Does Vocational Training Improve Business Practices Among Individual Microenterprises?

NUSRAT JIMI, State University of New York (Binghamton)
PLAMEN NIKOLOV, State University of New York (Binghamton) and Harvard Institute for Quantitative Social Studies

Using a "fuzzy" regression discontinuity design, we examine the short-run impacts of a large youth training program in Nepal on business ownership, business practices, and business performance using data from the Nepal Employment Fund training program. We find, approximately after twelve months of the start of the program, that the youth skill training generated a significant positive increase in business ownership of at least 8 percentage points (based on ITT estimates) and up to 30 percentage points for compliers (based on LATE estimates) from a baseline mean of 11 percent. There is an increase in sales record keeping practice among program participants compared to the control group. We also detect an average improvement of 7 to 31 percentage points (ITT and LATE estimates, respectively) in association with formal trade and business organizations. However, there is no significant positive increase on business performance in terms of monthly business earnings and having hired employee. Although training on average tend to improve business ownership, standard business practices, and association with formal trade organizations more for potential entrepreneurs compared to subsistence self-employed group, we do not find any statistically significant differential effect of training between them. The overall positive impacts on business ownership and business practices seem to be broadly driven by service and production-based businesses which are primarily operated by women – i.e., beautician, food producer, weaving and garments etc. Women, on average, benefit more compared to men in terms of gaining skills, applying standard business accounting practices, and getting involved into home-based businesses and self-employment activities that comply with the traditional social norms of Nepal.

I. Introduction

Entrepreneurship-support interventions have long been considered as a promising policy agenda to expand employment, earning opportunities and to reduce poverty (Schumpeter, 1912; Cho, Honorati 2013). The role of entrepreneurship in the development process is eliciting increasing attention from scholars and policymakers. This is particularly pressing in developing countries, where wage jobs are limited, youth unemployment is pervasive (40 percent of the world's unemployed) and the majority of jobs (53 to 80 percent) are created and operated in self-employment (Heinz and Valodia 2008; Haltiwanger et al., 2010; Ayyagari et al., 2011; Gindling, Newhouse, 2014; ILOSTAT, 2017). In this context, entrepreneurship skill and youth training programs are considered to have strong potentials to enable youth to gain skills and generate their own skilled jobs. Youth training programs are designed to build human capital and foster the acquisition of cognitive and non-cognitive skills, and the main expected outcome is improved employment. These programs can also facilitate business networking and alliance opportunities along with a platform to glean and share best practices that are relevant to particular businesses. Approximately 700 youth employment programs have been implemented in 100 countries till now, and around 82 percent of them offer skills training (World Bank IEG, 2013).

In this paper, we examine one of the largest youth training interventions in Nepal, which serves poor and disadvantaged young men and women by subsidizing skills training and employment placement services. We study the shot-run causal effects of the program on entry into entrepreneurship, business practices, business network, and business performance. In particular, our business ownership and business characteristics variables include business ownership and place of business operation. Outcomes reflecting standard business practices include business registration with government, keeping written records of business expenses, sales and inventory, and having separate business and personal account. Standard business accounting practices are considered to be the first link in the causal chain from skill training to business profitability and growth because training improves the knowledge and implementation of learned business practices by business owners (McKenzie and Woodruff 2013). For examining impact on business networking, we consider membership of formal trade and business association and business or work-related discussion with community members doing similar kind of business. Having at least one hired employee and monthly earnings from business are used as

indicators of business performance. We then examine the heterogeneity in program impacts by gender, type of entrepreneurs, baseline schooling and business skill. Finally, we examine how program impact varies by baseline business sectors to explore the types of businesses that benefit more from skill-training compared to other business sectors. We use the whole sample for examining impact on business ownership, whereas, for outcomes related to business characteristics, standard accounting practices, and business performance, we limit our attention to business owners only.

We use a panel data set of three consecutive cohorts of applicants to examine the impact of the program on business outcomes based on the regression discontinuity approach. Program eligibility was based on individual scores determined by a standardized application procedure 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 a treatment variable to estimate the programs intent-to-treat (ITT) effects. To distinguish between high and low baseline business skills/knowledge, we create a binary variable that reflects whether the individual had a baseline financial literacy score above or below the median.

Most of the self-employment microenterprises in developing countries generate the majority of non-farm jobs who work for themselves, earn little due to being rationed out of wage jobs (de Mel, McKenzie and Woodruff 2012; De Mel, McKenzie and Woodruff 2008) or because of their own preference for autonomy and flexibility (Maloney 2004). A small group among those self-employed become innovative and successful entrepreneurs with ambitions and potential for business growth, which in turn can translate to higher economic growth (Bennett and Estrin 2007; Ehrlich, Li, Liu 2017). Schoar (2010) distinguishes between the two types, calling them subsistence self-employed and entrepreneurs, where the first type typically operates small businesses as an alternative employment opportunity for themselves whereas the latter group – under the right circumstances – could grow their business enterprise rapidly and create jobs for others. To distinguish between subsistence self-employment and entrepreneurship, we first construct an index variable using principle component analysis, which is a scalar measure

combining various entrepreneurship-related traits based on Schoar (2010) and De Mel et al. (2010). We then use the index to create a binary indicator that enables us to distinguish between the two groups.

We find, approximately after twelve months of the start of the program, that the youth skill training generated a significant positive increase in business ownership of at least 8 percentage points (based on ITT estimates) and up to 30 percentage points for compliers (based on LATE estimates) from a baseline mean of 11 percent. Most of these businesses are new and initiated by women inside their homes. Program increases standard record keeping practices of business sales by 10 (ITT estimate) to 52 (LATE estimate) percentage points from a base of only 18 percent. There is an average improvement of 7 to 31 percentage points (based on ITT and LATE estimates, respectively) in program participant's association with formal trade and business organizations which may provide better networking and business alliance opportunities. However, we do not detect any evidence of statistically significant program effect on business registration with government as well as on having hired employee and monthly business earnings.

Our findings show that the effectiveness of the program differs by gender. Women seem to benefit more from the training program in terms of gaining skills, meeting new people, practicing standard business practice, and getting involved into economic activities that comply with the traditional social norms–i.e., starting home-based businesses and self-employment activities. Although training on average tend to improve business ownership, standard business practices, and association with formal trade organizations more for potential entrepreneurs compared to subsistence self-employed group, we do not find any statistically significant differential effect of training between the two groups. Next, we examine the impact heterogeneity by baseline years of schooling and business skill with the expectation that people with lower schooling and skill may benefit more from the training–i.e., training as a substitute of formal education among disadvantaged youth. We do not find any evidence that skill training is more beneficial to less educated and low-skilled business individuals except for business ownership–training participants with better baseline business skill experiences 20 percentage point more decrease in business ownership compared to low skilled individuals. Finally, we examine the program effects by sector of business by classifying them into four broad categories. Findings suggest that the overall positive impacts on business ownership and business practices tend to be

largely driven by the businesses in service and food producing sectors that are mostly run by women. Overall insignificant effect of skill training on monthly business income and hired employee might be coming from the opposite effects of manufacturing sector (negative effect) and food production and trading sector (positive effect).

This paper adds to the sparse knowledge base on the effects of youth training interventions in developing countries. Despite the rapid expansion of youth skill-enhancement vocational, managerial, and financial training programs around the developing world, evidence on the effectiveness of such entrepreneurship promotion programs is still scarce, and findings from existing impact evaluations are widely heterogeneous (Cho, Honorati 2013; Alzúa, Cruces and Lopez 2016). Most of the existing studies that attempt to examine heterogeneity in business practices and business performance outcomes by gender as well as other dimensions find statistically insignificant effect and offer little evidence, primarily due to low statistical power, on the question of what types of firms gain more from skill training (McKenzie and Woodruff 2013). Our study is among the few that rigorously evaluate the effect of a youth training program on business ownership and characteristics, business practices, and business performance in a low-income country context. We account for the heterogeneity among self-employed individuals and identify the impact of skill training program on business ownership, business practices, network, and business performance disaggregated by gender, owner's education, baseline business skill level, and business sectors. Moreover, we attempt to distinguish between subsistence self-employed and potential entrepreneurs to examine whether there is any differential impact across these two groups, and therefore, shed light on the question whether technical education and skill training are binding constraints on the growth prospects of entrepreneurs or not. Overall, these findings will provide useful information and guidance for effective targeting of youth skill training programs in poor economies.

The remainder of the paper is structured as follows. Section 2 provides background information on Nepal's Employment Fund training program and the intervention design. Section 3 describes the data, sampling strategy and sample characteristics. Section 4 presents the study design, and Section 5 describes the empirical strategy. Section 6 presents the results, and Section 7 provides some robustness checks. Section 8 concludes.

II. Background of Nepal's Employment Fund

Employment Fund (EF) is one of the largest skills training programs that funds vocational training along with placement services in Nepal. The program subsidizes short-term market-oriented training courses from existing training providers for young and disadvantaged individuals. Their primary objective is to place trainees into profitable employment upon completion of the training. Courses are announced publicly in local communities to encourage prospective candidates to apply. A standardized procedure based on eligibility criteria is used to select candidates among the application pool. Around 600 to 700 training courses are sponsored every year. As of 2010, the EF program covers 54 districts of Nepal, providing training for over 65 occupations and has become larger in the following years. Table 1 provides an overview of the total number of training events, trainees, and training providers for the time of this study period (2010-2012).

The program offers a bundle of services to the selected candidates. Technical training including certification and job-search-assistance are the primary components among those. Each technical training course lasts from one month to three months and vary across a wide range of trades (e.g., carpentry, welding, tailoring, incense stick rolling, and masonry). Trainees are also encouraged to complete a skills certification test offered by the National Skills Testing Board (NSTB) at the end of their course. After the classroom-based training, the program emphasizes on job placement services. Training providers are required to link trainees to trade-specific employers for six months of paid on-the-job-training. Usually, they use their networks of trainers and employers to find trade-specific internships for their trainees. This internship helps the trainees to acquire immediate work experience after the skill training. It also helps them to strengthen their contacts with potential employers. In addition to these core components, all female participants of the program receive a life-skills training of 40 hours (started in 2010 and fully implemented in 2011), which covered topics like workers' rights, sexual and reproductive health, negotiation skills, and discrimination response. A subset of trainees also received a short course in basic business skills.

The cost for training and employment services is pre-financed by the providers. The cost is reimbursed once training providers accomplish the pre-determined outcomes stipulated in the contract between them and the EF. Payments are made in three installments. A provider qualifies to receive the full price of the training and services provided plus a bonus based on whether the trainee belongs to a vulnerable group. The contract stipulates that all accepted trainees must complete their skills training successfully and sit for a skills test given by the NSTB. Providers obtain their first installment (40 percent of total payment) once the trainees complete the test. After completing the exam, providers need to ensure that graduates remain employed for the next six months and that they earn above 3000 NRs per month ("gainful employment"). The EF verifies every third and sixth months after training completion whether the trainees are gainfully employed or not. Once verified, the training provider obtains the second and the third installments which covers 25 percent and 35 percent of total payment, respectively.[1] Therefore, training providers carry a risk of losing their investments if trainees do not finish the skill training and providers fail to place them into gainful employment.

All EF-subsidized training programs use an eligibility criterion comprised of three factors to select candidates: age (from 16 to 35), education (less than ten years of formal education, or below SLC[2]) and low self-reported economic status. An applicant needs to meet all three criteria for being short-listed in the admission procedure. Providers also receive a bonus payment, as mentioned above, for successfully training and placing candidates from particular disadvantaged groups.[3] As long as the requirement for each installment is fulfilled, the bonus payment is calculated as a percentage of the full cost of training and services and issued together with the installment. 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. For poor women and poor-disadvantaged woman, the bonus payment is 70 percent and 80 percent of the base cost of training, respectively.[4]

[1] Employment status of a randomly selected sample of graduates is verified by EF field monitors.
[2] 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.
[3] Disadvantaged groups are defined by the EF as people belonging to the Dalit community, ex-combatants, internally displaced, widows (only women), disabled, HIV/AIDS infected, and formerly bonded laborers.
[4] Poverty is defined as less than six months of food sufficiency for farm households or less than 3,000 per capita family income, from non-farm based income.

