

Pre-analysis plan for a field experiment in the Republic of Congo: Addressing gender-based occupational segregation through the provision of sector-specific information on earnings

March 2019

Abstract

The authors are currently running a randomized field experiment testing the effect of providing information on sector-specific earnings on men and women's choice of trade and training outcomes. The young men and women included in the experimental sample are applicants to a vocational training program in the Republic of Congo. They are randomly selected to receive information on sector-specific earnings prior to picking the trade in which they want to be trained. This information makes it apparent that median earnings are typically higher in traditionally male-dominated sectors. While we observe the effect of this intervention on all applicants, we are particularly interested to test whether providing this information encourages women to cross over to non-traditional trainings. This pre-analysis plan provides an overview of the sample parameters, the analytical framework, the estimation strategy and the variables to be constructed to implement it.

1. Introduction

As part of a youth employment project launched by the government of the Republic of Congo, this study tests the effect of providing information on sector-specific earnings on the choice of trade and training outcomes of male and female applicants to a vocational training program. This intervention is primarily meant as a strategy to address gender-segregation in the labor market.

In the Republic of Congo as in many countries in Sub-Saharan Africa, there are substantial gender gaps in the youth labor market. Not only are young women less likely to work than young men, but when they are, they tend to engage in a very limited range of activities. At the national level, 40 percent of young women work in agriculture, and 33 percent work in trade. Young men, on the other hand, seem to find opportunities in multiple sectors including services, construction, transportation and manufacturing. Providing information on sector-specific earnings can help young women in some of the challenges they face. Specifically, young women's access to information can be more limited than that of young men. Providing information on earnings before applicants select a trade is a way to remediate this gap. Furthermore, this information can become a tool for young women facing pressure to conform to gender norms. Emphasizing

earnings potential might make it more acceptable for them to work or seek training in non-traditional sectors.

This study randomly selects male and female applicants to receive information on sector-specific earnings prior to picking the trade in which they want to be trained. One of the salient aspects of the information provided is that median earnings are typically higher in traditionally male-dominated sectors. While we observe the effect of this intervention on a range of sub-groups, one of the study's key outcome of interest is the effect of the information on the likelihood that young-women select a traditionally male-dominated trade. To make sure that we can also assess the potential adverse effects of inducing such crossovers, we also track the dropout rate across a broad set of sub-groups.

2. Description of the sample

Applicants had the opportunity to apply to a vocational training program (known as subcomponent 1.1 of the project) and an apprenticeship program (known as subcomponent 1.3). There was substantial overlap in the list of trades/activities available in each sector, although the two lists did not fully overlap.

Prior to the beginning of the application period, the government launched a four weeks communication campaign describing the trades available in each program, as well as the content and duration of each program. The communication materials also covered the eligibility criteria and application dates and procedures. Though the project targets out of school youth, it has specific eligibility criteria for each program. The criteria for the vocational training program are different from that of the apprenticeship program. The former targets youth who aged between 17 and 30 years and who completed at least the third grade of the lower secondary education but left school before obtaining the upper secondary education certification. The latter welcomes youth who are between 16 and 24 years and who dropped out of primary school or who have no education background.

Youth who wanted to apply to the program had to go to enrolment centers located in different areas of each city. In total 8 centers (4 in Brazzaville and 4 in Pointe Noire) were opened for 14 consecutive days. Roughly 30 officers worked in each center to help candidates with their applications. The enrollment process consisted of five steps: (i) Youth were welcomed and briefed on the project's training programs and selection criteria. At this step, the applicant had to choose between the apprenticeship program and the vocational training program; (ii) Youth met with a team for the identification of households that are already registered in the Social Registry (RSU), which is being developed by another World Bank financed project (Lisungi) for the assessment of the vulnerability of Congolese households. Those whose households were not in the RSU provided information for a future survey on vulnerability; (iii) the applicants met with surveyors who filled out electronic application forms on tablets for each application asking questions on youth demographic, educational background and economic characteristics; (iv) Applicants were then shown videos on trades available for training and apprenticeship. Half of the applicants were randomly selected to watch a video of about twenty-five minutes containing

the expected average monthly earnings per trade. The other half also watched a video, but without the information on incomes per trade. After the video, applicants had to select two trades; one as first choice and the other one as second choice. This step is the last for the applicants applying to the apprenticeship program. For those applying to the vocational training, there is a last step during which applicants took short French and Math tests deemed to assess their literacy and numeracy levels.

At the end of the enrollment period, the project was able to collect 14,639 applications of which 39 percent were submitted by young women. This study will be based on two datasets: application forms data and baseline data. The baseline data takes only into account applicants to the vocational training program, as this program is subject to an impact evaluation. The application form data is limited to the information required to assess the eligibility and the trade choice of the applicants, while the baseline data collected detailed information about the applicants in terms of their education history, their economic activities, information on their households and also information on their perception of gender norms.

3. Analytical framework

1) Average effect of the information intervention on estimates of earnings

- H1a: Applicants who have received information on sector-specific earnings give lower estimates of earnings in traditionally male-dominated trades.
- H1b: The difference between estimates of earnings in traditionally male-dominated trades and estimates of earnings in other trades are larger for applicants who have received information on sector-specific earnings.
- H1c: The change brought by the information on sector-specific earnings in terms of the estimated difference in earnings between traditionally male-dominated trades and other trades will be larger for women than for men.

