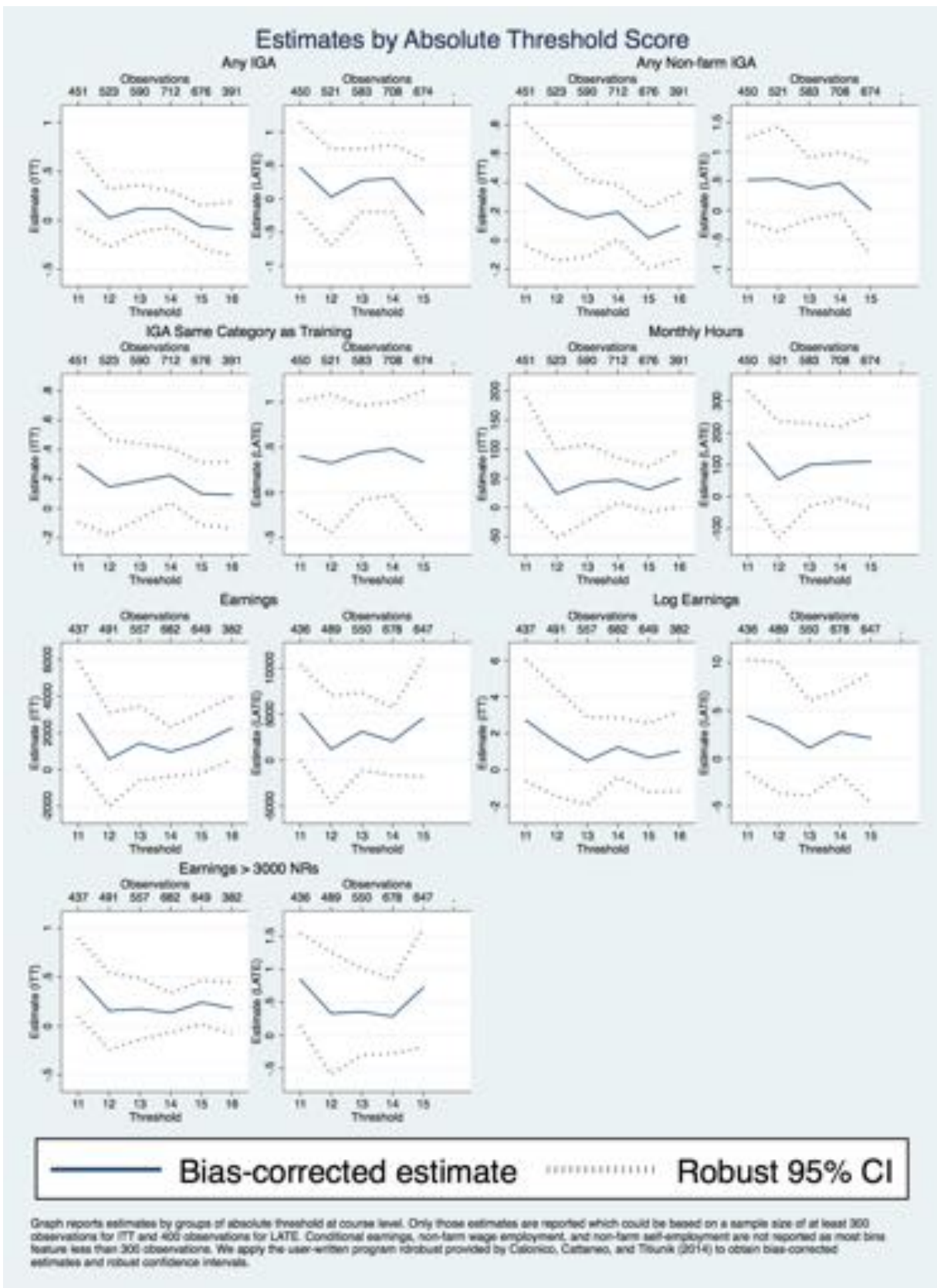**FIGURE A4: CROSS-VALIDATION FUNCTION FOR OUTCOMES IN LEVELS**

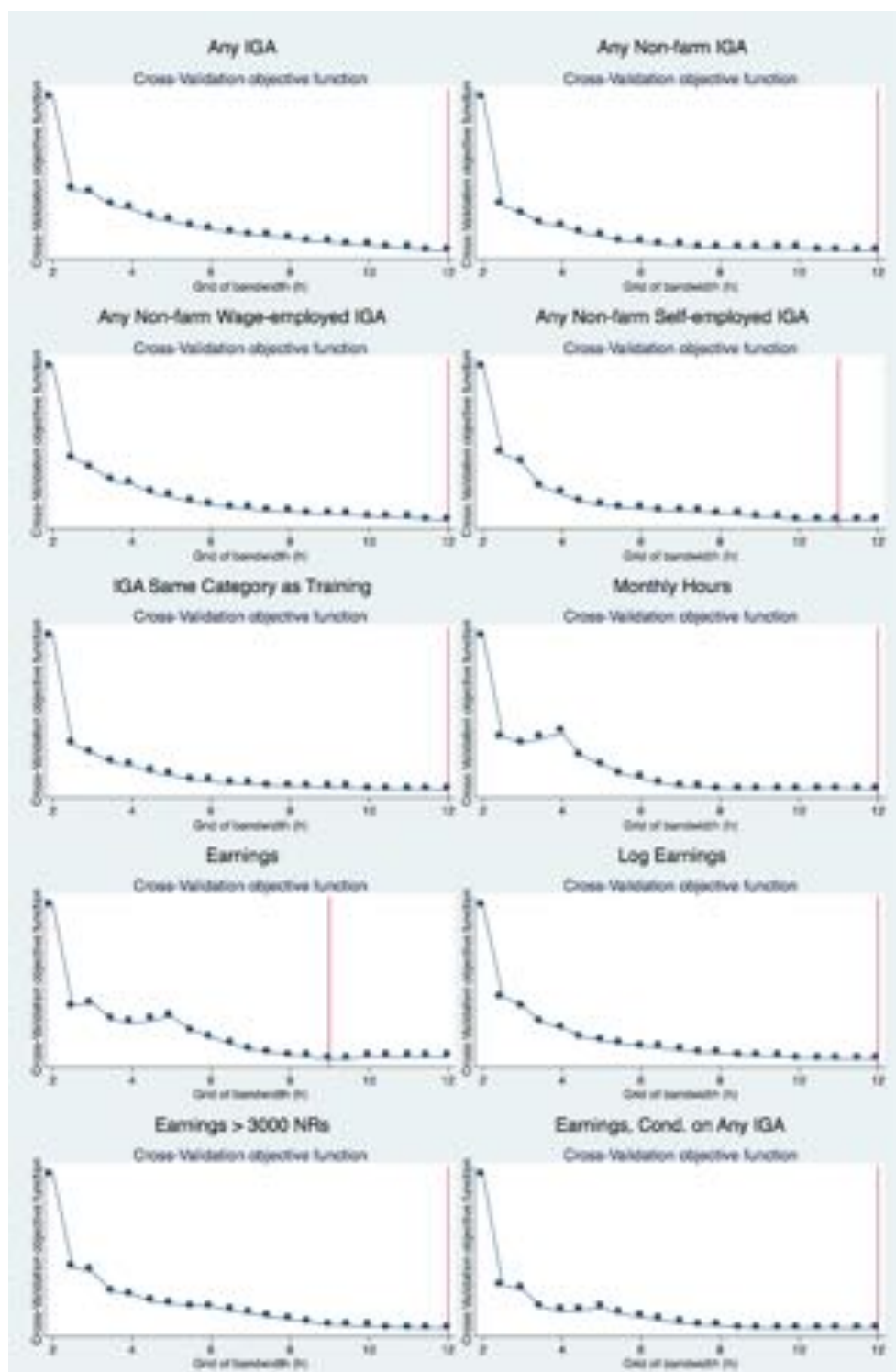**FIGURE A5: CROSS-VALIDATION FUNCTION FOR DIFFERENCED OUTCOMES**

Sub-Score	Total weight	Basis for evaluation	Indicators	Original Score	Reconstructed Score	
				Assigned Sub-Score	Assigned Sub-Score	Source
1	15%	Trade-specific education requirement	Compulsory prerequisite: All candidates must meet the minimum requirement for the training they applied to*	15 (0, 5 or 10 in exceptional cases)	Integers of 0 to 3	Predicted
2	20%	Economic poverty	Less than 3 months of food sufficiency	20	4	Application form data
			Less than 6 months of food sufficiency or less than 3000 per capita family income from non-farm based income	15	3	
			More than 6 months of food sufficiency and per capita family income from non-farm based income equal or more than 3000	0	0	
3	25%	Social caste	Women: Dalit women or women from the following special groups: widows; internally displaced; ex-combatants; physically disabled; HIV-infected infected	25	5	Application form data
