Pre-analysis plan for an Impact Evaluation of the Skills Development for Employability Project in Republic of Congo

This version:
November 2021

Abstract

The authors are currently running a randomized field experiment testing the effectiveness of a youth employment program providing access to a vocational skills training to young men and women in the cities of Brazzaville and Pointe Noire (Republic of Congo). This pre-analysis plan provides an overview of the estimation strategy and of the variables to be constructed to implement it.

1. Introduction

This document outlines a pre-analysis plan for a youth employment impact evaluation. Through one component of this project, randomly selected subsets of applicants received the opportunity to participate in six- to nine-month long vocational training courses, followed by a 3-month long 'support to transition' period during which beneficiaries interned in their trade of choice. A subset of eligible candidates was not offered the opportunity to be trained as part of this program because of the limited capacities of the training providers. In addition, as fully described in another pre-analysis plan submitted in March 2019[1], applicants to the program were randomly assigned to receiving information on trade-specific earnings.

The document includes 1) a description of the study sample and the data to be generated as part of the experiment, 2) a list of variables to be considered in the analysis, with the detail of their construction, and 3) the authors' plan for the econometric analysis of the impact of receiving the vocational training.

At the time of writing this pre-analysis plan, the authors had access to the data from the baseline, midline and endline surveys for the first cohort, and the data from the baseline and midline surveys[2] for the second cohort.

[1]Registered under "Pre-analysis plan for a field experiment in the Republic of Congo: Addressing gender-based occupational segregation through the provision of trade-specific information on earnings" in Rouanet, L. and Gassier, M., 2016. Impact evaluation of the Skills Development for Employability Project in the Republic of Congo. Registry for International Development for Impact Evaluations (RIDIE). Available at: 10.23846/ridie096.

[2] Midline impacts were not studied at the time of writing this document.

2. Description of the sample

Application process

In 2015, the government of the Republic of Congo implemented a cohort-based vocational training program in the two largest cities of the countries, Brazzaville and Pointe-Noire. Two cohorts of the program have already been organized: the first cohort started in January 2016 and the second cohort started in February 2019. This impact evaluation will rely on the sample of the second cohort as much as possible, for purpose of comparability. Results on the first cohort will be presented as robustness checks. For outcomes with insufficient statistical power, the first and second cohorts may be pooled together to estimate specifications below.

For each cohort, youths were invited to apply to the training program and place two ordered choices. Among these applicants, only a portion were considered eligible. In order to be eligible, applicants must i) be aged 17 to 30, ii) be out of school for at least a year at the time of application, iii) have completed at least the third grade of the lower secondary education cycle without obtaining the upper secondary education certification and, iv) not be formally employed.

For the first cohort, 2,170 applications were received (1,436 in Brazzaville and 734 in Pointe-Noire), among which 1,477 were considered eligible. For the second cohort, 14,643 applications were collected (6,150 in Brazzaville and 8,493 in Pointe-Noire), of which 12,016 were eligible.

Sampling and randomization

Subsets of eligible applicants were randomly selected to participate in the training program (treatment group) and be surveyed for evaluation (survey sample). Given that the sampling and randomization process slightly vary between the two cohorts, we present both separately.

- First cohort

To constitute this sample, we started from the universe of applicants to the program and excluded people that did not meet the eligibility criteria, which gives 1,477 eligible applicants. We then grouped applicants by their first-choice track. We calculated the number (N) of spots available in each track. We then selected 3xN eligible applicants who opted for this trade as first choice for our study sample, among which two-thirds were women and one-third were men whenever possible, i.e when enough women selected this trade. In cases whether there were more than 3xN applicants for a track (or more precisely more than 2xN female applicants or N male applicants), we ranked individuals by the ‘fitness score’, assigned to them by training providers when they submitted their application, and allocate excess applicants (those with the lowest fitness scores) to their second choice instead, if it was a less popular one to ensure that all available spots in the program were filled.

When too few applicants chose a trade, as first or second choice, the control group could not be constituted. Because of this, some trades were excluded from the evaluation sample. Among trades included in the sample, we randomly selected 3 groups of applicants of equal sizes, which would become a treatment group and a twice larger control group. We stratified the randomization by city, gender, allocated trades and whether the applicant has dependent children. At the end of this process, a balancing test on all groups was conducted to check whether key variables showed statistical differences. In the final step, the three groups were presented to the country project management team, who publicly selected the treatment group.