2) Average effect of the information intervention on trade choice, by gender

- H2a: Men who have received the information intervention are more likely to opt for a traditionally male-dominated trade
- H2b: Women who have received the information intervention are more likely to opt for a traditionally male-dominated trade

- H2c: Receiving the information intervention has a greater effect on women's likelihood of opting for a male-dominated trade than on men's.

3) Sub-groups analysis: Variations in the effect of the information intervention on trade choice

Demographic characteristics

Marital status (application form)

- H3a1: The effect on trade choice of receiving the information on sector-specific earnings is greater for women who are not married/partnered
- H3a2: The effect on trade choice of receiving the information on sector-specific earnings does not vary with men's marital status
- H3a3: Being married is associated with a greater decrease in the effect of receiving the information for women than it is for men

Dependents/Family status

- H3b1: The effect on trade choice of receiving the information on sector-specific earnings is greater for women who have no children
- H3b2: The effect on trade choice of receiving the information on sector-specific earnings does not vary when men have children
- H3b3: Having no children is associated with a greater increase in the effect of receiving the information for women than it is for men

Human capital/technical skills

Education (application form)

- H3c1: The effect on trade choice of receiving the information on sector-specific earnings is greater for women who are relatively better educated
- H3c2: The effect on trade choice of receiving the information on sector-specific earnings is greater for men who are relatively better educated
- H3c3: Being relatively better educated than other people of the same gender is associated with a greater increase in the effect of receiving the information for women than it is for men

Technical skills

- H3d1: The effect on trade choice of receiving the information on sector-specific earnings is lower for women who have strong technical skills
- H3d2: The effect on trade choice of receiving the information on sector-specific earnings is lower for men who have strong technical skills
- H3d3: Having strong technical skills is associated with a greater decrease in the effect of receiving the information for women than it is for men

Non-cognitive skills

Self-esteem

- H3e1: The effect on trade choice of receiving the information on sector-specific earnings is higher for women who have strong self-esteem
- H3e2: The effect on trade choice of receiving the information on sector-specific earnings does not vary with men's level of self-esteem
- H3e3: Having strong self-esteem is associated with a greater increase in the effect of receiving the information for women than it is for men

Self Assessments

- H3f1: The effect on trade choice of receiving the information on sector-specific earnings is greater for women who rank themselves as more competent/intelligent than other women
- H3f2: The effect on trade choice of receiving the information on sector-specific earnings is greater for men who rank themselves as more competent/intelligent than other men
- H3f3: Ranking oneself as more intelligent than other people of the same gender is associated with a greater increase in the effect of receiving the information for women than it is for men

Anticipated discrimination

- H3g1: The effect on trade choice of receiving the information on sector-specific earnings is greater for women who anticipate less gender-based discrimination in the labor market

H3g2: The effect on trade choice of receiving the information on sector-specific earnings does not vary with men's anticipation of gender-based discrimination in the labor market

H3g3: Anticipating gender-based discrimination in the labor market is associated with a greater decrease in the effect of receiving the information for women than it is for men

Gender beliefs

H3h1: The effect on trade choice of receiving the information on sector-specific earnings is lower for women who hold gendered beliefs regarding men and women's roles in society

H3h2: The effect on trade choice of receiving the information on sector-specific earnings does not vary with men's beliefs regarding men and women's roles in society

H3h3: Holding gendered attitudes regarding men and women's roles in society is associated with a greater decrease in the effect of receiving the information for women than it is for men

Networks

H3i1: The effect on trade choice of receiving the information on sector-specific earnings is lower for women who have close ties to someone working in a traditionally male-dominated sector

H3i2: The effect on trade choice of receiving the information on sector-specific earnings is lower for men who have close ties to someone working in a traditionally male-dominated sector

H3i3: Having close ties to someone working in a traditionally male-dominated sector is associated with a greater decrease in the effect of receiving the information for women than it is for men

Receiving advice (application form)

H3j1: The effect on trade choice of receiving the information on sector-specific earnings is lower for women who have been explicitly advised by people in their direct network prior to their choice

- H3j2: The effect on trade choice of receiving the information on sector-specific earnings is lower for men who have been explicitly advised by people in their direct network prior to their choice
- H3j3: Having been explicitly advised prior to their choice is associated with a greater decrease in the effect of receiving the information for women than it is for men

Job priorities

Earnings

- H3l1: The effect on trade choice of receiving the information on sector-specific earnings is higher for women who rank "earnings" as their top priority when looking for a job
- H3l2: The effect on trade choice of receiving the information on sector-specific earnings is higher for men who rank "earnings" as their top priority when looking for a job
- H3l3: Ranking "earnings" as their top priority when looking for a job is associated with a greater increase in the effect of receiving the information for women than it is for men

Multitasking

- H3m1: The effect on trade choice of receiving the information on sector-specific earnings is lower for women who rank "multitasking" as their top priority when looking for a job
- H3m2: The effect on trade choice of receiving the information on sector-specific earnings is lower for men who rank " multitasking " as their top priority when looking for a job
- H3m3: Ranking "multitasking" as their top priority when looking for a job is associated with a greater decrease in the effect of receiving the information for women than it is for men

**4) Among women who selected a male dominated trade:
Differences between "organic" and "encouraged" crossovers**

These tests are performed among women who selected male dominated trades only. They compare women who "spontaneously" picked a male dominated trade ("organic" crossovers) and women who did so after watching the video on sector-specific earnings ("encouraged" crossovers).