For making additional efforts to target young women aged 16 to 24, the Employment Fund partnered with the World Bank in 2010. Training under the Adolescent Girls Employment Initiative (AGEI) was the same as regular EF training. However, certain events were flagged ahead of time and radio and newspaper ads were used in addition to regular training course advertisement. Many of these ads specifically encouraged young women to sign up for traditionally male trades, such as mobile phone repair, electronics, or construction.

III. Data

A. Sampling

We use data of three consecutive cohorts of applicants (from 2010 to 2012) from the EF-sponsored training application forms and the applicant selection procedure. Additionally, we use survey data collected at individual (applicant) and household level with two rounds of information for each cohort (baseline and follow-up). Figure 1 shows the survey timeline.

(Figure 1 about here)

Our sampling procedure combines stratified, random, and convenience sampling. Two consecutive steps were followed in the sampling of each cohort–a selection of training events followed by selecting the individuals according to standardized procedures. All EF funded training events that occurred between January and April of each year consists the event sampling-frame of our study. For administrative reasons, events were grouped into clusters of close-by districts. We then randomly sampled up to 15 district clusters in each of the three years, respectively. Next, we randomly selected 20 percent of the listed training events that took place in these district clusters. Events that were more likely to include young women were purposely oversampled in 2011 and 2012. A complete event listing was not available in advance for 2010. Therefore, the events were chosen by convenience (based on scheduling and accessibility) for this year. Table 1 presents the sample of events for the three cohorts.

(Table 1 about here)

The 2010 sample covers 12 districts and 65 events. The 2011 sample comprises 182 events across 28 districts. However, only 69 out of the 182 events were included in the baseline survey. The remaining 113 events were dropped either because the survey team could not reach the event on the applicant selection day (usually due to weather conditions) or because the event was not "oversubscribed".[5] 85 out of 112 sampled events covering 26 districts were included in the 2012 study sample. Figure 2 shows the study areas by survey.

(Figure 2 about here)

On the day of the applicant selection, a survey team visited each sampled training event to sample applicants. The sampling of applicants was based on the standardized interview procedure that was used to determine training assignments. Applicants received scores in five different categories during the assignment procedure (explained below). Adding those scores form the total score. Then, the applicants were listed from top scorer to the bottom scorer using a ranking sheet. This listing indicated the minimum score or threshold for admission to the course. The individuals Our study sample comprise individuals who fell in the range of 20 percent below or above the threshold, i.e., a subset of the ranked individuals. Immediately following the sampling of applicants but before the results of the selection process were announced, A baseline survey was administered following the sampling of applicants but before the announcement of the selection process. The follow-up survey was done after one year.

(Table 2 about here)

At baseline, the sample covered 4,677 individuals across all three cohorts (see Table 2). 88 percent of the baseline survey respondents were tracked and interviewed successfully for the follow-up survey that gives a panel of 4,101 individuals for analysis.[6]

[5]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 three rejected candidates who could be sampled for the control group (i.e., at least three candidates that fell within 20 percent of the threshold score).

[6] 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), respondent migrated for work within Nepal or abroad (8 percent), respondent migrated after marriage (10 percent), or other (40 percent).

B. Sample Characteristics and Sample Selection

Our aim in this paper is to study the impact of the youth training program on business ownership, business practices, and business outcomes. While skills and connection gained through the training program might encourage some individuals to start their own business activities, others opt for getting job or paid employment using the new learned skill. Therefore, we consider the full sample of 4,101 individuals for studying program impact on business ownership and business start-ups. For outcomes related to standard business practices, business networking, and business performance, we subset our sample to individuals who own or operate a business and examine whether skill training has an effect on their business practices and performance. Sub-setting to business owners yields a sample of 1,279 individuals in the pooled cohorts of 2010, 2011, and 2012. Among them, 500 individuals belong to 2010 cohort, 436 individuals belong to 2011 cohort and rest of them are from the 2012 cohort.

Table 3, Panel A gives an overview of baseline demographic characteristics of our total pooled sample of 4,101 individuals (i.e., 2010-2012 cohorts). Because of the restrictions on age, education level and poverty status for being eligible to apply to the EF-funded training courses, individuals in the sample are on average young, low educated and poor. The average age of individuals is 24 years and 63 percent of them are female. 38 percent of them have either no education or only primary education. Approximately 61 percent of the pooled sample was engaged in some income-generating activity in the month prior to the survey. If we restrict to non-farm income-generating activities, the employment rate falls to 30 percent. Panel B represents our outcome variables. Approximately 11 percent of the sample population owned some businesses in baseline. As mentioned in the above paragraph, we consider the full sample for business ownership and new businesses. When we look at our subset sample of business owners, we find that on average 55 percent of the baseline businesses are operated from home. Only 9 percent businesses are registered with government. Practice of record keeping is also very low, ranging from 18 to 21 percent. Most of the business owners (94 percent) are not associated with formal trade and business associations. The average monthly earnings from business were 3,829.63 NRs (equivalent to about 51 USD) at baseline.

(Table 3 about here)

Generally, women on average own more businesses compared to men at baseline, i.e., 12 percent and 9 percent respectively. However, most of these businesses are operated from home (64 percent for women and 36 percent for men), unregistered with government (4 percent versus 17 percent for men), and follows very little of the standard business practices. These, together, suggest that businesses operated by women are primarily small-scale self-employment activities. Lower average community involvement of women compared to men and less association with formal trade and business organizations suggest the strict social norms and immobility faced by women in Nepal.

C. Study Outcomes

Our outcome variables are grouped into four broad categories: business ownership and characteristics, business practices, business network, and business performance. All of them are dummy variables except monthly business earnings measured in NRs. Variables under business ownership and characteristics include ownership of business and place of business operation. 'Business ownership' is a dummy that represents whether the individual is currently owning a business or not. Whether the business is home-based or not is represent by the variable 'operate business from home' which gives us an idea about the size and category of the business. Our business practice variables include registration with government, keeping written business expense record, written sales record, inventory record, and having a separate business and personal account. Registration with government is a dummy that takes a value of 1 if the current business is registered with government or in the process of registration. This variable gives idea about the formality of the business. Keeping written records of business expenses, sales and inventory are dummy variables, each taking a value of 1 if the business owner follows the standard business management practices and 0 otherwise. The category 'business network' includes two variables. One is 'membership of trade and business association' which takes a value of 1 if the business owner is member of any formal trade and business association. The second variable is 'business networking' which denotes whether or not a business owner

discusses and exchanges businesses or work-related ideas (at least once a week) with people within the community doing similar businesses. Finally, the 'business performance' category is comprised of two variables– a dummy variable that denotes whether the business has at least one hired worker or not and the respondent's total monthly earnings from the business measured in NRs. These two variables capture expansion or growth of the current business. There is another outcome variable named 'new business startup' which takes a value of 1 if the current business is being operated from no more than twelve months (i.e., established after the baseline). We use this variable to check skill training helps existing business owners to expand or open new business.

IV. Study Design

The decision of admission to the training program was based on each candidate's total score and the course-specific threshold score. Training providers followed a detailed guideline–a clear rubric for scoring, ranking the shortlisted candidates by score, and selecting the top-scoring candidates for admission to the program. Given a threshold score for each course, applicants above the score were assigned to the program, and applicants below the program were not assigned. Providers were advised to shortlist 50 percent more candidates than the available number of seats to form a sufficiently large quasi-experimental control group.

The total individual score consisted of five sub-scores that depends on: (a) applicants' trade-specific education (prerequisite, 15 points)[7], (b) economic status (up to 20 points), (c) applicants' gender, social caste, and special circumstances (up to 25 points), (d) development status of applicants' district of origin (up to 10 points), and (e) an interview score determined by a selection committee (up to 30 points). Information in the application form were used to determine scores for the first four components (Figure A1 in Appendix A). The application form lists the exact criteria upon which the distribution of scores was based. Tables A1 and A2 show how these criteria were converted into numeric scores within each of the four categories. Candidates were short-listed and invited for an interview based on the total scores in the four components. The interview committee had three to five members consists of representatives of

[7]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 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.

training institution and potential employers. This committee, usually in presence of donor institution and survey firm representatives, assessed each candidate's motivation, commitment, attitude, aptitude, and clear vision for employment and enterprising to assign the fifth sub-score. Aggregating the interview score with other four sub-scores, there was a total score for each that ranged from 0 of 100.

The total score was then used to rank the candidates and decide whether or not they were eligible for the training in a course. Available seats were assigned starting with the top-scorer. Therefore, the threshold in each course was based on two things: distribution of candidates' scores and the pre-determined number of available seats, and it varies for each course (we take this into account in the empirical strategy section). A sample ranking form used by training providers is presented in Figure A2.[8] Although eligibility for training, based on the actual score, influenced the likelihood of training course enrollment, individual assignment to training was not automatic, primarily for two reasons: non-compliance and manipulation of the assignment procedure.

A large mismatch was found after cross examining the training compliance with selection procedure. We found that approximately 30 percent of the training assigned group did not take-up the training, whereas 32 percent of the non-assigned applicants did. This situation can arise when individuals not originally assigned to training, but next in line as per the ranking form, were given the chance in absence of originally assigned individuals. We check whether and how the probability of treatment jumped at the threshold scores. We subtract each individual's assignment score by the course-specific threshold score and obtain a standardized relative score around the cut-off of zero. Then, we plot the probability of treatment against the relative assignment score. Figure 3's left graph shows the results. Although there is a clear jump in the probability of program participation at the cut-off, the jump is less than one. We account for this fact in our estimation strategy.

(Figure 3 about here)

[8] 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 on the list was 73. Hence, in this example, 73 would be the threshold score for this particular course.

When we plot the density of the relative score (figure 4, left graph), a discontinuity in the distribution is observed around zero

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.[9] In contrast, providers may have had the opportunity to precisely alter the scores of those candidates who seemed favorable to them.[10] Although the selection committee included several persons from different interest groups and was designed to avoid this type of manipulation on 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 specifically address this manipulation issue by reconstructing the assignment score, which we describe in more detail below.

(Figure 4 about here)

V. Empirical Strategy

We use a non-parametric regression discontinuity strategy by running local linear regressions for estimating the effect of the training program on business characteristics and practices. Specifically, we estimate local average treatment effects (LATE) for the people who comply with the assignment status, intent-to-treat (ITT) effects, and heterogeneous effects by gender, trade and type of entrepreneur within these two frameworks. We describe the estimation strategy and how we reconstruct the assignment score below.[11]

[9] Lee and Lemieux (2010) distinguish between *precise* and *imprecise* manipulation. While applicants were certainly able to manipulate the information that 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, which, as mentioned earlier, occurred after the application forms were 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).

[10] 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 in the manipulation.

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

A. Treatment Effect Estimators

We employ a fuzzy regression discontinuity set-up to address imperfect compliance to treatment and estimate the Local Average Treatment Effects (LATE).[12] We run the following first-stage equation:

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

where D_i is the treatment dummy that takes a value of 1 if applicant i has received training and 0 otherwise, and the excluded instrument T_i indicates whether or not an applicant i has been assigned to training (i.e., whether the absolute assignment score X_i of the applicant is greater than or equal to the threshold score t of the respective course c he or she applied to). 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:

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

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 a control variable to account for the heterogeneity in the cut-off values across courses. It is important to note that LATE estimates are based only on those candidates who comply with program assignment, and therefore, not necessarily equal to the population average treatment effects.

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. The overall Intent-to-treat effects (ITT) of the program is of policy relevance. We estimate the following reduced form equation:

[12] similar to the one proposed by Hahn, Todd, and Van der Klaauw (2001).

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

where T_i is an indicator variable taking a value of 1 if an applicant *i* has been assigned to training and 0 otherwise. 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 and the LATE estimates as the upper bound estimates of the program impacts.

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

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

where H_i is an indicator for the 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 $H_i D_i$.