			Women: Economically poor women not referred to above	20	4	
			Men: Dalit, Janjati, Madhesi men or men from the following special groups: internally displaced; ex-combatants; physically disabled; HIV-infected infected	15	3	
			Men: Economically poor men not referred to above	10	2	
			Neither of the above	0	0	
4	10%	Development status of district of origin	Least developed districts	10	2	Application form data
			Moderately developed districts	5	1	
			Developed districts	0	0	
Preliminary score for short-listing (Sub – total)				**70**	**14**	
5	30%	Interview	Commitment, Motivation, Attitude, Aptitude, Clear Vision for Employment and Enterprising	0-30	0-6	Predicted
Total score after interview				**100**	**20**	

Notes: *If candidates did not fulfill the course specific education prerequisites they were not eligible for short-listing and immediately rejected. Yet, in exceptional cases (approx. 9 percent of the sample) this criterion was not adhered to and instead applicants received 0, 5, or 10 points. When reconstructing this component, we therefore allow for integer values between 0 and 3.

TABLE A2: DEVELEOPMENT STATUS OF NEPAL DISTRICTS

Developed District	Moderately Developed District	Least Developed District
Kathmandu	Makawanpur	Ramechhap
Chitwan	Gulmi	Parsa
Jhapa	Surkhet	Rasuwa
Bhaktapur	Solukhumbu	Kapilbastu
Lalitpur	Banke	Bara
Kaski	Bhojpur	Dadeldhura
Dhankuta	Gorkha	Darchula
Palpa	Taplejung	Siraha
Syangja	Bardiya	Jajarkot
Manang	Kanchanpur	Rukum
Morang	Nuwakot	Sarlahi
Illam	Nawalparasi	Baitadi
Rupandehi	Khotang	Dailekh
Sunsari	Okhaldhunga	Rolpa
Kabhreplanchok	Kailali	Mahotari
Tanahu	Dolakha	Doti
Terhathum	Arghakhanchi	Dolpa
Sankhuwasabha	Udayapur	Rautahat
Mustang	Dhading	Jumla
Parbat	Salyan	Kalikot
Dang	Dhanusa	Bajura
Lamjung	Saptari	Achham
Panchthar	Sindhipalchok	Bajhang
Baglung	Sundhuli	Humla
Myagdi	Pyuthan	Mugu

Source: Districts of Nepal, Indicators of Development. Updated 2003. CBS/Nepal, ICIMOD. December, 2003