Furthermore, in the baseline survey, people were asked whether having watched the video influenced their choice. This creates the possibility of conducting other comparisons, between on the one hand women in male dominated trades who declared that they made this choice as a result of the video to all other women in male dominated trades

Tests to be conducted among women who selected male dominated trades only, considering two groups successively:

- All women who selected a male dominated trade (whether they applied to the apprenticeship or vocational training component of the project)
- Women who selected a male dominated trade AND were included in the main study sample

For women included in the study sample, we will run a second set of tests controlling for the score used to select individuals into this sample. In trades where there was excess demand (more applications than slots in the control and treatment group combined), we used a score capturing applicants' basic skills and motivations that was assigned to them as they submitted the application form to select individuals into the study sample. Furthermore, in order to maximize the experiment's statistical power, the study sample was selected so that there would be the same number of men and women who had received the information and who had not in each trade. In a range of male-dominated trades for which there was excess demand among demand, this excess was greater among women who had received the information on earnings than among men. As a result, the selection process was more competitive for the women who had received the information (they had to clear a higher score threshold). To control for this, we include the score that applicants received among the explanatory variables.

Demographic characteristics

- | | |
|------|--|
| H4a: | Women who saw the video and picked a male dominated tend to be younger than other women who picked a male dominated trade |
| H4b: | Women who saw the video and picked a male dominated are less likely to be single/unmarried than other women who picked a male dominated trade |
| H4c: | Women who saw the video and picked a male dominated are less likely to be the head of their household than other women who picked a male dominated trade |
| H4d: | Women who saw the video and picked a male dominated are more likely to have children than other women who picked a male dominated trade |

Human capital, technical skills and work experience

- H4e: Women who saw the video and picked a male dominated trade tend to have lower educational attainments than other women who picked a male dominated trade
- H4f: Women who saw the video and picked a male dominated trade tend to have lower cognitive skills than other women who picked a male dominated trade
- H4g: Women who saw the video and picked a male dominated trade tend to have lower technical skills skills than other women who picked a male dominated trade
- H4h: Women who saw the video and picked a male dominated tend to be more likely to be employed prior to the training than other women who picked a male dominated trade

Wealth

- H4i: Women who saw the video and picked a male dominated trade tend to be less wealthy than other women who picked a male dominated trade (asset index)
- H4j: Women who saw the video and picked a male dominated trade tend to have less income than other women who picked a male dominated trade (IGA module)
- H4k: Women who saw the video and picked a male dominated trade tend to be more credit constraints than other women who picked a male dominated trade (credit constraints from GIL quests)

Non-cognitive skills

- H4l: Women who saw the video and picked a male dominated trade tend to have lower self-esteem scores than other women who picked a male dominated trade

Self-assessment

- H4m: Women who saw the video and picked a male dominated trade are less likely to rate themselves as more intelligent than others of their own gender, compared other women who picked a male dominated trade
- H4n: Women who saw the video and picked a male dominated trade are less likely to rate themselves as more ambitious than others of their own gender, compared other women who picked a male dominated trade

Networks

H4o: Women who saw the video and picked a male dominated trade tend to be less likely to know someone in the trade they picked than other women who picked a male dominated trade

Priorities

H4p: Women who saw the video and picked a male dominated trade are more likely to have ranked "earnings" as their first priority in selecting a job than other women who picked a male dominated trade

Aspirations

H4q: Women who saw the video and picked a male dominated trade are more likely to have higher earnings aspirations than other women who picked a male dominated trade

5) Differences in dropout rates by gender, trade choice and information treatment status

In the analysis of the predictors of participants' likelihood of dropping out, we consider three distinct measures of participation

- *Whether the individual finished the training and was placed in an internship (binary variable)*
- *Whether the individual "graduated" (binary variable)*
- *The individual attendance rate, computed using administrative data (continuous variable)*

Tests (series 5 & 6) to be conducted among all applicants invited to a training

H5a: Women are more likely to dropout than men

H5b: Women in male dominated trades are more likely to drop out than women in other trades

H5c: Men in male dominated trades are not more likely to drop out than men in other trades

H5d: Women in male dominated trades are more likely to drop out than men in similar trades

H5e: Women in non-male dominated trades are more likely to drop out than men in similar trades

- H5f: Having selected a male dominated trade increases women's likelihood of dropping out more than having selected a non-male dominated trade increases men's
- H5g: Having received the information on earnings increases women's likelihood of dropping out, but does not increase men's
- H5h: Having selected a male dominated trade increases women's likelihood of dropping out more than having selected a non-male dominated trade increases men's

6) Subgroup analysis: Variations in dropout rates

In order to observe variations in dropout rates across subgroups, interacting the subgroup characteristics with 1) gender 2) whether the respondent has received information on sector specific earnings, we will run the following succession of two tailed test:

- H6a1: Women with characteristic X have a similar/higher/lower likelihood of dropping out than other women
- H6a2: Men with characteristic X have a similar/higher/lower likelihood of dropping out than other Men
- H6a3: Women with characteristic X have a similar/higher/lower likelihood of dropping out than men with the same characteristic
- H6a4: Women with characteristic X being trained in male dominated trades have a higher likelihood of dropping out than women with the same characteristic being trained in other trades.
- H6a5: Men with characteristic X being trained in male dominated trades have a similar likelihood of dropping out than men with the same characteristic being trained in other trades.
- H6a6: Women with characteristic X being trained in male dominated trades have a similar/higher/lower likelihood of dropping out than other women being trained in the same trades
- H6a7: Women with characteristic X being trained in male dominated trades have a similar/higher/lower likelihood of dropping out than other women being trained in the same trades
- H6a8: Women with characteristic X being trained in male dominated trades have a similar/higher/lower likelihood of dropping out than comparable men