As a result, for the first cohort, 404 individuals were invited to participate in a training (180 in Brazzaville and 224 in Pointe-Noire). The baseline survey sample included 880 individuals (439 in Brazzaville and 441 in Pointe-Noire).

- Second cohort

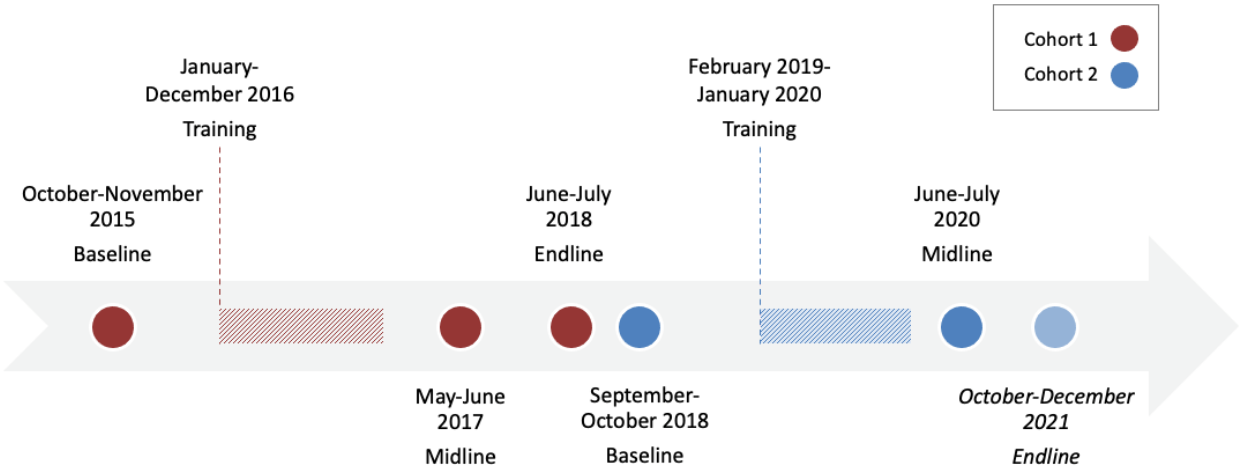
The randomization for the second cohort followed the same requirements as the first cohort, with two main changes.
First, we selected 2.5xN eligible applicants instead of 3xN, since applicants would have one less opportunity to apply to the program's subsequent cohorts.
Second, for the second cohort, we set up a new experiment that tests the effect of providing information on trade-specific earnings on trade choice, which implies new requirements in the randomization process. At the time of application, youths were assigned to one of two groups "on the spot": applicants who were to receive information[3] on trade-specific earnings before placing their applications; and applicants who were not. This experiment was conducted for all applicants in each application center. The application tools (tablets) in the centers were set to automatically allocate applicants into receiving information on earnings or not.
Given this setting, the randomization process of the survey sample for the second cohort included the information treatment as one additional randomization stratum in order to minimize imbalances within each trade between individuals who had received the information on trade-specific earnings and individuals who had not received it. This was done to maximize power when later measuring the effect of this information on employment and earnings.

As a result, 1,784 individuals were randomly selected to participate in the training, among which 777 were selected in Brazzaville and 1,007 in Pointe-Noire, and we identified a total of 2,642 individuals as control group. To prevent attrition, a replacement list was also randomly selected and provided to the project team.
At baseline for the second cohort, survey data was collected for a total of 4,426 applicants, of which 1,539 individuals in Brazzaville and 2,887 individuals in Pointe-Noire.

[3] All applicants were invited to visualize a video presenting the different trades offered. Applicants assigned to the information treatment had additional information in the video about median earnings for each trade in urban Congo.

Data collection

Figure 1: Timeline of the study



Three rounds of data collection are planned for each cohort. As displayed in Figure 1, the baseline survey was conducted in October-November 2015 for the first cohort and September-October 2018 for the second cohort. The trainings were launched in January 2016 for the first cohort's treatment group and in November 2018 for the second cohort's treatment group. Midline data was collected in May-June 2017 for the first cohort and June-July 2020 for the second cohort. Endline data was collected in June-July 2018 for the first cohort and is expected to be collected from mid-October to December 2021 for the second cohort.

Additionally, monitoring data was collected and shared by training centers. This data includes daily attendance rates, internship completion rates and final graduation rates.

3. Analytical framework

We make a distinction between primary and secondary outcomes. The success of the intervention is determined by considering its effect on primary outcomes. The secondary outcomes provide additional information on how the intervention may transform the beneficiaries' lives through other channels than improved employment prospects. For each primary outcome, we consider a single measure. Other measures will however be included as robustness checks.