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

B. 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 should theoretically generate an exogenous change in the probability of training, holding individual characteristics constant. However, and as mentioned previously in Section IV, 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 favored by providers, in which case our estimates would be upward-biased. On the other hand, it is also conceivable that providers favored disadvantaged 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 such as 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 or 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 the 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.

To overcome this challenge, we follow the approach by Miller et al. (2013) who reconstructed 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).[13] 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 with 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

[13] 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.

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 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 are 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' assessments 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.

C. 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.[14] To construct a valid instrument, we re-estimate the threshold scores for each course following the approach proposed by Miller et al. (2013).[15] The authors' proposed approach depends 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.

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

D. 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 estimated eligibility thresholds. To examine whether the 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, except 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 Figures 5, 7, and 8 that the subgroup indicators we use to estimate heterogeneous treatment effects (i.e., an applicant's gender, type of entrepreneur, baseline education, business skill, and business sector) 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)

(Figure 8 about here)

E. Distinguishing Entrepreneurs from the Subsistence Self-employed Type

We allow for heterogeneity among self-employed individuals and separate them into two groups. The first group we attempt to identify are those who become self-employed as a means of providing subsistence income, who do not grow or create employment opportunities for other workers in the economy–which we call the subsistence self-employed. The second group, which we call potential entrepreneurs, are the self-employed who act like entrepreneurs and aim to create larger and more dynamic ventures that eventually tend to grow well beyond the scope of their subsistence needs. these two groups differ in multidimensional set of domains–skills, ability, economic objectives, and their role in the economy. To distinguish between the two types, we build on a framework by Schoar (2010) and De Mel et al. (2010) who examine and identify several dimensions that separates the two groups: human capital, cognitive skills, financial literacy and attitudes towards risk. The authors find that the transformational entrepreneurs scored much higher on different measures of IQ, willingness to take risk, motivation, and the level of managerial and financial literacy.

Based on information from our baseline survey, we construct an index using several variables that relate to entrepreneurship-related traits based on Schoar (2010) and De Mel et al. (2010), which we then use to create a binary indicator that enables us to distinguish subsistence self-employment from entrepreneurship. In particular, we include variables reflecting individual's years of schooling, managerial and financial literacy,[16] analytical ability,[17] entrepreneurial ability,[18] and attitude towards risk of participants. The index was constructed using the Principal Component Analysis (PCA). PCA is used to reduce the dimensionality of correlated variables that capture various domains of the same intended concept–entrepreneurship in our case. Reduction in variable dimensionality is achieved by estimating *n* weighted linear combinations

[16] Our financial literacy score is based on responses in the following areas: whether one how to make a budget for the household, whether one had actually written a budget before, whether one kept track of his or her money, ability to compute simple and compound interest, and knowledge of bank services and products. The score ranges from 0 to 23 and baseline mean score is 7.00.

[17] Measured by individual's score in the cognitive Raven's progressives test based on six questions. The questions consisted of visual geometric design with a missing piece and each respondent was asked to fill in the missing piece based on a logical pattern from six given choices to pick from. The score ranges from 0 to 5 and baseline mean score is 2.91.

[18] Calculated from a survey module which assessed each respondent's ability in several entrepreneurship-related areas: to run a business, to work in a team, identify income generating activities, to obtain credit from a financial institution, to manage financial accounts and to collect the money someone owes to him or her. The score ranges from 0 to 24 and baseline mean score is 9.50.

containing the proxy variables. The linear transformation produces *n* uncorrelated components (linear combinations) that are the eigenvectors of the system; combined they contain the same information as the original variables. The eigenvalue (variance) for each principal component indicates the percentage of variation in the total data explained. By design, the first component contains the most information (largest eigenvalue), whereas the last contains the least. We create an index by retaining the component with the largest overall variance (eigenvalue). Since PCA is sensitive to scaling and our variables don't have similar scales, we use the correlation matrix in the weighting procedure.

In our analysis, we find that the first component explained 59 percent of the overall variance and has an eigenvalue of 1.43 (total variance is 2.44). The first two components explain about 61 percent of the total variance. The weights for the first component enter positively in the linear equation describing the composite score, which suggests that high levels of the variables, within the first component, translate into higher levels of the composite index. We opt for the first component as a collective proxy of entrepreneurship in our analysis. Table A6 in Appendix A exhibits the PCA component loadings based on the four measures discussed previously. To create the binary indicator for distinguishing entrepreneurship from subsistence self-employment, we divide the index into five equal sections and participants in the top 20th percentile is assigned a value of 1 and termed as entrepreneurs.

VI. Short-Run Results on Enterprise-Related Outcomes

In this section, we present our findings on the impact of the vocational training program on business related outcomes for the pooled 2010, 2011, and 2012 samples based on the specifications described in Section 5. We focus on impact estimates based on the reconstructed score because the estimates using the original score might be biased due to possible manipulation of original score variable that led to unbalanced individual characteristics and outcome variables at baseline.

A. Impacts on Business Involvement, Business Ownership, and Business Characteristics

Table 4 presents the estimates of the impact of vocational training on business ownership, business practices, and business performance. Panel A shows the local average treatment effects estimated from equations (1) and (2)–impact of the program on individuals who complied with the training assignment on differenced outcome variables using the original score. We find significant positive impact of 23 percentage points on business ownership. In terms of business practices, we do not find any significant positive effects of the training program on record keeping practices (i.e. business expenses, sales, and inventory record) and having separate business and personal account. We also detect no statistically significant positive effects on registration with government and membership of trade and business association. There is a positive effect of around 6,919.43 NRs on monthly earnings from business, however, high standard error leads to statistical insignificance of the coefficient. As mentioned before, estimates in panel A might be biased due to the possible manipulation of original score variable and therefore, we turn our attention to the analysis using the reconstructed score in Panel B.

(Table 4 about here)

Panel B shows the local average treatment effects on differenced outcomes using the reconstructed score. We find statistically significant increase of 30 percentage points of business ownership (from the base of 11 percent) as compared to 23 percentage points from using the manipulated score (column 1, Panel A). Businesses of training participants are not significantly more likely to be operated from home compared to the businesses of our control group. Using the reconstructed score, we find significant positive effect of 52 and 38 percentage points of the training on written sales record keeping and maintaining separate business and personal accounts, respectively. The EF program led to improvements in trainee's association with trade and business organizations, an increase of 31 percentage points from the baseline of only 6 percent (column 8, Panel B), which will give them formal networking opportunities along with a platform to glean and share best practices that are relevant to their business. However, training participants are not more likely to meet and discuss their businesses and work-related things with people

doing similar businesses within their community (column 9, Panel B) compared to the control group. Unlike panel A, we find no evidence of positive effect of the training program on business earnings using the reconstructed score and the coefficient is statistically insignificant.

We use differenced outcomes to address the (remaining) disparities in participant characteristics we observed at baseline in Panels A and B. In Panel C, we alter this feature by using outcomes in levels. Comparing the estimates using reconstructed score (Panel B and C), we find that the effect sizes for business ownership and business characteristics outcomes are higher in the level form (Column 1 and 2, Panel C) than the differenced outcomes. The reduced effect size in panel B might have come from the fact that our credit assigned group is slightly younger on average and older individuals may be more involved in business activities compared to young people. Also, they are less likely to be Dalit than the non-assigned group. Individuals of Dalit ethnicity usually face substantial labor market discrimination in the context of Nepal, and therefore, it might make them more likely to be self-employed and operate small businesses from home compared to non-Dalits.

Results reported in Panel A, B and C are based on individuals who complied with their treatment assignment status and, and therefore, may differ from individuals in the full sample. In Panel D, we examine how altering this feature changes our findings, and estimate the intent-to-treat (ITT) effects based on Equation (3) in Section 5. Our ITT coefficients are much lower compared to the LATE estimates. However, they still indicate that training led to a significant increase of 8 percentage points in business ownership, 10 percentage points in keeping written sales record, 7 percentage points in membership of trade and business association (from the base of 1 percent), and 17 percentage points in networking with community people doing similar businesses. Complier-population in our sample might differ from the non-compliers–i.e., compliers may be a subgroup of individuals for whom the returns to training participation are different than for the overall sample. Although program eligibility criteria 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. For example, these particular groups of applicants are likely to have lower baseline educational levels and higher returns to additional education as compared to the rest of the

sample. Presence of caste- and gender-based discrimination and extreme poverty can be another issue. Therefore, we interpret the ITT estimates as lower-bound of the program effects and LATE coefficient estimates as upper-bound effects.

Appendix table A5 shows the impact estimates on business ownership, business practices, and business outcomes of training participants who had businesses in baseline. Findings for existing business owners are similar to our findings in table 4. We find significant positive effect of 70 percentage point on business ownership (column 1 Panel A) and 85 percentage points (column 2 Panel A) on opening of new businesses, respectively. This finding suggests that the skill training do not make existing business owners to switch from business to wage employment, and rather, it helps them to expand their businesses. There is 55 percentage point increase in written sales record keeping practice (from the base of 11 percent). Training participants with existing businesses are significantly more likely to be associated with formal trade and business associations, i.e., 31 percentage point increase from the base of only 7 percent. Unlike the overall sample of business owners, we find positive effect of training on having hired employee and business earnings of existing businesses. However, the standard errors are high which lead to statistical insignificance of the estimates.

B. Impacts Disaggregated by Gender

Tables 5 shows the impact of the program disaggregated by men and women based on the specifications described in section 5. Panel A shows the LATE estimated from equation (4).

(Table 5 about here)

Our findings reveal that the effectiveness of the training program differs by gender. The LATE estimates of business ownership and business practices are larger and statistically significantly different for women as compared to men. We find significant positive impact of 32 percentage point increase in business ownership for female program participants compared to male members. Businesses owned by women are significantly more likely to be operated from home (column 2, Panel A) which can be explained by stricter social norms about responsibilities and mobility faced by Nepali women. One interesting finding is that skill training changes

women's current business practices more compared to male (column 4 to 6, Panel A). Women are more likely to follow standard business practices (i.e., keeping written business expenses, sales and inventory records) that are considered to be precondition of improved business outcome and performance. By examining the types of businesses, we find that women with the EF training are heavily involved in service and production related businesses–i.e., beautician, food producer, weaving and garments etc. Therefore, women seem to benefit more from the training program in terms of gaining skills, applying standard managerial practices, and getting involved into income generating activities that comply with the traditional social norms–i.e., starting home-based businesses and self-employment activities. However, we do not detect much differences in business networking in terms of association with formal trade and business association and business-related discussion with community members (column 8 and 9, Panel A). There is also no significant difference in business earnings across the two groups. Our ITT results in Panel B show smaller effect size compared to panel A. However, most of the outcomes are not statistically significantly different between the two groups except business ownership which suggest that women who complied with their treatment assignment status might be different from women in the full sample.

C. Impacts Disaggregated by Entrepreneur-Type

Next, we examine whether there is any difference in the effectiveness of skill training program for subsistence self-employment and potential entrepreneurs. Subsistence self-employed individuals may have very low level of education and business skills and therefore, it should be relatively easy for them to learn through training, implement the practices taught, and improve performance. On the contrary, skill training might be more helpful for potential entrepreneurs to start businesses with longer-term commitment and investment than subsistence self-employed group who run businesses only because they cannot find a wage job and are less interested or able to implement the practices taught. For effective targeting of training programs, it is important to know the type of micro-enterprises in which skills are binding constraint on business performance and growth, especially in the context of poor developing economies. In table 6, we represent the impact estimates of the training program disaggregated by the two groups.

(Table 6 about here)

Potential entrepreneurs tend to own more businesses compared to the subsistence self-employed group. Their businesses tend to be more market-based, registered with the government, follows standard business practices, and they are more likely to be associated with formal trade and business organizations. For subsistence self-employed individuals, training seems like to have more of a poverty alleviating effect–i.e., increase in community-based business networking and ownership of home-based businesses, as opposed to acting as a catalyst to remove constraints for further business growth. However, differences in outcomes between the two groups are not statistically significant due to small sample size and large standard errors.