H6a9: Women with characteristic X being trained in male dominated trades have a similar/higher/lower likelihood of dropping out than men with the same characteristics being trained in non-male dominated trades

The subgroup characteristics we will consider are as follows:

1) *Demographic characteristics*

- Age
- Marital status
- Head of household
- Children

2) *Human capital and work experience*

- Education
- Cognitive skills
- Technical skills
- Work experience

3) *Wealth*

- Assets
- Income (high income is defined successively in two ways: 1) higher than the median in the full study population and 2) higher than the median for the respondent's own gender)

4) *Non cognitive skills*

- Grit
- Self-esteem

5) *Sexual behavior*

- Age at first relation

6) *Networks*

- Knows someone in the trade selected

7) *Agency*

- Agency scale

8) *Priorities*

- Ranked "earnings" as their priority

9) *Aspirations*

- Aspires to high earnings in ten years

10) Training conditions

- Distance from training center
- Proportion of men in classes

11) Choices

- Two choices
- Were assigned to first choice

7) Among women who selected a male dominated trade: Differences in predictors of dropping out between "organic" and "encouraged" crossovers

Tests to be conducted among women who selected a male dominated trade and were invited to participate in a training

In the fourth set of hypotheses, we explore differences between women in male dominated trade who had and had not received the information on earnings, considering a range of characteristics. In this seventh set of hypotheses, we consider how these characteristics affect the dropout rates of both groups of women. As mentioned, the selection process itself, based on a score, may have exacerbated or minimized the differences between these two groups. For this reason, the analysis control for the score used to select individual into the study sample.

Demographic characteristics	
------------------------------------	--

- | | |
|------|--|
| H7a: | Younger women in male-dominated trades who have received the information are more likely to drop out than younger women in the same trades who have not received this information |
| H7b: | Married women in male-dominated trades who have received the information are more likely to drop out than married women in the same trades who have not received this information |
| H7c: | Women who are the head in their household, trained in male-dominated trades and who have received the information are more likely to drop out than similar women in the same trades who have not received this information |
| H7d: | Women who are the head in their household, trained in male-dominated trades and who have received the information are more likely to drop out than similar women in the same trades who have not received this information |

Human capital and work experience

- H7e: Women with high educational attainments who are trained in male dominated trades and who have received the information are not more likely to drop out than women with similar attainment, who are in the same trades, and who have not received this information
- H7f: Women with strong cognitive skills who are trained in male dominated trades and who have received the information are not more likely to drop out than women with similar cognitive skills, who are in the same trades, and who have not received this information
- H7g: Women with strong technical skills who are trained in male dominated trades and who have received the information are not more likely to drop out than women with similar technical skills, who are in the same trades, and who have not received this information
- H7h: Women who worked prior to the training, trained in male-dominated trades and who have received the information are more likely to drop out than similar women, who are in the same trades, and who have not received this information

Wealth

- H7i: Wealthy women in male-dominated trades who have received the information are more likely to drop out than wealthy women in the same trades who have not received this information
- H7j: Women with high incomes in male-dominated trades who have received the information are more likely to drop out than women with similar incomes in the same trades who have not received this information
- H7k: Women who are less credit constrained, trained in male-dominated trades and who have received the information are more likely to drop out than women with similar credit constraints, who are in the same trades and who have not received this information

Non cognitive skills

- H7l: Women who scored high on the grit scale, trained in male-dominated trades, and who have received the information are more likely to drop out than women with similar grit scores, who are in the same trades and who have not received this information

H7m: Women who scored high on the self-esteem scale, trained in male-dominated trades, and who have received the information are more likely to drop out than women with similar self-esteem scores, who are in the same trades and who have not received this information

Sexual behavior

H7n: Women who were sexually active at a young age, trained in male-dominated trades, and who have received the information are more likely to drop out than similar women in the same trades, who have not received this information

Networks

H7o: Women who know someone in the trade they selected, trained in male-dominated trades, and who have received the information are not more likely to drop out than similar women in the same trades, who have not received this information

Agency

H7p: Women who scored high on the agency scale, trained in male-dominated trades, and who have received the information are not more likely to drop out than women with similar agency scores who are in the same trades, and who have not received this information

Priorities

H7q: Women who ranked "earnings" as their priority, trained in male-dominated trades, and who have received the information are more likely to drop out than women with similar priorities, who are in the same trades and who have not received this information

Aspirations

H7r: Women who aspire to high earnings in ten years, trained in male-dominated trades, and who have received the information are more likely to drop out than women with similar aspirations, who are in the same trades and who have not received this information

Training conditions

H7s: Women who leave far from the training center, trained in male-dominated trades, and who have received the information are not more likely to drop out than women who live at a similar distance, who are in the same trades and who have not received this information

H7t: Women who attend classes where men are the majority, trained in male-dominated trades, and who have received the information are not more likely to drop out than women who attend similar classes, who are in the same trades and who have not received this information

Choices

H7u: Women who listed two choices when applying, trained in male-dominated trades, and who have received the information are more likely to drop out than women who similarly listed two choices, who are in the same trades and who have not received this information

H7v: Women who were assigned to their first choice, trained in male-dominated trades, and who have received the information are more likely to drop out than women who were also assigned to their first choice, who are in the same trades and who have not received this information

4. Estimation strategy

A - Internal validity checks

a) Test of balance

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

$$Y_{i0} = \alpha + \gamma_1 \text{Group}B_i + \epsilon_i \quad (1)$$

In which:

- Y_{i0} is the baseline value of the outcome variable considered.
- $\text{Group}B_i$ captures the treatment assignment (see variable creation in section 5).