A - Primary outcomes of interest

1) Main measures

H1a: Receiving access to the training increases employment in the past 7 days
H1b: Receiving access to the training without having received information on trade-specific earnings increases employment in the past 7 days
H1c: Receiving information on trade-specific earnings without receiving access to the training increases employment in the past 7 days
H1d: Receiving both the information on trade-specific earnings and the training increases employment in the past 7 days
H1e: Receiving both the information on trade-specific earnings and the training compared to only the training increases employment in the past 7 days

Main outcome measure	emp_7d	A dummy variable equal to 1 if the respondent has worked in the past 7 days

H2a: Receiving access to the training increases earnings
H2b: Receiving access to the training without having received information on trade-specific earnings increases earnings
H2c: Receiving information on trade-specific earnings without receiving access to the training increases earnings
H2d: Receiving both the information on trade-specific earnings and the training increases earnings
H2e: Receiving both the information on trade-specific earnings and the training compared to only the training increases earnings

Main outcome measure	earn_30d*	Earnings generated by all activities in the past 30 days

H3a: Receiving access to the training intensifies job search
H3b: Receiving access to the training without having received information on trade-specific earnings intensifies job search
H3c: Receiving information on trade-specific earnings without receiving access to the training intensifies job search
H3d: Receiving both the information on trade-specific earnings and the training intensifies job search

* This variable and all other variables marked with * are modified to deal with outliers, following the methodology described in section *D – Outliers.*

H3e: Receiving both the information on trade-specific earnings and the training compared to only the training intensifies job search

Main outcome measure	jobS_30d	A dummy variable equal to 1 if the respondent has looked for work in the past 30 days

H4a: Receiving access to the training increases self-employment
H4b: Receiving access to the training without having received information on trade-specific earnings increases self-employment
H4c: Receiving information on trade-specific earnings without receiving access to the training increases self-employment
H4d: Receiving both the information on trade-specific earnings and the training increases self-employment
H4e: Receiving both the information on trade-specific earnings and the training compared to only the training increases self-employment

Main outcome measure	self_emp	A dummy variable for whether the respondent reports a self-employed activity as his/her main income-generating activity

H5a: Receiving access to the training improves the respondent's empowerment and decision-making
H5b: Receiving access to the training without having received information on trade-specific earnings improves the respondent's empowerment and decision-making
H5c: Receiving information on trade-specific earnings without receiving access to the training improves the respondent's empowerment and decision-making
H5d: Receiving both the information on trade-specific earnings and the training improves the respondent's empowerment and decision-making
H5e: Receiving both the information on trade-specific earnings and the training compared to only the training improves the respondent's empowerment and decision-making

Main measure	empwr_sc	Recipient's empowerment score based on her involvement in making decisions for herself and her household

2) Additional measures (robustness checks)

For H1

Additional outcome measures	usual_emp	A dummy variable equal to 1 when the respondent declares being usually employed
	iga_30d	A dummy variable equal to 1 if the respondent has had any income-generating activity in the past 30 days
	iga_hrs_month	Total number of hours usually spent by the respondent on all income-generating activities in a typical month

For H2

Additional outcome measure	earn_main_30d*	Earnings generated by the main activity in the past 30 days

For H3

Additional outcome measures	jobapp_30d	Number of job applications submitted by the respondent in the past 30 days
	jobS_7d	A dummy variable equal to 1 if the respondent has looked for work in the past 7 days
	jobS_hrs_7d*	Number of hours the respondent has spent looking for work in the past 7 days

For H4

Additional outcome measure	self_emp_future	A dummy variable equal to 1 if the respondent wants to start a new business in the next year

B - Secondary outcomes

H6a: Receiving access to the training increases the recipient's expenditures
H6b: Receiving access to the training without having received information on trade-specific earnings increases the recipient's expenditures
H6c: Receiving information on trade-specific earnings without receiving access to the training increases the recipient's expenditures
H6d: Receiving both the information on trade-specific earnings and the training increases the recipient's expenditures

H6e: Receiving both the information on trade-specific earnings and the training compared to only the training increases the recipient's expenditures

Secondary outcome measure	tot_exp*	Total value of the respondent's expenditures in the past 30 days
Additional outcome measure	noness_exp*	Value of the respondent's non-essential expenditures including tobacco, entertainment and beauty products in the past 30 days
	ess_exp*	Value of the respondent's essential expenditures including rent, health, education, food and utilities in the past 30 days