D. Impacts Disaggregated by Business Owner's Education and Baseline Business Skill

Skill training programs usually aim to create an opportunity for individuals who drop out of the formal education system to obtain and upgrade job-specific skills that can improve their employment prospects. Therefore, vocational training is sometimes considered as a substitute of formal education among disadvantaged youth. In this case, one can expect that people with lower schooling and skill may benefit more from the training program. We examine this issue by disaggregating the EF program effect by individual's baseline education and business skill. In terms of schooling, we divide individuals into two groups–with primary or lower level education and above primary education. To distinguish individuals in terms of high and low baseline business skills or knowledge, we create a binary variable that reflects whether the individual had a baseline financial literacy score above or below the median.[19] Table 7 presents the estimates of the disaggregated program impact.

(Table 7 about here)

Panel A and C show the heterogeneous LATE (coefficient of the interaction term) estimated from equation (4) by owner's baseline education and business skills, respectively. We do not find any significant difference in program effectiveness by education for almost all the outcome

[19] Creation of the financial literacy score is explained in sub-section E of section 5.

variables. On the other hand, training program has a significant negative effect of 20 percentage point more decrease in business ownership of individuals with better business skills compared to low business skilled individuals (column 1 Panel C). ITT estimates in Panel B and D show similar results to the LATE estimates with smaller effect sizes. Overall, we do not detect any evidence that skill training is more beneficial to less educated and low-skilled business owners.

E. Impacts Disaggregated by Baseline Business Sectors

We classify businesses into four broad categories: food producer, manufacturing, buying and selling of goods, and service, and examine the program effects by sector of business. Rest of the businesses are put into others category. Table 8 presents the impact estimates of the training program by baseline business sectors.

(Table 8 about here)

The food producing sector shows strong positive impacts on business ownership (a 65-percentage point increase), having at least one hired employee, and monthly business earnings (Panel A). On the other hand, service-based businesses show significant effect on home-based business start-ups and standard business record keeping practices with statistically insignificant positive effect on monthly business earnings. For the sectors of manufacturing and buying and selling goods, we do not detect conclusive and significant impacts. Businesses under the manufacturing sector seem to be less home-based with negative effect on business network and business outcomes. On the other hand, buying and selling goods sector shows negative effect on businesses ownership but significant positive effects on membership of formal business association, having hired employee, and monthly business earnings. Overall, results in Tables 8 reflects significant heterogeneity on business-related outcomes across the four broad sectors. The overall positive impacts on business ownership and business practices discussed previously seem to be broadly driven by the businesses in service and food producing sectors that are primarily operated by women. Insignificant effect of skill training on monthly business income and hired employee might be due to the cancellation of positive effects from food production and trading

sector by the manufacturing sector. Our ITT estimates in Panel B shows similar findings to the LATE estimates with smaller and in some cases insignificant effect sizes.

VII. Robustness Checks

A. Differential Attrition

Group-wise differential attrition can affect the findings of our study. For example, results will be bias upward if trainees who did not participate to the training program are more capable candidates of the control group and they go to abroad for work. In Table A4 of Appendix A, we explore the possibility of "differential attrition" and do not find any 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.

B. Alternative Bandwidths

To investigate the robustness of results relative to bandwidth choice, we re-estimate equations (1), (2) and (3) using alternative bandwidths– 4, 6, 7, 8 and 10. Table A5 of Appendix A presents the results. Findings are more or less stable both in terms of the coefficient magnitudes and statistical significance. Magnitudes of the effect sizes usually increase with higher bandwidth choice. The effects for the female, and entrepreneur sub-samples also remain similar at different bandwidths (not shown in table).

VIII. Discussion and Concluding Remark

References

Alzúa, M. L., Cruces, G., & Lopez, C. (2016). Long-Run Effects of Youth Training Programs: Experimental Evidence from Argentina. Economic Inquiry, 54(4), 1839-1859.

Ayyagari, M., A. Demirgüç-Kunt and V. Maksimovic (2011). "Young vs. Small Firms across the World: Contribution to Employment, Job Creation, and Growth." World Bank Policy Research Working Paper No. 5631. Washington DC: World Bank.

Bennett, J., & Estrin, S. (2007). Entrepreneurial entry in developing economies: Modeling interactions between the formal and informal sector.

Cho, Y., & Honorati, M. (2013). Entrepreneurship programs in developing countries: A meta regression analysis. The World Bank.
Gindling, T. H. and D.L. Newhouse (2012). "Self-Employment in the Developing World" World Bank Policy Research Working Paper No. 6201.

De Mel, S., McKenzie, D., & Woodruff, C. (2009). Are women more credit constrained? Experimental evidence on gender and microenterprise returns. American Economic Journal: Applied Economics, 1(3), 1-32.

De Mel, S., McKenzie, D., & Woodruff, C. (2008). Who are the microenterprise owners? Evidence from Sri Lanka on Tokman v. de Soto. The World Bank.

De Mel, S., McKenzie, D., & Woodruff, C. (2012). Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. The World Bank.

Ehrlich, Isaac, Dunli Li, and Zhiqiang Liu. "The role of entrepreneurial human capital as a driver of endogenous economic growth." Journal of Human Capital 11, no. 3 (2017): 310-351.

Haltiwanger, J., R. Jarmin, and J. Miranda (2010). "Who Creates Jobs? Small vs. Large vs. Young," NBER Working Paper No. 16300. Cambridge, Mass.: National Bureau of Economic Research.

Heintz, J., & Valodia, I. (2008). Informality in Africa: A review. Background paper for the Swedish International Development Cooperation Agency (Sida), Unpublished Working Paper.

International Labour Office (ILO). (2017). World employment and social outlook: Trends 2017. Geneva: ILO.

Maloney, W. F. (2004). Informality revisited. World development, 32(7), 1159-1178.

McKenzie, D., & Woodruff, C. (2013). What are we learning from business training and entrepreneurship evaluations around the developing world?. The World Bank Research Observer, 29(1), 48-82.

McKenzie, D. (2017). How effective are active labor market policies in developing countries? a critical review of recent evidence. The World Bank Research Observer, 32(2), 127-154.

Schoar, A. (2010). The divide between subsistence and transformational entrepreneurship. Innovation policy and the economy, 10(1), 57-81.

Schumpeter, J. (2003). Theory of economic development. In Joseph Alois Schumpeter (pp. 5-59). Springer, Boston, MA.

World Bank Group. Independent Evaluation Group (IEG). (2013). Youth employment programs: an evaluation of World Bank and International Finance Corporation support.