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

B - Differences in assessments of sector specific earnings

We will test for differences in estimates of sector-specific earnings between applicants who received the information and applicants who did not by estimating the following model:

$$\text{Estim}_i = \alpha + \gamma_1 \text{Group}B_i + \pi_{ic} + \epsilon_i \quad (2)$$

Interpretation:

Hypothesis	Test
H1a: Effect of information on estimates of earnings in male dominated trades	$H_0: \gamma_1 = 0$

We will test for differential changes – by gender – in estimates of the ratio of earnings in male dominated trades over earnings in non-male dominated trades. This ratio is a "level" measure of the respondent's perception of the gap between earnings in the two types of trades. To perform this test, we estimate the following models:

$$RatioEstim_i = \alpha + \gamma_1 GroupB_i + \pi_{ic} + \epsilon_i \quad (3)$$

Interpretation:

Hypothesis	Test
H1b: Effect of information on ratio of estimates of earnings in male dominated trades over estimates of earnings in non-male dominated trade	$H_0: \gamma_1 = 0$

$$RatioEstim_i = \alpha + \gamma_1 W_i + \gamma_2 GroupB_i + \gamma_3 GroupB_i W_i + \pi_{ic} + \epsilon_i \quad (4)$$

Interpretation:

Hypothesis	Test
H1c: Gender difference in the effect of information on ratio of estimates of earnings in male dominated trades over estimates of earnings in non-male dominated trade	$H_0: \gamma_3 = 0$

In these models:

- $Estim_i$ is a numerical variable capturing the respondent's estimate of typical earnings in different trades.
- $RatioEstim_i$ is a numerical variable capturing the ratio of the respondents' estimates of typical earnings in male dominated trades over their estimates of earnings in other trades.
- $MaleTrade_i$ is a categorical variable equal to one if the applicant selected a male trade as his/her first choice.
- W_i is a dummy variable equal to one if the individual is a woman.
- $GroupB_i$ captures the treatment assignment (see variable creation in section 5).
- π_{ic} is a vector of control variables including the individual's age.

C - Impact on trade selection

We will test the effect of providing information on sector-specific earnings on two different measures of trade choice, given that each applicant has the opportunity of selecting up to two trades, with the second choice being optional.

- The first measure (*MaleTrade*) is a dummy variable equal to one if the applicant selects a traditionally male dominated trade as his or her first choice. With this measure, we capture the effect of the information on applicants' first choices.
- The second measure (*N_MaleTrade*) is a categorical variable capturing the number of traditionally male dominated trades that each applicant has selected. This measure takes the values 0, 1 or 2. With this measure we capture the effect of the information on the proportion of traditionally male dominated trade as a share of all choices made.

To estimate the effect of the information intervention on trade selection, we estimate the following ANOVA specification:

$$D_i = \alpha + \gamma_1 W_i + \gamma_2 \text{Group}B_i + \gamma_3 \text{Group}B_i W_i + \pi_{ic} + \epsilon_i \quad (5)$$

Interpretation:

Hypothesis	Test
H2a: Effect of information on men's trade choices	$H_0: \gamma_2 = 0$
H2b: Effect of information on women's trade choices	$H_0: (\gamma_2 + \gamma_3) = 0$
H2c: Gender difference in effect of information on applicants' trade choices	$H_0: \gamma_3 = 0$

In this model:

- D_i is either the first (binary: *MaleTrade*) or second (categorical, *N_MaleTrade*) measure of trade choice.
- W_i is a dummy variable equal to one if the individual is a woman.
- $\text{Group}B_i$ captures the treatment assignment (see variable creation in section 5).
- π_{ic} is a vector of control variables including the individual's age and the center where he or she applied.

Other parameters:

- In the core specification, we estimate the model separately for applicants to vocational trainings (subcomponent 1.1) and applicants to apprenticeships (subcomponent 1.3). However, we will also do a robustness check including the full sample

- In the core specification, we include applicants from both cities in the same regression. By controlling for applications centers, we are also controlling for whether the individual applied in Brazzaville or Pointe Noire. However, in order to provide more specific information to the institutions who collected application in each city, we will also generate estimates separately for each of them.
- Standard errors will be clustered by application centers and half-days during the period of application.
- For the first, binary measure of trade-choice, we will estimate both a linear and a probit model as a robustness check. For the second, categorical measure of trade choice, we will estimate a linear model as well as an ordinal logistic regression.

D - Sub-group analysis - Impact on trade selection

We plan to run several heterogeneity tests, comparing the sub-groups defined in section 3 of the analytical framework, checking for variations in the effect of the information intervention on applicants' trade selection. To test for heterogeneous effects, we estimate the following specification, separately for men and women:

$$D_i = \alpha + \gamma_1 Z_i^0 + \gamma_2 GroupB_i + \gamma_3 GroupB_i Z_i^0 + \pi_{ic} + \epsilon_i \quad (6)$$

Interpretation:

Hypothesis	Test
H3Z1: Difference in effect of information between women with and without characteristic Z	$H_0: \gamma_{3(among\ women)} = 0$
H3Z2: Difference in effect of information between men with and without characteristic Z	$H_0: \gamma_{3(among\ men)} = 0$
H3Z3: Gender difference in how characteristic Z changes the effect of information on applicants' trade choices	$H_0: \gamma_{3(among\ women)} = \gamma_{3(among\ men)}$

In this model:

- D_i is either the first (binary) or second (categorical) measure of trade choice.
- $GroupB_i$ captures the treatment assignment (see variable creation in section 5)
- Z_i^0 denotes the covariates to be interacted with the treatment assignment variables.
- π_{ic} is a vector of control variables including the individual's age and the center where he or she applied.