H7a: Receiving access to the training increases the recipient's savings
H7b: Receiving access to the training without having received information on trade-specific earnings increases the recipient's savings
H7c: Receiving information on trade-specific earnings without receiving access to the training increases the recipient's savings
H7d: Receiving both the information on trade-specific earnings and the training increases the recipient's savings
H7e: Receiving both the information on trade-specific earnings and the training compared to only the training increases the recipient's savings

Secondary outcome measure	tot_sav*	Total value of the respondent's savings
Additional outcome measure	has_sav	A dummy variable equal to 1 if the respondent has savings

H8a: Receiving access to the training increases the recipient's employment stability
H8b: Receiving access to the training without having received information on trade-specific earnings increases the recipient's employment stability
H8c: Receiving information on trade-specific earnings without receiving access to the training increases the recipient's employment stability
H8d: Receiving both the information on trade-specific earnings and the training increases the recipient's employment stability
H8e: Receiving both the information on trade-specific earnings and the training compared to only the training increases the recipient's employment stability

Secondary outcome measure	duration_iga	Duration of longest income-generating activity

H9a: Receiving access to the training does not increase the recipient's perception of harassment
H9b: Receiving access to the training without having received information on trade-specific earnings does not increase the recipient's perception of harassment
H9c: Receiving information on trade-specific earnings without receiving access to the training does not increase the recipient's perception of harassment
H9d: Receiving both the information on trade-specific earnings and the training does not increase the recipient's perception of harassment
H9e: Receiving both the information on trade-specific earnings and the training compared to only the training does not increase the recipient's perception of harassment

Secondary outcome measure	harass_freq	A dummy variable equal to 1 if the respondent states that harassment is frequent among his family and friends
Additional outcome measure	harass_wrk_fq	A dummy variable equal to 1 if the respondent states that harassment is frequent at school or work among women in his community

H10a: Receiving access to the training does not increase the recipient's fertility
H10b: Receiving access to the training without having received information on trade-specific earnings does not increase the recipient's fertility
H10c: Receiving information on trade-specific earnings without receiving access to the training does not increase the recipient's fertility
H10d: Receiving both the information on trade-specific earnings and the training does not increase the recipient's fertility
H10e: Receiving both the information on trade-specific earnings and the training compared to only the training does not increase the recipient's fertility

Secondary outcome measure	dep_child	A dummy variable equal to 1 if the respondent has at least one dependent child

C - Mechanisms

H11a: Receiving access to the training improves exposure to role models
H11b: Receiving access to the training without having received information on trade-specific earnings improves exposure to role models
H11c: Receiving information on trade-specific earnings without receiving access to the training improves exposure to role models
H11d: Receiving both the information on trade-specific earnings and the training improves exposure to role models
H11e: Receiving both the information on trade-specific earnings and the training compared to only the training improves access to network

Main measure	role_model	A dummy variable equal to 1 if the recipient has a role model
Additional measure	male_role_model	A dummy variable equal to 1 if the recipient has a male role model

H12a: Receiving access to the training increases capital investment in one's business
H12b: Receiving access to the training without having received information on trade-specific earnings increases capital investment in one's business
H12c: Receiving information on trade-specific earnings without receiving access to the training increases capital investment in one's business
H12d: Receiving both the information on trade-specific earnings and the training increases capital investment in one's business
H12e: Receiving both the information on trade-specific earnings and the training compared to only the training increases capital investment in one's business

Main measure	prod_assets*	Total value of productive assets
Additional measure	prod_assets_new*	Total value of productive assets acquired less than a year ago

H13a: Receiving access to the training improves technical knowledge and skills
H13b: Receiving access to the training without having received information on trade-specific earnings improves technical knowledge and skills
H13c: Receiving information on trade-specific earnings without receiving access to the training improves technical knowledge and skills
H13d: Receiving both the information on trade-specific earnings and the training improves technical knowledge and skills

H13e: Receiving both the information on trade-specific earnings and the training compared to only the training improves technical knowledge and skills

Main measure	tech_index	Index of technical knowledge and experience
Additional measures	tech_know	A dummy variable equal to 1 if the respondent has technical knowledge
	tech_exp	A dummy variable equal to 1 if the respondent has technical experience