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

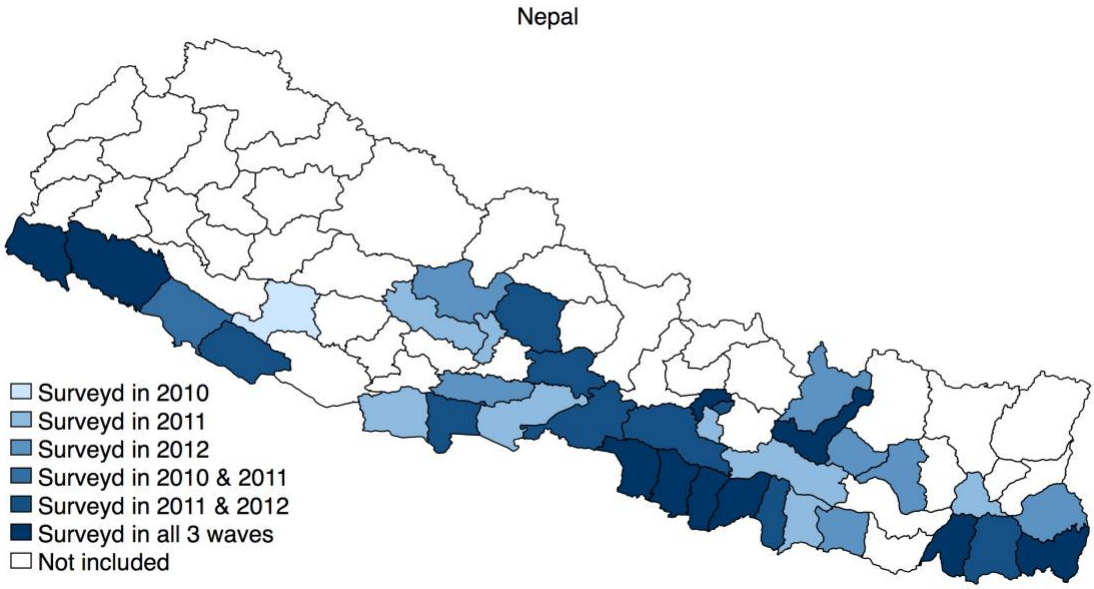


FIGURE 3: PROBABILITY OF PROGRAM PARTICIPATION

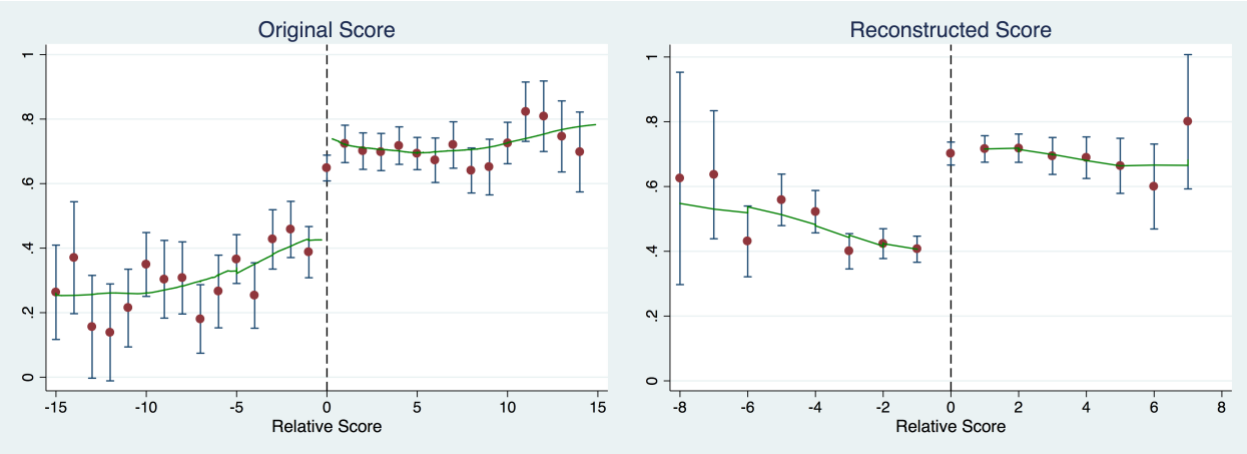


FIGURE 4: DENSITY OF FORCING VARIABLE

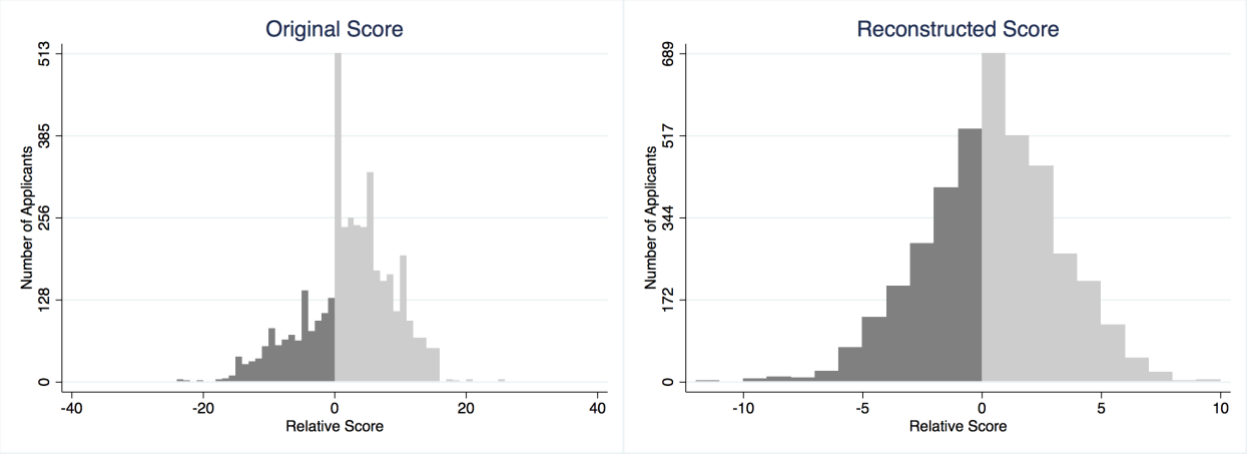


FIGURE 5: CONTINUITY OF COVARIATES AROUND THE CUT-OFF AT BASELINE

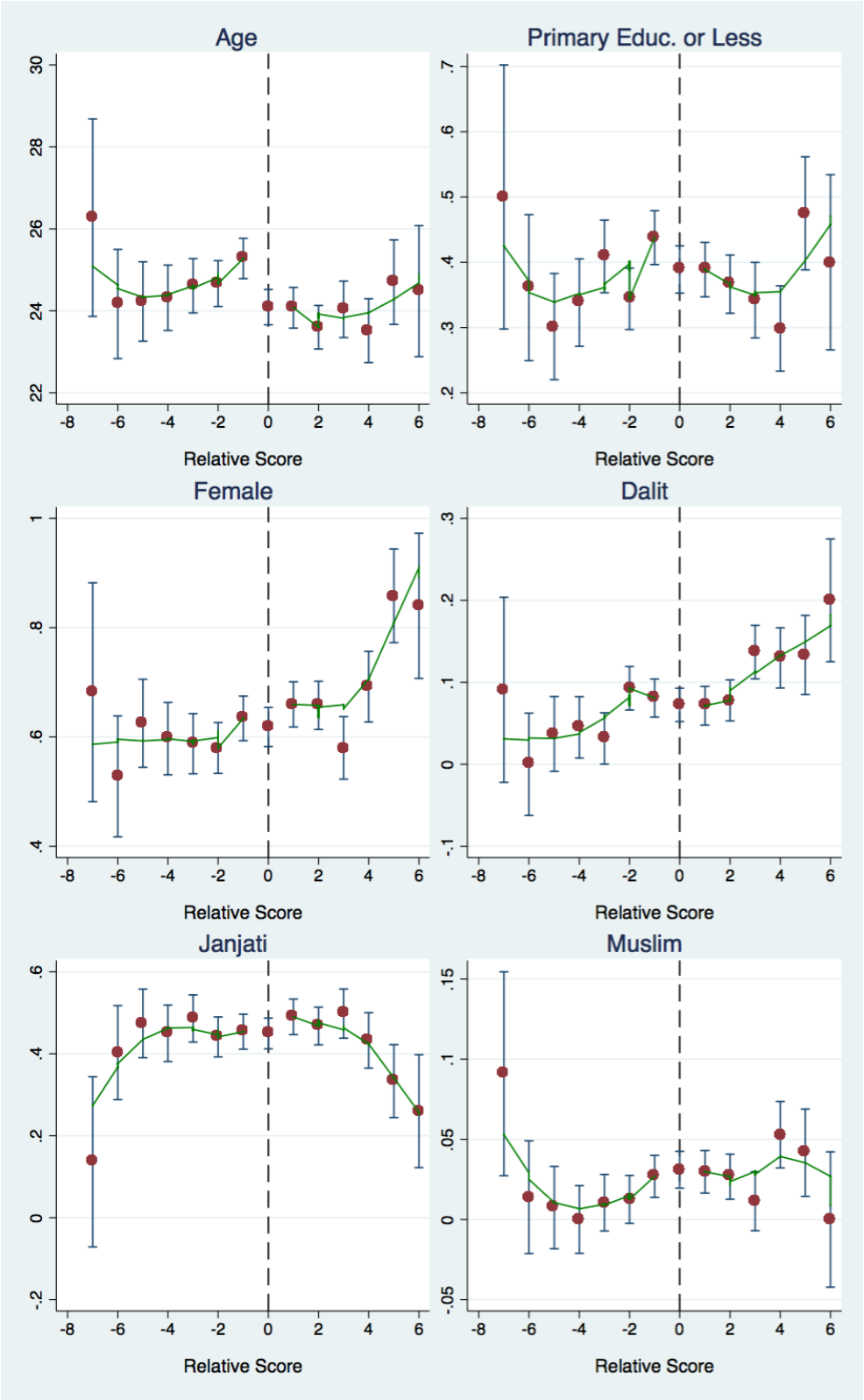


FIGURE 6: CONTINUITY OF OUTCOMES AROUND THE CUT-OFF AT BASELINE

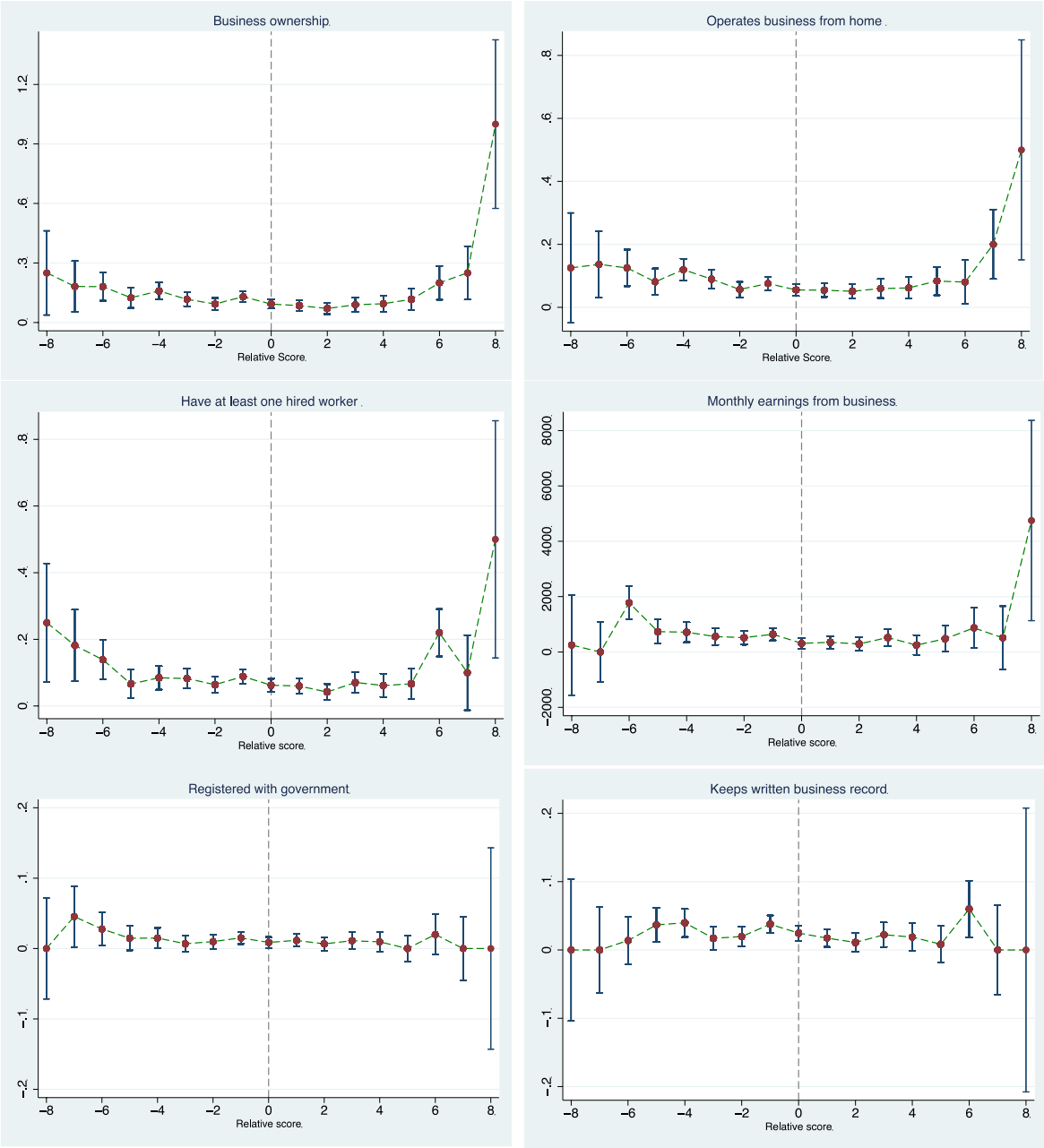


FIGURE 6 (CONT.): CONTINUITY OF OUTCOMES AROUND THE CUT-OFF AT BASELINE

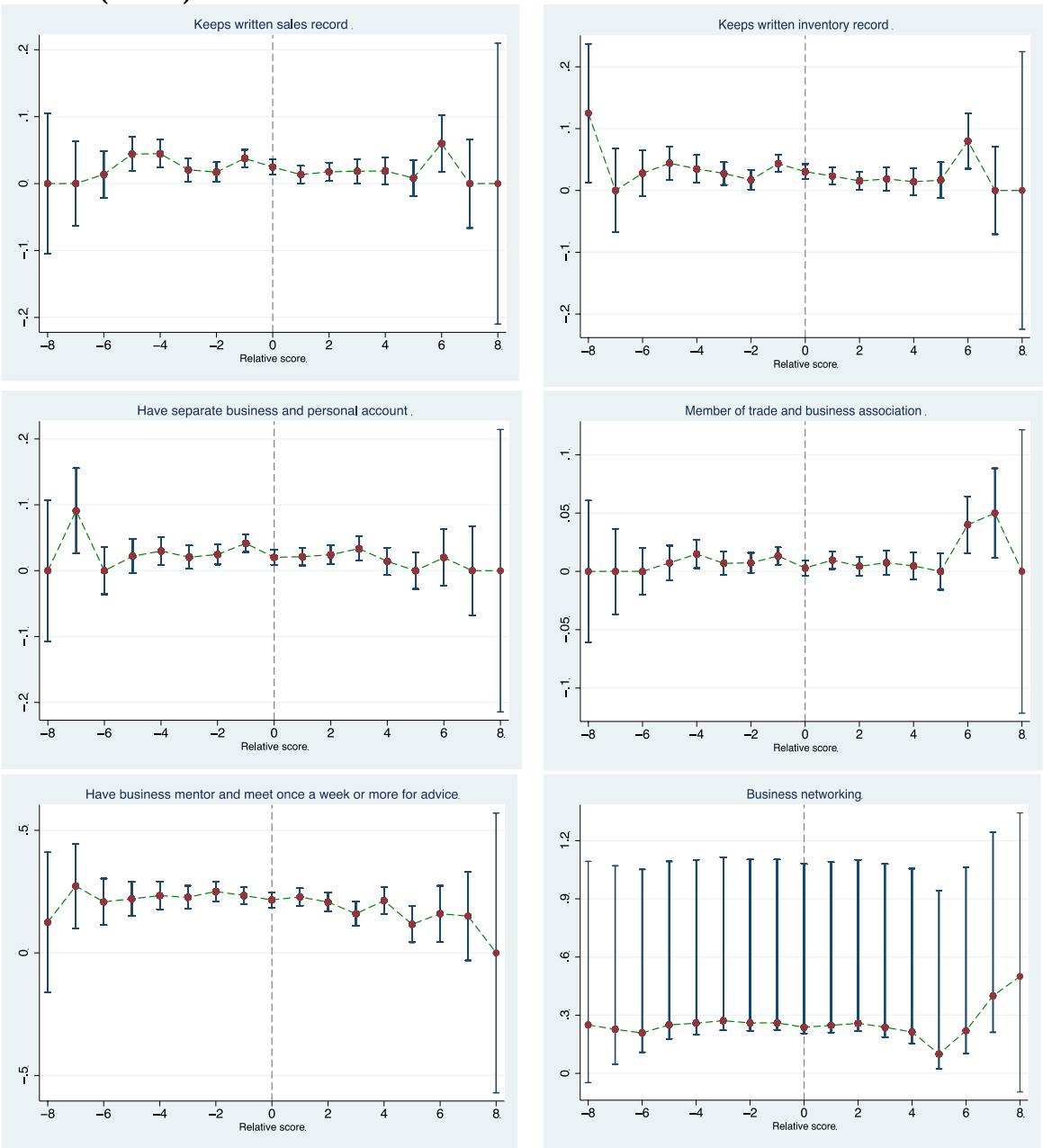


FIGURE 7: CONTINUITY OF BUSINESS SECTORS AROUND THE CUT-OFF

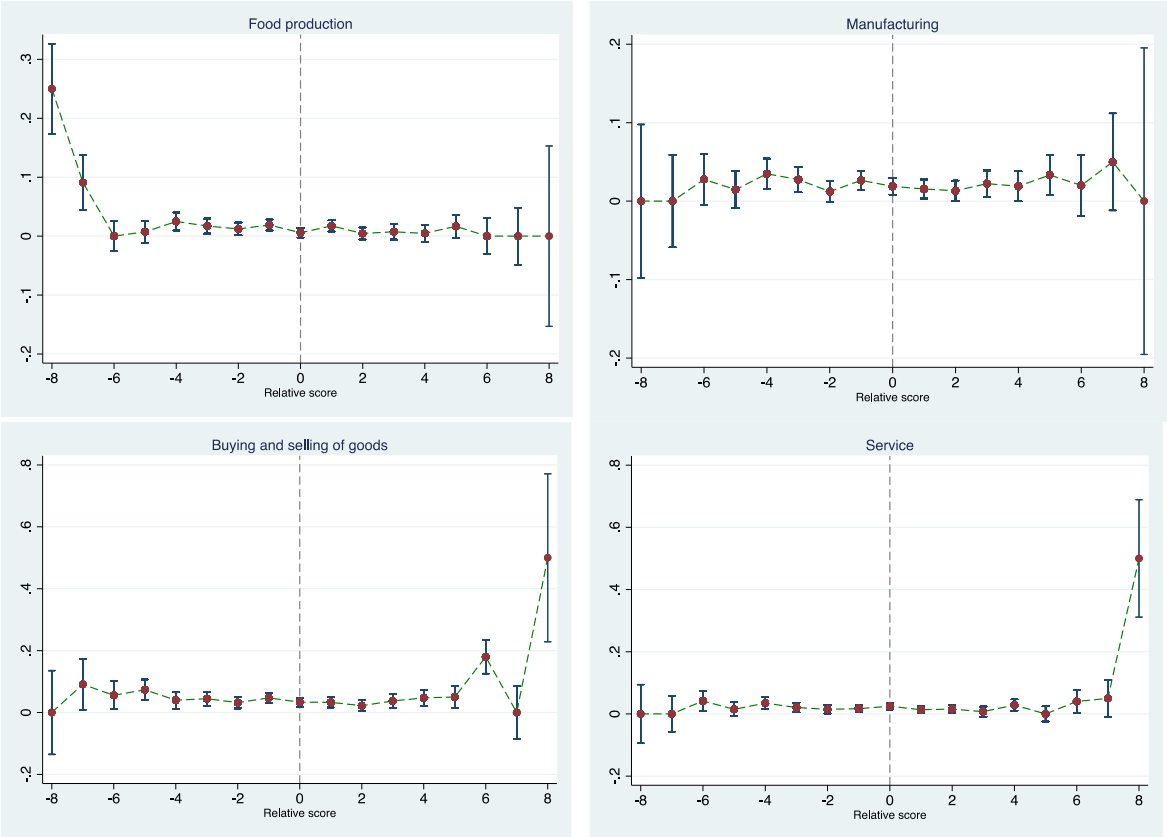
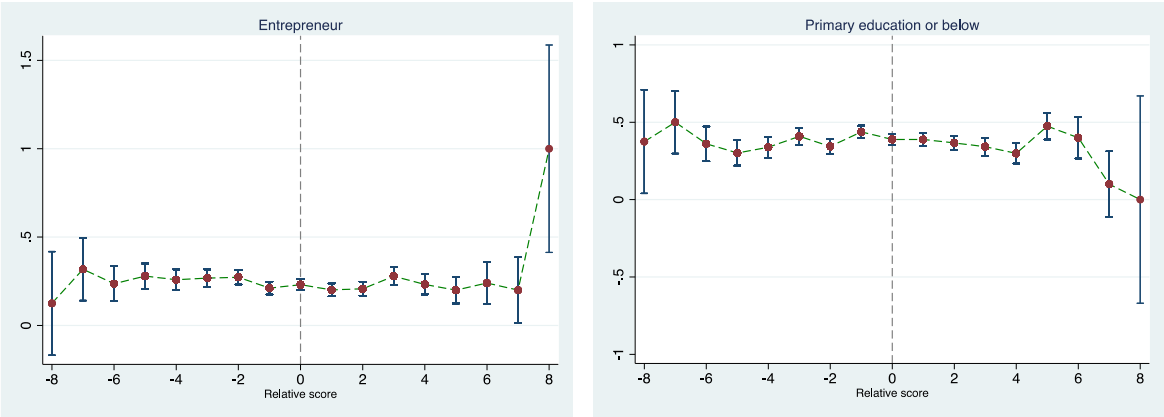


FIGURE 8: CONTINUITY OF SUB-GROUPS USED FOR HETEROGENEITY ANALYSIS AT BASELINE



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.

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	
Panel A: 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)
Entrepreneurs (1=Yes)	0.23	0.18	0.36	0.24	0.23	0.27	0.22	0.01	(0.08)	0.01	(0.02)	0.02	(0.07)	-0.01	(0.02)
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)
Panel B: Outcome Variables															
Business ownership[a] (1=yes)	0.11	0.12	0.09	0.11	0.11	0.11	0.11	0.00	(0.06)	0.00	(0.02)	-0.09	(0.07)	-0.04	(0.03)
Operates business from home[b] (1=yes)	0.55	0.64	0.36	0.53	0.56	0.58	0.52	0.44*	(0.23)	0.12	(0.08)	0.02	(0.27)	0.01	(0.07)
Registered with government[c] (1=yes)	0.09	0.04	0.17	0.09	0.08	0.09	0.08	-0.01	(0.13)	-0.02	(0.05)	0.04	(0.14)	0.01	(0.04)
Keeps written business expense record (1=yes)	0.18	0.13	0.28	0.18	0.18	0.20	0.17	0.03	(0.18)	0.00	(0.07)	-0.18	(0.29)	-0.06	(0.06)
Keeps written sales record (1=yes)	0.18	0.12	0.30	0.20	0.17	0.21	0.17	0.11	(0.18)	0.02	(0.07)	-0.15	(0.27)	-0.05	(0.06)
Keeps written inventory record (1=yes)	0.21	0.15	0.33	0.22	0.21	0.23	0.20	0.11	(0.19)	0.03	(0.07)	-0.04	(0.34)	-0.01	(0.06)
Separate business and personal account (1=yes)	0.13	0.11	0.17	0.13	0.13	0.15	0.12	-0.06	(0.16)	-0.01	(0.06)	0.02	(0.23)	0.01	(0.06)
Member of trade and business association (1=yes)	0.06	0.03	0.11	0.08	0.05	0.07	0.06	-0.15	(0.12)	-0.04	(0.05)	-0.14	(0.11)	-0.05	(0.04)
Business networking[d] (1=yes)	0.40	0.34	0.52	0.40	0.40	0.42	0.38	0.15	(0.22)	0.07	(0.08)	-0.23	(0.25)	-0.08	(0.07)
Have at least one employee (1=yes)	0.57	0.51	0.68	0.59	0.55	0.58	0.55	0.33	(0.23)	0.13	(0.09)	-0.07	(0.22)	-0.02	(0.07)
Monthly earnings from business[e] (in NRs)	3,829.63	29,93.05	5,478.74	4,393.30	3,458.36	4,562.94	3,149.39	-1,221.60	(2,786.75)	-549.84	(1,006.12)	-2,722.73	(3,020.95)	-928.79	(979.68)

Notes: [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is being operated from home. [c]Takes a value of 1 if the business is registered with government or is under the process of registration. [d]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [e]Respondent's total individual earnings from this activity during the past month. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 4: BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES

	Business Ownership and Characteristics		Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	Operate business from home (1=yes)	Registered with Govern -ment (1= yes)[b]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventor y record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business network -ing[c]	Have at least one hired worker (1=yes)	Monthly earnings from business[d] (in NRs)
Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
LATE, DD	0.23*	-0.58	-0.18	0.03	0.26	0.28	0.01	-0.24	0.52	-0.27	6,919.43
(Original Score)	(0.12)	(0.39)	(0.18)	(0.30)	(0.29)	(0.30)	(0.26)	(0.18)	(0.43)	(0.36)	(4,544.59)
First Stage F-statistic	27.83	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00
Baseline mean	0.11	0.55	0.09	0.18	0.18	0.21	0.13	0.06	0.40	0.57	3,829.63
Bandwidth	10	10	10	10	10	10	10	10	10	10	10
Observations	3504	864	864	864	864	864	864	919	919	919	919
Panel B											
LATE, DD	0.30***	0.31	0.05	0.40	0.52*	0.32	0.38*	0.31*	0.37	-0.28	-3,785.89
(Reconstr. Sc.)	(0.10)	(0.22)	(0.13)	(0.26)	(0.29)	(0.22)	(0.22)	(0.16)	(0.27)	(0.23)	(3,841.72)
First Stage F-statistic	27.84	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70
Baseline mean	0.11	0.55	0.09	0.18	0.18	0.21	0.13	0.06	0.40	0.57	3,829.63
Bandwidth	7	7	7	7	7	7	7	7	7	7	7
Observations	3968	994	994	994	994	994	994	994	994	994	994
Panel C											
LATE, Level	0.32***	0.48**	0.16	0.29	0.43*	0.18	0.30	0.30*	0.27	-0.12	-3,785.89
(Reconstr. Sc.)	(0.11)	(0.22)	(0.12)	(0.21)	(0.25)	(0.18)	(0.20)	(0.16)	(0.19)	(0.20)	(3,841.72)
First Stage F-statistic	27.87	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70
Baseline mean	0.11	0.55	0.09	0.18	0.18	0.21	0.13	0.06	0.40	0.57	3,829.63
Bandwidth	7	7	7	7	7	7	7	7	7	7	7
Observations	3977	994	994	994	994	994	994	994	994	994	994

Notes: : [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is registered with government or is under the process of registration. [c]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [d]Respondent's total individual earnings from this activity during the past month. 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. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 4 (CONT.): BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES

	Business Ownership and Characteristics		Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	Operate business from home (1=yes)	Registered with Govern -ment (1= yes)[b]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventory record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business network -ing[c]	Have at least one hired worker (1=yes)	Monthly earnings from business[d] (in NRs)
Panel D	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
ITT, DD	0.08***	0.10	0.01	0.07	0.10*	0.08	0.08	0.07**	0.17**	-0.10	-1,486.61
(Reconst. Sc.)	(0.03)	(0.06)	(0.04)	(0.06)	(0.06)	(0.06)	(0.05)	(0.03)	(0.08)	(0.07)	(1,188.92)
Group means at follow-up (Within given Bandwidth of Panel B)											
Non-treated	0.17	0.40	0.14	0.28	0.30	0.30	0.22	0.06	0.52	0.59	4,585.39
Treated	0.25	0.51	0.12	0.36	0.38	0.38	0.28	0.07	0.58	0.58	6,096.31
Non-assigned	0.19	0.46	0.12	0.28	0.30	0.31	0.23	0.07	0.54	0.59	6,110.56
Assigned	0.23	0.48	0.13	0.36	0.38	0.38	0.28	0.07	0.57	0.58	5,218.91

Notes: : [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is registered with government or is under the process of registration. [c]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [d]Respondent's total individual earnings from this activity during the past month. Standard errors adjusted for clustering at event level reported in parentheses. Each cell in Panel D 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. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 5: BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES BY GENDER