E - Differences between "organic" and "encouraged" crossovers

To detect significant differences between women who selected male dominated trades after getting information on sector-specific earnings ("encouraged" crossovers) and women who selected these trades in the absence of any prompting ("organic" crossovers), we estimate the following specification among women who selected male dominated trades only, whether or not they were included in the study sample:

$$Char_i = \alpha + \gamma_1 GroupB_i + \sum_{n=1}^N \beta_n Center_n + \epsilon_i \quad (7A)$$

Interpretation:

Hypothesis	Test
H4: Among women who selected a male dominated trade: difference along characteristics i between individuals who have received the information on sector specific earnings and individuals who have not	$H_0: \gamma_1 = 0$

When conducting the same test among women in male dominated trades included in the study sample, we estimate:

$$Char_i = \alpha + \gamma_1 GroupB_i + \sum_{n=1}^N \beta_n Center_n + \theta_1 Score_i + \epsilon_i \quad (7A)$$

(adding the score used for selecting individuals into the study sample)

In these models:

- $Char_i$ is the dimension along which we compare the two categories of women
- $GroupB_i$ captures the treatment assignment (see variable creation in section 5)
- $Center_n$ denotes the different centers in which the individual submitted her application
- $Score_i$ denotes the score assigned to the individual when she submitted her application

Standard errors will be clustered by application centers and half-days during the period of application.

F - Predictors of dropping out

To detect significant differences in dropout rates by gender, choice of trade, and access to information on sector-specific earnings we estimate the following specifications among all applicants who have been invited to participate in a training:

$$Dropout_i = \alpha + \gamma_1 W_i + \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i \quad (8)$$

Interpretation:

Hypothesis	Test
H5a: Gender difference in likelihood of dropping out	$H_0: \gamma_1 = 0$

Separately for men and women:

$$Dropout_i = \alpha + \gamma_1 MaleTrade_i + \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i \quad (9)$$

Interpretation:

Hypothesis	Test
H5b: Difference in dropout rates between women who picked a male dominated trades and women who picked other trades	$H_0: \gamma_1(\text{among women}) = 0$
H5c: Difference in dropout rates between men who picked a male dominated trades and men who picked other trades	$H_0: \gamma_1(\text{among men}) = 0$

We estimate again equation 8 separately for individual in male dominated trades and in other trades, which yields the following tests:

Hypothesis	Test
H5d: Gender difference in dropout rates among individuals in male dominated trades	$H_0: \gamma_1(MD \text{ trades}) = 0$
H5e: Gender difference in dropout rates among individuals in non-male dominated trades	$H_0: \gamma_1(NON MD \text{ trades}) = 0$

In order to further explore the determinants of dropout rates we then estimate a range of additional specifications:

$$Dropout_i = \alpha + \gamma_1 W_i + \gamma_2 OppTrade_i + \gamma_3 OppTrade_i W_i + \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i \quad (10)$$

Interpretation:

Hypothesis	Test
H5f: Difference in dropout rates between women who picked a male dominated trades and men who picked other trades	$H_0: \gamma_1 + \gamma_3 = 0$

$$\begin{aligned}
Dropout_i &= \alpha + \gamma_1 W_i + \gamma_2 GroupB_i + \gamma_3 GroupB_i W_i \\
&+ \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i
\end{aligned} \tag{11}$$

Interpretation:

Hypothesis	Test
H5g: Gender difference in effect of receiving the information on dropout rates	$H_0: \gamma_3 = 0$

Among women only:

$$\begin{aligned}
Dropout_i &= \alpha + \gamma_1 MaleTrade_i + \gamma_2 GroupB_i + \gamma_3 GroupB_i MaleTrade_i \\
&+ \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i
\end{aligned} \tag{12}$$

Interpretation:

Hypothesis	Test
H5h1: Effect of information on dropout rates of women in non male dominated trades	$H_0: \gamma_2 = 0$
H5h2: Effect of information on dropout rates of women in male dominated trades	$H_0: (\gamma_2 + \gamma_3) = 0$
H5h3: Difference in effect of information on dropout rates between women who picked a male dominated trade and women who picked other trades	$H_0: \gamma_3 = 0$

In these models:

- $Dropout_i$ is one of the three measures of participation considered.
- W_i is a dummy variable equal to one if the individual is a woman
- $MaleTrade_i$ is a categorical variable equal to one if the applicant selected a male trade as his/her first choice.
- $GroupB_i$ captures the treatment assignment (see variable creation in section 5).
- $TrainingInst_n$ denotes the training institution to which the individual has been assigned
- $OppTrade_i$ a categorical variable equal to one for women in male dominated trades and men in non-male dominated trades

Standard errors will be clustered by training institutions

G - Sub-group analysis - variations in dropout rates

We plan to run several heterogeneity tests, comparing dropout rates between the sub-groups defined in section 6 of the analytical framework. To test for heterogeneous effects, we estimate the following two specifications among all applicants who have been invited to participate in a training:

$$Dropout_i = \alpha + \gamma_1 W_i + \gamma_2 Z_i^0 + \gamma_3 W_i Z_i^0 + \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i \quad (13)$$