H14a: Receiving access to the training improves specific trade knowledge
H14b: Receiving access to the training without having received information on trade-specific earnings improves specific trade knowledge
H14c: Receiving information on trade-specific earnings without receiving access to the training does not improve specific trade knowledge
H14d: Receiving both the information on trade-specific earnings and the training improves specific trade knowledge
H14e: Receiving both the information on trade-specific earnings and the training compared to only the training does not improve specific trade knowledge

Main measure	trade_know_sc	Recipient's score assessing knowledge on the trade he was assigned to

H15a: Receiving access to the training improves aspirations and expectations
H15b: Receiving access to the training without having received information on trade-specific earnings improves aspirations and expectations
H15c: Receiving information on trade-specific earnings without receiving access to the training improves aspirations and expectations
H15d: Receiving both the information on trade-specific earnings and the training improves aspirations and expectations
H15e: Receiving both the information on trade-specific earnings and the training compared to only the training improves aspirations and expectations

Main measure	earn_asp_10y*	Monthly earnings which the respondent aspires to in ten years
Additional measure	earn_expect_10y*	Monthly earnings which the respondent expects to receive in ten years
	work_asp_10y	A dummy variable equal to 1 if the respondent aspires to have a job in ten years

H16a: Receiving access to the training improves self-esteem
H16b: Receiving access to the training without having received information on trade-specific earnings improves self-esteem
H16c: Receiving information on trade-specific earnings without receiving access to the training does not improve self-esteem
H16d: Receiving both the information on trade-specific earnings and the training improves self-esteem
H16e: Receiving both the information on trade-specific earnings and the training compared to only the training improves self-esteem

Main measure	self_esteem_sc	Score to the Rosenberg 3-item self-esteem scale
Additional measure	grit_sc	Score to the Duckworth et al. grit scale

H17a: Receiving access to the training improves socio-emotional skills
H17b: Receiving access to the training without having received information on trade-specific earnings improves socio-emotional skills
H17c: Receiving information on trade-specific earnings without receiving access to the training does not improve socio-emotional skills
H17d: Receiving both the information on trade-specific earnings and the training improves socio-emotional skills
H17e: Receiving both the information on trade-specific earnings and the training compared to only the training does not improve socio-emotional skills

Main measure	ses_score	Aggregate of principal component scores of subjective measures of 12 socio-emotional skills
Additional measures	intra_ses_score	Aggregate of principal component scores of subjective measures of 5 intrapersonal socio-emotional skills
	inter_ses_score	Aggregate of principal component scores of subjective measures of 7 interpersonal socio-emotional skills

H18a: Receiving access to the training improves the likelihood that the respondent works in the trade that he was assigned to
H18b: Receiving access to the training without having received information on trade-specific earnings improves the likelihood that the respondent works in the trade that he was assigned to
H18c: Receiving information on trade-specific earnings without receiving access to the training improves the likelihood that the respondent works in the trade that he was assigned to

H18d: Receiving both the information on trade-specific earnings and the training improves the likelihood that the respondent works in the trade that he was assigned to
H18e: Receiving both the information on trade-specific earnings and the training compared to only the training improves the likelihood that the respondent works in the trade that he was assigned to

Main measure	work_train_sect	A dummy variable equal to 1 if the respondent works in the trade that he was assigned to

H19a: Receiving access to the training improves the likelihood that the respondent works in a male-dominated trade
H19b: Receiving access to the training without having received information on trade-specific earnings improves the likelihood that the respondent works in a male-dominated trade
H19c: Receiving information on trade-specific earnings without receiving access to the training improves the likelihood that the respondent works in a male-dominated trade
H19d: Receiving both the information on trade-specific earnings and the training improves the likelihood that the respondent works in a male-dominated trade
H19e: Receiving both the information on trade-specific earnings and the training compared to only the training improves the likelihood that the respondent works in a male-dominated trade

Main measure	work_MDT	A dummy variable equal to 1 if the respondent works in a male-dominated trade

H20a: Receiving access to the training increases access to credit
H20b: Receiving access to the training without having received information on trade-specific earnings increases access to credit
H20c: Receiving information on trade-specific earnings without receiving access to the training increases access to credit
H20d: Receiving both the information on trade-specific earnings and the training increases access to credit
H20e: Receiving both the information on trade-specific earnings and the training compared to only the training increases access to credit

Main measure	LI_amt*	Total amount still owed by the respondent
Additional measure	LI_large	A dummy variable equal to 1 if the respondent owes above 5,000 FCFA