TABLE A3: CORRELATION OF 5TH SUB-SCORE WITH OTHER SCORE COMPONENTS

	Sub-Score 5 (1)	Sub-Score 5 (2)	Sub-Score 5 (3)
Aggregated Sub-Scores 1 to 4	-0.220***		
	(0.021)		
Sub-Score 1 (Trade-specif. Edu.)		0.117*	2.752***
		(0.062)	(0.431)
Sub-Score 2 (Econ. Poverty)		-0.182***	1.046*
		(0.029)	(0.551)
Sub-Score 3 (Social Caste)		-0.348***	1.013***
		(0.031)	(0.350)
Sub-Score 4 (Development Status of Dist. Of Origin)		-0.473***	3.560***
		(0.095)	(1.354)
SS 1 x SS 2			-0.120***
			(0.039)
SS 1 x SS 3			-0.136***
			(0.027)
SS 1 x SS 4			-0.379***
			(0.101)
SS 2 x SS 3			-0.049
			(0.030)
SS 2 x SS 4			-0.194**
			(0.096)
SS 3 x SS 4			-0.278***
			(0.087)
SS 1 x SS 2 x SS 3			0.006***
			(0.002)
SS 1 x SS 2 x SS 4			0.019***
			(0.007)
SS 1 x SS 3 x SS 4			0.026***
			(0.006)
SS 2 x SS 3 x SS 4			0.013**
			(0.006)
SS 1 x SS 2 x SS 3 x SS 4			-0.001***
			(0.000)
N	4090	4090	4090
Adj. R2	0.38	0.39	0.40

Notes: All models include event dummies. ***, **, and * denote significant at the 1, 5, and 10 percent level, respectively.

TABLE A4: ATTRITION

	Pooled Cohorts, 1st Follow-Up				2010 Cohort, 2nd Follow-Up			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
"Above Threshold"	0.008	0.017	0.007	0.016	-0.011	-0.047	-0.004	-0.052
	(0.010)	(0.019)	(0.010)	(0.019)	(0.027)	(0.045)	(0.029)	(0.047)
Female X "Above Threshold"		-0.020		-0.014		0.059		0.080
		(0.021)		(0.021)		(0.066)		(0.066)
Female		0.077***	0.066***	0.074***		-0.026	0.026	-0.019
		(0.021)	(0.014)	(0.020)		(0.060)	(0.035)	(0.056)
Age			0.004***	0.004***			0.007**	0.007**
			(0.001)	(0.001)			(0.003)	(0.003)
Parent			-0.003	-0.003			-0.006	-0.009
			(0.019)	(0.019)			(0.045)	(0.045)
Married			0.011	0.011			0.006	0.010
			(0.019)	(0.019)			(0.043)	(0.043)
Dalit			-0.052**	-0.052**			-0.052	-0.053
			(0.023)	(0.023)			(0.062)	(0.062)
Janjati			-0.014	-0.014			0.109***	0.110***
			(0.012)	(0.012)			(0.032)	(0.032)
Any IGA at baseline			0.017	0.017			0.024	0.023
			(0.011)	(0.011)			(0.027)	(0.027)
N	4585	4585	4585	4585	1547	1547	1547	1547
Training provider dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All models include district and training provider dummies. In Columns (3), (4), (7), and (8) indicators are included which flag missing values for each of the control variables, respectively. All models are estimated based on the reconstructed score. The difference in sample size between the initial baseline sample and the sample we use in this analysis arises due to missing values in the variables that were necessary to reconstruct the score variable, which determines assignment. All standard errors are clustered at event level. ***, **, and * denote significant at the 1, 5, and 10 percent level, respectively.