Interpretation:

Hypothesis	Test
H6Z1: Effect of characteristic Z on men's dropout rates	$H_0: \gamma_2 = 0$
H6Z2: Effect of characteristic Z on women's dropout rates	$H_0: (\gamma_2 + \gamma_3) = 0$
H6Z3: Gender difference in effect of characteristic Z on participants' dropout rates	$H_0: \gamma_3 = 0$

Estimated separately for men and women:

$$Dropout_i = \alpha + \gamma_1 MaleTrade_i + \gamma_2 Z_i^0 + \gamma_3 MaleTrade_i Z_i^0 + \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i \quad (14)$$

Interpretation:

Hypothesis	Test
H6Z4: Among women with characteristic Z: difference in dropout rates between individuals in male dominated trades and individuals in other trades	$H_0: \gamma_3 \text{ (among women)} = 0$
H6Z5: Among men with characteristic Z: difference in dropout rates between individuals in male dominated trades and individuals in other trades	$H_0: \gamma_3 \text{ (among men)} = 0$

We then estimate equation 13 for individuals in male dominated trades only, which yields the following tests:

Hypothesis	Test
H6Z6: Among women in male dominated trades: difference in dropout rates between individuals with and without characteristic Z	$H_0: (\gamma_2 + \gamma_3) = 0$
H6Z7: Among men in male dominated trades: difference in dropout rates between individuals with and without characteristic Z	$H_0: \gamma_2 = 0$
H6Z8: Among individuals in male dominated trades: gender difference in effect of characteristic Z on dropout rates	$H_0: (\gamma_1 + \gamma_3) = 0$

Finally, we estimate:

$$\begin{aligned}
Dropout_i = & \alpha + \gamma_1 W_i + \gamma_2 OppTrade_i + \gamma_3 OppTrade_i W_i \\
& + \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i
\end{aligned} \tag{15}$$

Interpretation:

Hypothesis	Test
H6Z9: Difference in dropout rates between women with characteristic Z in male dominated trades and men with characteristic Z in non-male dominated trades	$H_0: \gamma_1 + \gamma_3 = 0$

In these models:

- $Dropout_i$ is one of the three measures of participation considered.
- W_i is a dummy variable equal to one if the individual is a woman.
- $MaleTrade_i$ is a categorical variable equal to one if the applicant selected a male trade as his/her first choice.
- Z_i^0 denotes the covariates to be interacted with the trade choice and gender variables.
- $TrainingInst_n$ denotes the training institution to which the individual has been assigned
- $OppTrade_i$ a categorical variable equal to one for women in male dominated trades and men in non-male dominated trades

Standard errors will be clustered by training institutions

H - Differences in predictors of dropping out between "organic" and "encouraged" crossovers

An additional series of heterogeneity tests is meant to detect differences between the determinants of dropping out among "organic" and "encouraged" crossovers. To do so, we estimate the following specification among women who selected a male-dominated trade AND were invited to attend the training:

$$Dropout_i = \alpha + \gamma_1 Z_i^0 + \gamma_2 GroupB_i + \gamma_3 GroupB_i Z_i^0 + \theta_1 Score_i + \sum_{n=1}^N \beta_n TrainingInst_n + \epsilon_i \quad (16)$$

Interpretation:

Hypothesis	Test
H7: Among women in male dominated trades: Difference in effect of characteristics Z between women who have received the information and women who have not.	$H_0: \gamma_3 = 0$

In this model:

- $Dropout_i$ is one of the three measures of participation considered.
- $GroupB_i$ captures the treatment assignment (see variable creation in section 5).
- Z_i^B denotes the covariates to be interacted with the treatment variable.
- $Score_i$ denotes the score assigned to the individual when she submitted her application
- $TrainingInst_n$ denotes the training institution to which the individual has been assigned

Standard errors will be clustered by training institutions.

I - Risk of attrition

There is no risk of differential attrition by treatment arm (received the information or not) for the analysis of the effect of information on applicants' choice of trade, since this analysis relies on application and baseline data.

For the analysis of dropout rates, the risk of differential attrition between individuals who received the information and individuals who did not is limited, since all individuals considered were invited to participate in the training. Nonetheless, this risk exists, and will be managed using Lee bounds.

J - Corrections for multiple hypotheses testing

When testing differences in dropout rates across subgroups, we are using three different measures of participation (see introduction of the 5th set of hypotheses). This means that we are running four separate tests for each specification considered and must adjust the p-values of these tests accordingly. To do so, we will control the *familywise error rate* (FWER), that is the probability of rejecting one or more true null hypothesis. We will use the Romano Wolf procedure to account of the dependence structure of the p-values of these different tests.

When exploring dimensions of heterogeneity other than gender, following Fink et al. (2014) we will report corrected p-value calculated using the False Discovery Method. Specifically, we will correct the p-value to reflect the number of heterogeneity dimensions considered for each of the following three sets of hypotheses:

- "H3", capturing heterogeneity in the effect the intervention on trade selection
- "H6" capturing heterogeneity in dropout rates
- "H7" capturing heterogeneity in the difference in dropout rate between treated and untreated women among those who picked a male dominated trade

5. Variable creation

ADMINISTRATIVE DATA - OUTCOMES

(Available for both individuals applying to 1.1 and 1.3)

Variable	Definition	Source/Construction
<i>MaleTrade</i>	A dummy variable =1 if the respondent's first choice is a trade that is traditionally male dominated	Application form
<i>N_MaleTrade</i>	A categorical variable including the number of male trades that the applicant selected, given that each applicant could make up to two choices	Application form

The trades categorized as being traditionally male dominated include: mechanics, auto-mechanics, electricity, welding, carpentry, civil engineering, plumbing, topography, refrigeration and air conditioning, crane operator, construction, glass/metallic construction, painting/tiling, mechanical engineering and industrial piping.