H21a: Receiving access to the training increases net amounts of transfers received
H21b: Receiving access to the training without having received information on trade-specific earnings increases net amounts of transfers received
H21c: Receiving information on trade-specific earnings without receiving access to the training increases net amounts of transfers received
H21d: Receiving both the information on trade-specific earnings and the training increases net amounts of transfers received
H21e: Receiving both the information on trade-specific earnings and the training compared to only the training does not impact net amounts of transfers received

Main measure	TI_amt *	Total amount of transfers that the respondent received from individuals in the past 30 days

D - Dealing with outliers

Outliers often arise in the distribution of monetary values such as revenue and savings. Our approach will be to winsorize at the appropriate percentile based on the distribution of the outcomes. We will also rely on an inverse sine hyperbolic transformation (ISTH)[4] to approximate a log specification without dropping the zeroes, as is common in the literature with wealth and income data (see Callen et al. 2019, Dupas, 2018).
The following variables will be winsorized or transformed:

- earn_30d: Earnings generated by all activities in the past 30 days
- earn_main_30d: Earnings generated by the main activity in the past 30 days
- jobS_hrs_7d: Number of hours the respondent has spent looking for work in the past 7 days
- tot_exp: Total value of the respondent's expenditures in the past 30 days
- noness_exp: Value of the respondent's non-essential expenditures including tobacco entertainment and beauty products in the past 30 days
- ess_exp: Value of the respondent's essential expenditures including rent, health, education, food and utilities in the past 30 days
- tot_sav: Total value of the respondent's savings
- prod_assets: Total value of productive assets
- prod_assets_new: Total value of productive assets acquired less than a year ago
- earn_asp_10y: Monthly earnings which the respondent aspires to in ten years
- earn_expect_10y: Monthly earnings which the respondent expects to receive in ten years

[4] $\log(y_i+(y_i^2+1)^{1/2})$

- LI_amt: Total amount still owed by the respondent
- TI_amt: Total amount of transfers that the respondent received from individuals in the past 30 days

E - Variables for subgroup analysis

a) Core dimensions of heterogeneity analysis

We set up our main equation with the goal of disaggregating the estimation of the effect by **gender** and by the duration of the **period of time** that has elapsed since the end of the training i.e., after around 19-21 months versus 32-37 months.

b) Further heterogeneity analysis

We will also estimate the interactions of the treatment with specific covariates that may affect the effectiveness of the training. These covariates include demographic characteristics, human capital and work experience, wealth, non-cognitive skills, sexual behavior, agency, priorities, aspirations, and training conditions.

1) *Demographic characteristics*
   - Age
   - Marital status (married or partnered)
   - Head of household
   - Children
2) *Human capital and work experience*
   - Education
   - Cognitive skills
   - Technical skills
   - Work experience
3) *Wealth*
   - Household assets
4) *Training conditions*
   - Distance from training center

4. Estimation Strategy

Specifications will be estimated on the sample of the second cohort. Results on the first cohort will be presented as robustness checks. For outcomes with insufficient statistical power, the first and second cohorts may be pooled together to estimate specifications below. In the pooled estimation, a cohort dummy will be added as an additional control.

A - Internal validity checks

a) Test of balance

To check that the randomization was successful, we will estimate the following equation for each outcome variable:

$$Y_{i0} = \alpha \quad + \beta_1 T_i + \lambda_1 \pi_{i0} + \varepsilon_{it} \tag{1}$$

In which:

- Y_{i0} is the baseline value of the outcome variable considered,
- T_i is the indicator of training treatment status, equal to 1 if the individual *i* is in the treatment group, i.e. was invited to participate in a training program,
- π_{i0} is a vector of control variables which correspond to the stratification variables used in the randomization, i.e. gender, trade selection, whether the applicant has a dependent child, and whether the applicant received the information on trade-specific earnings (information treatment).

We will similarly check for balance for the characteristics considered for sub-group analysis.

b) Attrition

We will report the overall attrition, checking whether it is differentiated along the key dimensions considered in our analysis. These include primary and secondary outcomes of interest, as well as variables considered in the heterogeneity analysis. To do this, we estimate the following:

$$Char_i = \alpha \quad + \beta_1 Attrit_i + \varepsilon_i \tag{2}$$

In which:

- $Char_i$ is the baseline value of the dimension on which we want to test for differential attrition,
- $Attrit_i$ is a dummy variable equal to 1 for respondents who were not interviewed after the baseline (neither at midline nor at endline).