	Business Ownership and Characteristics		Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	Operate business from home (1=yes)	Registered with Govern -ment (1= yes)[b]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventory record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business network -ing[c]	Have at least one hired worker (1=yes)	Monthly earnings from business[d] (in NRs)
Panel A: HLATE, DD (Reconstructed Score)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Women	0.46***	0.68***	-0.09	0.33**	0.39***	0.46***	0.04	0.10	0.38*	-0.13	-3,000.17
	(0.09)	(0.20)	(0.08)	(0.15)	(0.15)	(0.16)	(0.15)	(0.07)	(0.21)	(0.16)	(2,540.94)
Men	0.11	0.19	0.2	0.24	0.03	0.16	0.05*	0.20**	0.10	0.30*	5474.77
	(0.09)	(0.15)	(0.13)	(0.17)	(0.17)	(0.17)	(0.15)	(0.09)	(0.23)	(0.18)	(4,218.50)
Difference	0.32***	0.44**	-0.11	0.09**	0.36*	0.30**	-0.01	-0.10	0.37	-0.09	-3,840.90
	(0.11)	(0.21)	(0.12)	(0.20)	(0.20)	(0.21)	(0.18)	(0.11)	(0.25)	(0.21)	(3,680.53)
Bandwidth	7	7	7	7	7	7	7	7	7	7	7
Observations	3,968	994	994	994	994	994	994	994	994	994	994
Group means at Baseline (Within a given Bandwidth)											
Women	0.12	0.64	0.04	0.13	0.13	0.15	0.11	0.03	0.34	0.52	3,044.76
Men	0.09	0.36	0.18	0.29	0.32	0.35	0.18	0.12	0.52	0.68	5,493.33
Panel B: Heterogeneous ITT, DD (Reconstructed Score)											
Women	0.11***	0.07	-0.03	0.08	0.08	0.10	0.02	0.08**	0.16*	-0.16**	-2,381.25**
	(0.03)	(0.08)	(0.03)	(0.07)	(0.07)	(0.07)	(0.07)	(0.03)	(0.09)	(0.08)	(1,098.52)
Men	0.03	0.13	0.04	0.04	0.03	0.01	0.01	0.01	0.02	0.08	3,079.21
	(0.04)	(0.10)	(0.09)	(0.12)	(0.13)	(0.13)	(0.10)	(0.07)	(0.17)	(0.13)	(2,362.50)
Difference	0.06**	0.08	-0.01	0.08	0.05	0.09	-0.03	0.07	0.14	-0.01	-2,268.64
	(0.03)	(0.07)	(0.06)	(0.08)	(0.08)	(0.08)	(0.07)	(0.05)	(0.10)	(0.09)	(1,767.11)

Notes: [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is registered with government or is under the process of registration. [c]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [d]Respondent's total individual earnings from this activity during the past month. 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. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 6: BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES BY TYPE OF ENTREPRENEURS

	Business Ownership and Characteristics		Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	Operate business from home (1=yes)	Registered with Govern -ment (1= yes)[b]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventory record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business network -ing[c]	Have at least one hired worker (1=yes)	Monthly earnings from business[d] (in NRs)
Panel A: HLATE, DD (Reconstructed Score)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Entrepreneurs	0.30**	0.31	0.04	0.19	0.47*	0.45**	0.50**	0.17	-0.02	-0.17	-790.62
	(0.14)	(0.22)	(0.20)	(0.25)	(0.25)	(0.25)	(0.25)	(0.19)	(0.30)	(0.22)	(3,580.64)
Subsistence	0.28***	0.57***	-0.01	0.44***	0.41***	0.45***	0.12	0.05	0.32*	-0.07	-14.92
	(0.07)	(0.15)	(0.06)	(0.13)	(0.13)	(0.14)	(0.12)	(0.06)	(0.17)	(0.14)	(2,152.57)
Difference	0.02	0.37	0.14	-0.13	0.07	0.01	0.29	0.20	-0.26	-0.24	-5,427.17
	(0.14)	(0.27)	(0.16)	(0.23)	(0.23)	(0.23)	(0.21)	(0.13)	(0.30)	(0.26)	(5,164.76)
Bandwidth	7	7	7	7	7	7	7	7	7	7	7
Observations	3,968	728	728	728	728	728	728	728	728	728	728
Group means at Baseline (Within a given Bandwidth)											
Entrepreneurs	0.13	0.46	0.17	0.35	0.38	0.39	0.28	0.08	0.41	0.69	5073.10
Subsistence	0.10	0.58	0.05	0.12	0.11	0.15	0.08	0.05	0.39	0.52	3357.88
Panel B: Heterogeneous ITT, DD (Reconstructed Score)											
Entrepreneurs	0.08	0.23*	-0.03	0.04	0.18	0.02	0.14	0.10	-0.06	-0.08	-4,933.59**
	(0.06)	(0.12)	(0.10)	(0.13)	(0.12)	(0.12)	(0.10)	(0.07)	(0.16)	(0.12)	(2,413.54)
Subsistence	0.06***	0.05	0.01	0.09	0.16	0.01	0.05	0.07**	0.27***	-0.10	-156.47
	(0.03)	(0.07)	(0.03)	(0.07)	(0.07)	(0.07)	(0.06)	(0.03)	(0.09)	(0.08)	(1,322.74)
Difference	0.02	0.18	0.04	-0.05	0.02	0.01	0.08	0.04	-0.11	-0.09	-1,509.09
	(0.04)	(0.08)	(0.06)	(0.08)	(0.08)	(0.08)	(0.07)	(0.05)	(0.10)	(0.09)	(1,740.95)

Notes. [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is registered with government or is under the process of registration. [c]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [d]Respondent's total individual earnings from this activity during the past month. 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. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 7: BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES BY EDUCATION AND BASELINE BUSINESS SKILL

	Business Ownership and Characteristics		Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	Operate business from home (1=yes)	Registered with Govern -ment (1= yes)[b]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventory record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business network -ing[c]	Have at least one hired worker (1=yes)	Monthly earnings from business[d] (in NRs)
Business Owner's Education	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: HLATE, DD (Reconstructed Score)											
Primary or Below-Above Primary	0.08	0.14	0.11	-0.09	-0.27	-0.19	-0.01	-0.04	0.05	0.31	2935.06
	(0.13)	(0.25)	(0.12)	(0.19)	(0.19)	(0.20)	(0.18)	(0.11)	(0.28)	(0.23)	(3,893.90)
Panel B: Heterogeneous ITT, DD (Reconstructed Score)											
Primary or Below-Above Primary	-0.04	0.05	-0.03	-0.06	-0.13*	-0.09	-0.05	-0.04	-0.01	-0.09	-313.18
	(0.03)	(0.08)	(0.04)	(0.07)	(0.07)	(0.07)	(0.06)	(0.04)	(0.09)	(0.08)	(1,310.99)
Baseline Business Skill											
Panel C: HLATE, DD (Reconstructed Score)											
Skilled-Low Skilled	-0.20*	0.17	-0.07	0.00	0.25	0.11	0.19	-0.09	0.19	-0.15	-2,702.41
	(0.12)	(0.23)	(0.10)	(0.18)	(0.18)	(0.18)	(0.16)	(0.09)	(0.25)	(0.20)	(3,491.08)
Panel D: Heterogeneous ITT, DD (Reconstructed Score)											
Skilled-Low Skilled	-0.03	0.07	-0.01	0.01	0.04	-0.01	0.04	-0.08**	0.05	-0.05	-1,026.16
	(0.03)	(0.07)	(0.04)	(0.07)	(0.07)	(0.07)	(0.06)	(0.04)	(0.09)	(0.08)	(1,491.65)

Notes. [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is registered with government or is under the process of registration. [c]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [d]Respondent's total individual earnings from this activity during the past month. Standard errors adjusted for clustering at event level reported in parentheses. Panel A and C present second stage results. In all four panels, each cell represents coefficient of the interaction term with the group variable 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. For comparability, all estimates are conducted within the same bandwidths as in Panel B, Table 4. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE 8: BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES BY BASELINE BUSINESS SECTOR

	Business Ownership and Characteristics		Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	Operate business from home (1=yes)	Registered with Govern-ment (1= yes)[b]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventory record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business network -ing[c]	Have at least one hired worker (1=yes)	Monthly earnings from business[d] (in NRs)
Panel A: HLATE, DD (Reconstructed Score)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Food Producer	0.65**	0.53	-0.23	0.22	-0.07	0.21	0.16	-0.14	0.64	0.79*	11,315.21*
	(0.27)	(0.42)	(0.22)	(0.36)	(0.48)	(0.36)	(0.37)	(0.14)	(0.49)	(0.43)	(6,490.80)
Manufacturing	0.32	-1.23**	0.09	-0.36	-0.11	-0.56*	0.50	0.15	-0.65*	-1.02*	-2,180.55
	(0.21)	(0.53)	(0.10)	(0.35)	(0.42)	(0.31)	(0.46)	(0.14)	(0.34)	(0.60)	(4,512.07)
Buying and selling of goods	-0.05	-0.10	0.25	-0.18	-0.20	0.17	-0.36	0.59*	-0.68	0.64*	15,715.17**
	(0.37)	(0.35)	(0.27)	(0.44)	(0.36)	(0.41)	(0.42)	(0.34)	(0.45)	(0.37)	(7,390.20)
Service	0.65**	0.61*	1.04	0.70	1.01**	1.10*	0.43	0.83	1.50	0.80	3,791.12
	(0.28)	(0.36)	(0.76)	(0.51)	(0.49)	(0.64)	(0.39)	(0.61)	(1.09)	(0.65)	(9,789.70)
Panel B: Heterogeneous ITT, DD (Reconstructed Score)											
Food Producer	0.21	0.31	-0.66*	0.06	-0.16	-0.41	-0.45	0.20	1.11**	0.50	-13,145.12
	(0.31)	(0.51)	(0.29)	(0.42)	(0.58)	(0.54)	(0.51)	(0.25)	(0.42)	(0.51)	(26,508.51)
Manufacturing	0.31	0.13	0.22	0.35	0.44	0.08	0.36	0.08	0.00	-0.30	-5,009.93
	(0.21)	(0.42)	(0.18)	(0.44)	(0.50)	(0.54)	(0.23)	(0.28)	(0.56)	(0.36)	(10,690.87)
Buying and selling of goods	0.20	-0.20	-0.22	-0.07	0.22	0.29	0.18	0.20	-0.23	-0.25	4,384.52
	(0.15)	(0.20)	(0.19)	(0.28)	(0.21)	(0.24)	(0.22)	(0.18)	(0.24)	(0.19)	(5,187.07)
Service	0.06	0.49**	-0.1	0.52*	0.58**	0.72**	-0.12	0.12	0.74	0.01	70.75
	(0.22)	(0.24)	(0.17)	(0.28)	(0.27)	(0.33)	(0.29)	(0.26)	(0.61)	(0.52)	(8,643.42)

Notes. [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is registered with government or is under the process of registration. [c]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [d]Respondent's total individual earnings from this activity during the past month. 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. ***, **, and * denote significant at the 1, 5, and 10 percent level.

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	
District:	
Municipality/ VDC:	
Tole:	Ward No:
Phone Number (Land line):	
Phone Number (Mobile):	
Father's Name:	
Mother's Name:	
Citizenship No.:	
Issued District:	

Current Address	
District:	
Municipality/ VDC:	
Tole:	Ward No:
Phone Number (Land line):	
Phone Number (Mobile):	

In case of getting information of you:

Reference person:
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

Training Information

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	Entry Requirement (Y/N)			Selection Criteria (Individual Scores)					Final Score	Rank
						1-4. Short-listing (70%)				5. Interview (30%)		
			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)		**TOTAL** SCORE (100) *List candidates from highest to lowest*	
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 percent. The shaded area represents candidates who would have been sampled for the baseline survey based on this example.