TABLE A5: BANDWIDTH SENSITIVITY

	Any IGA (1=Yes)	Any non-farm IGA (1=Yes)	Non-farm wage IGA (1=Yes)	Non-farm self IGA (1=Yes)	Trade-specific IGA (1=Yes)	Hours worked in past month	Earnings	Logged earnings	Earnings > 3000 NRs. (1=Yes)	Earnings, conditional on any IGA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: LATE, DD (Reconstructed Score)										
2 Index Scores	0.19	0.50**	0.35	0.14	0.43**	121.8***	3387**	2.33	0.43**	3058
	(0.21)	(0.20)	(0.31)	(0.27)	(0.20)	(42.98)	(1341)	(1.63)	(0.21)	(2254)
Observations	2133	2133	1364	1364	2133	2133	2049	2049	2049	1012
F-statistic	27.50	27.50	12.22	12.22	27.50	27.50	27.83	27.83	27.83	12.85
3 Index Scores	0.13	0.38**	0.28	0.18	0.45***	77.66**	2319**	1.52	0.34**	1722
	(0.16)	(0.15)	(0.19)	(0.17)	(0.16)	(31.35)	(1083)	(1.27)	(0.16)	(1645)
Observations	2874	2874	1862	1862	2874	2874	2759	2759	2759	1411
F-statistic	36.18	36.18	20.31	20.31	36.18	36.18	37.95	37.95	37.95	17.43
4 Index Scores	0.09	0.31**	0.15	0.25*	0.45***	58.86**	1755**	1.26	0.24**	1258
	(0.12)	(0.12)	(0.13)	(0.14)	(0.13)	(24.97)	(872)	(1.06)	(0.12)	(1097)
Observations	3340	3340	2145	2145	3340	3340	3202	3202	3202	1622
F-statistic	62.42	62.42	33.90	33.90	62.42	62.42	65.38	65.38	65.38	37.29
5 Index Scores	0.07	0.29***	0.10	0.25**	0.41***	52.67**	1666**	1.29	0.24**	1393
	(0.11)	(0.10)	(0.12)	(0.12)	(0.12)	(20.54)	(744)	(0.92)	(0.10)	(994)
Observations	3685	3685	2391	2391	3685	3685	3530	3530	3530	1778
F-statistic	77.89	77.89	42.88	42.88	77.89	77.89	79.88	79.88	79.88	42.02
10 Index Scores	0.11	0.29***	0.09	0.28**	0.41***	45.13**	1685**	1.53*	0.28***	1662
	(0.10)	(0.10)	(0.11)	(0.12)	(0.12)	(20.99)	(705)	(0.89)	(0.11)	(1017)
Observations	4000	4000	2604	2604	4000	4000	3834	3834	3834	1929
F-statistic	83.98	83.98	51.84	51.84	83.98	83.98	82.98	82.98	82.98	38.61
Panel B: ITT, DD (Reconstructed Score)										
2 Index Scores	0.06	0.15**	0.08	0.03	0.13**	36.23***	1033***	0.71	0.13**	885
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(12.15)	(392)	(0.51)	(0.06)	(602)
3 Index Scores	0.03	0.11**	0.07	0.05	0.13***	21.29**	662**	0.42	0.10**	455
	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(8.66)	(309)	(0.38)	(0.04)	(437)
4 Index Scores	0.03	0.10**	0.05	0.07*	0.15***	18.97**	587**	0.42	0.08**	435
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(8.26)	(295)	(0.37)	(0.04)	(386)
5 Index Scores	0.02	0.10***	0.03	0.08**	0.14***	17.88**	579**	0.45	0.09**	473
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(7.20)	(264)	(0.33)	(0.04)	(344)
10 Index Scores	0.03	0.09***	0.03	0.08**	0.13***	14.11**	532**	0.49*	0.09***	494
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(6.60)	(223)	(0.29)	(0.03)	(305)

Notes: Standard errors adjusted for clustering at event level reported in parentheses. Panel A presents second stage results. In both panels, each cell represents an estimate from a separate regression, which includes the total score, the relative score (forcing variable), and an interaction of the relative score with the assignment variable as counterfactuals, as well as a constant. *Non-farm wage IGA* and *Non-farm self IGA* are only available for 2011 and 2012 cohorts. ***, **, and * denote significant at the 1, 5, and 10 percent level.