Then we will examine the possibility of differential attrition between treatment arms by estimating the following:

$$Attrit_i = \alpha \quad + \beta_1 T_i + \varepsilon_i \tag{3}$$

In which:

- $Attrit_i$ is a dummy variable equal to 1 for respondents who were not interviewed after the baseline (neither at midline nor at endline),
- T_i is the indicator of training treatment status, equal to 1 if the individual *i* is in the treatment group, i.e. was invited to participate in a training program.

Research question	**Test**
Differential attrition between treatment arms	H_0: β_1= 0

If attrition is uneven between groups, we will use Lee bounds the address the biases associated with this selective attrition.

B - Impact

To estimate the effect of the intervention on the outcomes included in our framework, we use an ANCOVA specification[5], controlling for the stratification variables used in the randomization. The randomization was stratified by four variables: gender, trade selection, whether the applicant received the information on earnings by trade (information treatment) and whether the applicant has a dependent child.

For outcomes without baseline values, a simple OLS estimation will be used, considering the same specifications below without the baseline value of these outcomes.

As our core estimation, we will estimate the following equation in order to estimate the impact of training on the outcomes of interest, capturing the midline and endline impacts while controlling for the stratification variables:

$$Y_{it} = \alpha \quad + \beta_1 T_i + \delta_1 Y_{i0} + \lambda_1 \pi_{i0} + \varepsilon_{it} \quad \text{where } t = \{1, 2\} \tag{4}$$

In which:

- Y_{it} is an outcome variable for individual *i* at midline (*t*=1) or at endline (*t*=2),
- Y_{i0} is the baseline value of the outcome variable of interest,

[5] Following McKenzie (2012), ANCOVA specification, controlling the baseline value of the outcome variable in the analysis, is found to be more efficient.

- T_i is the indicator of training treatment status, equal to 1 if the individual *i* is in the treatment group, i.e. was invited to participate in a training program,
- π_{i0} is a vector of control variables which correspond to the stratification variables used in the randomization, i.e. gender, trade selection, whether the applicant has a dependent child, and whether the applicant received the information on trade-specific earnings (information treatment).

Interpretation:

Research question	**Test**
Effect of training on outcome of interest	H_0: β_1= 0

In addition, we will interact the gender dummy variable with the training treatment variable to differentiate the impact of training by gender:

$$Y_{it} = \alpha \quad + \beta_1 T_i + \beta_2 W_i + \beta_3 T_i W_i + \delta_1 Y_{i0} + \lambda_1 \pi_{i0} + \varepsilon_{it} \quad \text{where } t = \{1, 2\} \qquad (5)$$

In which W_i is a dummy variable equal to one if the individual is a woman.

Interpretation:

Research question	**Test**
Effect of training on men's outcome of interest	H_0: β_1= 0
Effect of training on women's outcome of interest	H_0: ($\beta_1 + \beta_3$) = 0
Gender difference in effect of training on applicants' outcomes of interest	H_0: β_2= 0

In addition, we estimate the following equation to characterize how the impact of training varies depending on whether individuals received the information treatment (before application), in three samples: the whole sample, and men and women separately.

$$Y_{it} = \alpha \quad + \beta_1 T_i + \beta_2 GroupB_i + \beta_3 T_i GroupB_i + \delta_1 Y_{i0} + \lambda_1 \pi_{i0} + \varepsilon_{it} \quad \text{where } t=\{1, 2\} \quad (6)$$

In which:

- Y_{it} is an outcome variable for individual *i* at midline (*t*=1) or at endline (*t*=2)
- Y_{i0} is the baseline value of the outcome variable of interest

- T_i is the indicator of training treatment status, equal to 1 if the individual *i* is in the treatment group, i.e. was invited to participate in a training program
- $GroupB_i$ is a dummy variable equal to one if the individual received the information on trade-specific earnings (information treatment).
- π_{i0} is a vector of control variables which correspond to the stratification variables used in the randomization, i.e. gender, trade selection, and whether the applicant has a dependent child.

Interpretation:

Research question	**Test**
Effect of training for applicants who did not receive information on trade-specific earnings	H_0: $\beta_1 = 0$
Effect of training for women who did not receive information on trade-specific earnings	H_0: $\beta_{1\ (among\ women)} = 0$
Effect of training for men who did not receive information on trade-specific earnings	H_0: $\beta_{1\ (among\ men)} = 0$
The tests above address hypothesis Hnb as described in Section 3 for each outcome.	
Effect of information on trade-specific earnings for applicants who did not receive access to the training	H_0: $\beta_2 = 0$
Effect of information on trade-specific earnings for women who did not receive access to the training	H_0: $\beta_{2\ (among\ women)} = 0$
Effect of information on trade-specific earnings for men who did not receive access to the training	H_0: $\beta_{2\ (among\ men)} = 0$
The tests above address hypothesis Hnc as described in Section 3 for each outcome.	
Effect of receiving both the training and information on trade-specific earnings	H_0: $\beta_1 + \beta_2 + \beta_3 = 0$
Effect of receiving both the training and information on trade-specific earnings for women	H_0: $(\beta_1 + \beta_2 + \beta_3)_{(among\ women)} = 0$
Effect of receiving both the training and information on trade-specific earnings for men	H_0: $(\beta_1 + \beta_2 + \beta_3)_{(among\ men)} = 0$
The tests above address hypothesis Hnd as described in Section 3 for each outcome.	
Additional effect of receiving information on trade-specific earnings compared to only the training	H_0: $\beta_2 + \beta_3 = 0$

Additional effect of receiving information on trade-specific earnings compared to only the training for women	$H_0: (\beta_2 + \beta_3)_{(among\ women)} = 0$
Additional effect of receiving information on trade-specific earnings compared to only the training for men	$H_0: (\beta_2 + \beta_3)_{(among\ men)} = 0$
The tests above address hypothesis Hne as described in Section 3 for each outcome.	

Finally, we will estimate the following equation to test the sustainability of the impact of the training, and to see if the impact persists in the medium term:

$$Y_{it} = \alpha \quad + \beta_1 T_i + \beta_2 N_t + \beta_3 T_i * N_t + \delta_1 Y_{i0} + \lambda_1 \pi_{i0} + \varepsilon_{it} \quad \text{where } t = \{1,2\} \qquad (7)$$

In which:

- Y_{it} is an outcome variable for individual *i* at midline (*t*=1) or at endline (*t*=2)
- Y_{i0} is the baseline value of the outcome variable of interest
- T_i is the indicator of training treatment status, equal to 1 if the individual *i* is in the treatment group, i.e. was invited to participate in a training program
- N_t is a time period indicator, equal to 1 if the observation is an endline observation (*t*=2)
- π_{i0} is a vector of control variables which correspond to the stratification variables used in the randomization, i.e. trade selection, whether the applicant has a dependent child, gender, and whether the applicant received the information on trade-specific earnings (information treatment).

Here, the parameter β_3 will identify the change between midline and endline in the treatment effect of the training on the relevant outcome of interest.

C - Heterogeneity analysis

We plan to run several heterogeneity tests, comparing the sub-groups defined in sub-section E of the analytical framework, checking for variations in the effects of the program on all primary and secondary outcomes.

To test for heterogeneous effects, we estimate the following specification:

$$Y_{i1} = \alpha \quad + \beta_1 T_i + \beta_2 Z_{i0} + \beta_3 T_i Z_{i0} + \delta_1 Y_{i0} + \lambda_1 \pi_{i0} + \varepsilon_{it} \qquad (8)$$

In which:

- Y_{i1} is an outcome variable for individual *i* at midline (*t*=1)

- Y_{i0} is the baseline value of the outcome variable of interest
- T_i is the indicator of training treatment status, equal to 1 if the individual *i* is in the treatment group, i.e. was invited to participate in a training program
- Z_{i0} denotes the baseline value of the covariate to be interacted with the treatment assignment variable.
- π_{i0} is a vector of control variables which correspond to the stratification variables used in the randomization, i.e. trade selection, whether the applicant has a dependent child, and whether the applicant received the information on trade-specific earnings (information treatment).

Here, the parameter β_3 will identify the marginal impact of the treatment for individuals with characteristics Z_{i0} at baseline.

D - Dealing with multiple measures for each outcome

We describe in this section the strategy used in this study to limit Type I error (false positive or false discovery).
To do so, we kept only one main outcome per family of primary outcomes (employment, earnings, job search, self-employment, empowerment) and secondary outcomes (expenditures, savings, employment stability, harassment, fertility). We plan on running robustness checks using additional measures for each one of our core outcomes.
Additionally, to be able to look at a larger range of outcomes while limiting false discoveries, we will rely as much as possible on summary indices. These summary indices will provide a single outcome measure per family of outcomes.

We have also selected a range of intermediate outcomes, which we will test the effects of the intervention on, in order to identify the mechanisms through which the intervention operates. If we were to detect a significant effect on these intermediate outcomes but not on the core outcomes, this would not change how we answer our main research questions on these core outcomes. Therefore, the implications of rejecting the null hypothesis are limited.