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 to 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. 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	-0.22***		
	(0.02)		
Sub-Score 1 (Trade-specif. Edu.)		0.12*	2.75***
		(0.06)	(0.43)
Sub-Score 2 (Econ. Poverty)		-0.18***	1.05*
		(0.03)	(0.55)
Sub-Score 3 (Social Caste)		-0.35***	1.01***
		(0.03)	(0.35)
Sub-Score 4 (Development Status of Dist. Of Origin)		-0.47***	3.56***
		(0.09)	(1.35)
SS 1 x SS 2			-0.12***
			(0.04)
SS 1 x SS 3			-0.14***
			(0.03)
SS 1 x SS 4			-0.38***
			(0.10)
SS 2 x SS 3			-0.05
			(0.03)
SS 2 x SS 4			-0.19**
			(0.09)
SS 3 x SS 4			-0.28***
			(0.09)
SS 1 x SS 2 x SS 3			0.01***
			(0.01)
SS 1 x SS 2 x SS 4			0.02***
			(0.01)
SS 1 x SS 3 x SS 4			0.03***
			(0.01)
SS 2 x SS 3 x SS 4			0.01**
			(0.01)
SS 1 x SS 2 x SS 3 x SS 4			-0.00***
			(0.00)
Observations	4,090	4,090	4,090
Adj. R2	0.38	0.39	0.40

Notes: All models include event dummies. ***, **, and * denote significance 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)
Observations	4,585	4,585	4,585	4,585	1,547	1,547	1,547	1,547
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 significance at the 1, 5, and 10 percent level, respectively.

TABLE A5: BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES OF BASELINE BUSINESS-OWNER

	Business Ownership and Characteristics			Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	New Business (1=yes)[b]	Operate business from home (1=yes)	Registered with Govern ment (1= yes)[c]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventory record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business networ -king[d]	Have at least one hired worker (1=yes)	Monthly earnings from business[e] (in NRs)
Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
LATE, DD	0.70**	0.85*	0.32	0.00	0.33	0.55*	0.22	0.29	0.31*	0.27	0.20	4,369.05
(Reconstr. Sc.)	(0.30)	(0.49)	(0.32)	(0.20)	(0.29)	(0.30)	(0.32)	(0.32)	(0.18)	(0.71)	(0.38)	(6,015.72)
First Stage F-statistic	10.91	10.91	10.91	10.91	10.91	10.91	10.91	10.91	10.91	10.91	10.91	10.91
Baseline mean	1.00	0.43	0.62	0.09	0.20	0.20	0.23	0.15	0.07	0.38	0.62	4551.53
Bandwidth	7	7	7	7	7	7	7	7	7	7	7	7
Observations	418	418	418	418	418	418	418	418	418	418	418	418
Panel B												
ITT, DD	0.20**	0.22**	0.10	0.01	0.10	0.16*	0.05	0.07	0.10*	0.02	0.08	1,590.32
(Reconst. Sc.)	(0.08)	(0.11)	(0.10)	(0.06)	(0.09)	(0.09)	(0.09)	(0.08)	(0.05)	(0.12)	(0.10)	(1,729.99)

Notes: : [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is being operated from no more than last twelve months (established after the intervention). [c]Takes a value of 1 if the business is registered with government or is under the process of registration. [d]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [e]Respondent's total individual earnings from this activity during the past month. 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. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE A6: BANDWIDTH SENSIVITY (BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES)

	Business Ownership and Characteristics		Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	Operate business from home (1=yes)	Registered with Govern -ment (1= yes)[b]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventory record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business network -ing[c]	Have at least one hired worker (1=yes)	Monthly earnings from business[d] (in NRs)
Panel A: LATE, DD (Reconstructed Score)	(1)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
4 Index Scores	0.18*	0.22	0.11	0.13	0.14	0.09	0.27	0.35**	0.17	-0.07	270.93
	(0.11)	(0.22)	(0.13)	(0.22)	(0.21)	(0.21)	(0.19)	(0.14)	(0.28)	(0.25)	(4,602.13)
Observations	3549	861	861	861	861	861	861	861	861	861	861
F-statistic	27.22	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20
6 Index Scores	0.25***	0.41	0.15	0.38	0.39	0.3	0.36	0.27*	0.35	-0.22	-4,308.95
	(0.09)	(0.32)	(0.18)	(0.31)	(0.27)	(0.25)	(0.25)	(0.17)	(0.26)	(0.23)	(4,175.77)
Observations	3926	973	973	973	973	973	973	973	973	973	973
F-statistic	27.53	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
7 Index Scores	0.30***	0.31	0.05	0.4	0.52*	0.32	0.38*	0.31*	0.37	-0.28	-3,785.89
	(0.10)	(0.22)	(0.13)	(0.26)	(0.29)	(0.22)	(0.22)	(0.16)	(0.27)	(0.23)	(3,841.72)
Observations	3968	994	994	994	994	994	994	994	994	994	994
F-statistic	27.84	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70	10.70
8 Index Scores	0.29***	0.28	0.04	0.35	0.51*	0.28	0.35*	0.28**	0.28	-0.25	-3,762.16
	(0.10)	(0.21)	(0.12)	(0.22)	(0.26)	(0.19)	(0.21)	(0.14)	(0.27)	(0.21)	(3,675.50)
Observations	3978	997	997	997	997	997	997	997	997	997	997
F-statistic	27.78	12.30	12.30	12.30	12.30	12.30	12.30	12.30	12.30	12.30	12.30
10 Index Scores	0.21***	0.24	0.03	0.26	0.41**	0.23	0.32**	0.24**	0.3	-0.26	-3,796.47
	(0.08)	(0.18)	(0.10)	(0.17)	(0.20)	(0.16)	(0.15)	(0.11)	(0.23)	(0.19)	(3,335.01)
Observations	4001	1004	1004	1004	1004	1004	1004	1004	1004	1004	1004
F-statistic	28.09	12.60	12.60	12.60	12.60	12.60	12.60	12.60	12.60	12.60	12.60

Notes. [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is registered with government or is under the process of registration. [c]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [d]Respondent's total individual earnings from this activity during the past month. Standard errors adjusted for clustering at event level reported in parentheses. Panel A and C present second stage results. In all four panels, each cell represents coefficient of the interaction term with the group variable 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. For comparability, all estimates are conducted within the same bandwidths as in Panel B, Table 4. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE A6 (CONT.): BANDWIDTH SENSIVITY (BUSINESS CHARACTERISTICS AND BUSINESS PRACTICES)

	Business Ownership and Characteristics		Business Practices					Business Network		Business Performance	
	Business ownership (1=yes)[a]	Operate business from home (1=yes)	Registered with Govern -ment (1= yes)[b]	Keeps written business expense record (1=yes)	Keeps written sales record (1=yes)	Keeps written inventory record (1=yes)	Have separate business and personal account (1=yes)	Member of trade and business association (1=yes)	Business network -ing[c]	Have at least one hired worker (1=yes)	Monthly earnings from business[d] (in NRs)
Panel B: ITT, DD (Reconstructed Score)	(1)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
4 Index Scores	0.05*	0.07	0.02	0.02	0.04	0.02	0.08	0.11**	0.11	-0.06	-517.49
	(0.03)	(0.08)	(0.04)	(0.07)	(0.07)	(0.08)	(0.06)	(0.04)	(0.09)	(0.08)	(1,563.00)
6 Index Scores	0.07***	0.09	0.01	0.07	0.10	0.07	0.09	0.07**	0.14*	-0.08	-1,564.99
	(0.03)	(0.06)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)	(0.03)	(0.08)	(0.07)	(1,259.00)
7 Index Scores	0.08***	0.10	0.00	0.07	0.10*	0.08	0.08	0.07**	0.17**	-0.1	-1,486.61
	(0.03)	(0.06)	(0.04)	(0.06)	(0.06)	(0.06)	(0.05)	(0.03)	(0.08)	(0.07)	(1,188.92)
8 Index Scores	0.08***	0.10	0.00	0.06	0.10*	0.07	0.07	0.06**	0.15*	-0.09	-1,478.53
	(0.03)	(0.06)	(0.04)	(0.06)	(0.06)	(0.06)	(0.05)	(0.03)	(0.08)	(0.07)	(1,177.84)
10 Index Scores	0.06***	0.11*	0.00	0.07	0.10*	0.07	0.09*	0.06**	0.14*	-0.09	-1,478.53
	(0.02)	(0.06)	(0.03)	(0.06)	(0.06)	(0.06)	(0.05)	(0.03)	(0.08)	(0.06)	(1,158.38)

Notes. [a]Captured by survey responses to the question of whether they currently own a business or not. [b]Takes a value of 1 if the business is registered with government or is under the process of registration. [c]Takes a value of 1 if the individuals meet and discuss about their businesses and work-related things with people within the community once a week or more. [d]Respondent's total individual earnings from this activity during the past month. Standard errors adjusted for clustering at event level reported in parentheses. Panel A and C present second stage results. In all four panels, each cell represents coefficient of the interaction term with the group variable 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. For comparability, all estimates are conducted within the same bandwidths as in Panel B, Table 4. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE A7: LOANS AND SAVINGS

	Loans					Savings				
	Total Loans (in NRs)	Loan from Bank	Loan from Credit & Savings Groups	Loan from Money Lenders	Loan from Friends & Family (1=yes)	Total Savings (in NRs)	Savings at Bank (1=yes)	Savings at Credit & Savings Group (1=yes)	Savings at Coopera -tives (1=yes)	Savings at Home (1=yes)
Panel A	(1)	(2)	(3)			(5)	(6)	(7)	(8)	(9)
LATE, DD (Reconst. Sc.)	3,805.92	20,515.19	-6,824.80	-13,498.41	3,613.94	251.29	-4,891.85	2,661.11	8,022.56	344.59
	(33,442.21)	(13,389.28)	(21,863.61)	(13,629.77)	(10,534.38)	(62,578.77)	(46,014.87)	(4,116.81)	(21,127.55)	(2,936.46)
Baseline Mean	16,404.93	3,997.04	3,508.90	3,266.54	5,632.46	10,067.18	2,855.12	2,417.97	1,583.70	1,442.12
Panel B: **By Gender**										
HLATE, DD (Reconstructed Score)										
Women-men	-39,939.00	10,159.15	-11,546.44	-10,463.90	-28,087.81**	51,919.18	42,010.67*	1,110.98	2,888.87	1,038.38
	(25,116.64)	(17,366.93)	(10,747.29)	(11,024.81)	(11,019.32)	(31,818.50)	(22,783.59)	(2,968.99)	(8,540.34)	(2,257.70)
Group means at Baseline										
Women	14,876.82	3,093.21	3,921.28	2,506.41	5,355.92	7,501.97	1,558.59	2,446.56	771.67	1,314.94
Men	19,417.24	5,778.74	2,695.98	4,764.94	6,177.59	15,123.91	5,410.92	2,361.61	3,184.43	1,692.82
Panel C: **By Entrepreneurship**										
HLATE, DD (Reconstructed Score)										
Entrepreneur-Subsistence	39,620.81	-13,730.62	-180.70	14,537.64	38,994.49***	-69,950.11	-57,477.13	-3,799.25	-7,477.62	257.70
	(33,485.56)	(18,551.11)	(19,838.91)	(13,539.95)	(13,237.42)	(56,359.62)	(41,681.72)	(3,634.14)	(7,319.38)	(1,108.68)
Group means at Baseline										
Entrepreneur	20,710.60	4,372.85	4,106.62	5,747.68	6,483.44	14,783.71	6,222.52	2,468.48	3,146.69	1,571.85
Subsistence	14,628.55	3,842.00	3,262.30	2,242.90	5,281.37	8,121.30	1,465.83	2,397.13	938.86	1,388.59

Notes: Panel A presents second stage results where 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. In Panel B and C, each cell represents coefficient of the interaction term with the group variable 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. For comparability, all estimates are conducted within the same bandwidths as in Panel B, Table 4. ***, **, and * denote significant at the 1, 5, and 10 percent level.

TABLE A8. PRINCIPAL COMPONENTS TO DEFINE TRANSFORMATIONAL ENTREPRENEURS. INITIAL FACTOR METHOD

EIGENVALUES OF THE CORRELATION MATRIX: TOTAL= 4 AVERAGE= 1

	Eigenvalue	Difference	Proportion	Cumulative
1	1.405	0.40	0.35	0.35
2	1.008	0.14	0.25	0.60
3	0.869	0.15	0.22	0.82
4	0.719			

FACTOR PATTERN

	PCA Component 1	PCA Component 2
Years of schooling	76*	-7
Financial literacy score	68*	2
Self-reported entrepreneurial skills	-12	97*
Raven's Score	60*	26

Notes: Rotation Method: Varimax. Robust standard errors. Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 are flagged by an "*".