SURVEY DATA - ESTIMATES OF SECTOR-SPECIFIC EARNINGS

(Available for individuals applying to 1.1 only)

Variable	Definition	Source/Construction
<i>earnings_meca</i>	Respondent's assessment of earnings of self-employed auto-mechanics (monthly)	PP11A (adjusting for PP11B)
<i>earnings_welder</i>	Respondent's assessment of earnings of self-employed welder (monthly)	PP12A (adjusting for PP12B)
<i>earnings_maleDT</i>	Respondent's assessment of earnings in male dominated trades (monthly)	PP11A + PP12A/2 (adjusting for PP11B and PP12B)
<i>earnings_hairdresser</i>	Respondent's assessment of earnings of self-employed hair-dresser (monthly)	PP13A (adjusting for PP13B)
<i>earnings_tailor</i>	Respondent's assessment of earnings of self-employed tailor (monthly)	PP14A (adjusting for PP14B)
<i>earnings_nonmaleDT</i>	Respondent's assessment of earnings in non male dominated trades (monthly)	PP13A + PP14A/2 (adjusting for PP13B and PP14B)
<i>RatioEstim</i>	Ratio of the respondent's estimates of earnings in male dominated trades over earnings in non-male dominated trades	$(PP11A + PP12A) / (PP13A + PP14A)$ (adjusting for PP11B, PP12, PP13B and PP14B)

ADMINISTRATIVE DATA - PARAMETERS OF CORE SPECIFICATION - CHOICE OF TRADE

(Available for both individuals applying to 1.1 and 1.3)

Variable	Definition	Source/Construction
<i>Subcomponent1</i>	A dummy variable equal to 1 if the respondent applied to the vocational training rather than apprenticeship program	Application form
<i>GroupB</i>	A dummy variable equal to 1 if the applicant received the information on earnings by trade before applying to the program	in-the-field randomization

<i>City</i>	A categorical variable equal to 1 if the applicant is based in Brazzaville and 2 if the applicant is based in Pointe Noire	Application form
<i>application_half_day</i>	A categorical variable capturing the half day on which the applicant submitted his/her form	Application form
<i>Center</i>	A categorical variable capturing the center where the applicant submitted his/her form	Application form
<i>Woman (W)</i>	A dummy variable capture the respondent's gender	Application form
<i>age</i>	Age of the respondent	Application form

SURVEY DATA - SUB-GROUP ANALYSES

(Available for individuals applying to 1.1 only)

Variable	Definition	Source/Construction
<i>family status</i>		
<i>partnered</i>	A dummy variable for whether the respondent is married or in a committed relationship	SE2=2 SE2=3
<i>head_hh</i>	A dummy variable for whether the respondent is the head of his/her household	LM1=1
<i>dep_child</i>	A dummy variable for whether the respondent has at least one dependent child	SE4!=0
<i>education</i>		
<i>AL_high_school</i>	A dummy variable equal to 1 if the respondent attended <u>at least</u> high school	ED7=12 ED7=13 ED7=14 ED7=16 ED7=17
<i>work experience</i>		
<i>work_expce</i>	A dummy variable equal to 1 if the respondent was working prior to the training	EM1.11=1 EM1.11=2 EM2=1 AP1=1

<i>wealth</i>		
<i>high_assets</i>	A dummy variable equal to 1 if the respondent is from a household wealthier than the median. Wealth is measured through an mca analysis of the list of assets and characteristics of the respondent's house	BP1 to BP24
<i>high_tot_rev1</i>	A dummy variable equal to 1 if the respondent's earnings in the past 30 days are higher than the median among the full study sample	AR8
<i>high_tot_rev2</i>	A dummy variable equal to 1 if the respondent's earnings in the past 30 days are higher than the median among study participants of his/her own gender	AR8
<i>gender beliefs</i>		
<i>gender_beliefs1</i>	A score capturing the extent to which the respondent holds gendered beliefs on men and women's roles	TBD (NG1-NG4)
<i>gender_beliefs2</i>	A score capturing the extent to which the respondent holds gendered beliefs on men and women's roles	TBD (NG1-NG4)
<i>dist_gender</i>	An average of the differences between the scores that the respondent assigns to men and women in general, across three qualities (intelligence, competence, ambition)	$((PP6C-PP6B)+(PP8C-PP8B)+(PP10C-PP10B))/3$

<i>self -assessment</i>		
<i>diff_intelligence</i>	A numerical variable capturing the difference between the intelligence score that the respondent assigns to himself/herself and the score he/she assigns to people of his own gender	PP6A-PP6B/PP6C
<i>diff_competence</i>	A numerical variable capturing the difference between the competence score that the respondent assigns to himself/herself and the score he/she assigns to people of his own gender	PP8A-PP8B/PP8C
<i>diff_ambition</i>	A numerical variable capturing the difference between the ambition score that the respondent assigns to himself/herself and the score he/she assigns to people of his own gender	PP10A-PP10B/PP10C
<i>diff_assess_scores</i>	An average of the differences between the scores that the respondents assigns to themselves and to people of their own gender, across three qualities	$((PP6A-PP6B/C)+(PP8A-PP8B/C)+(PP10A-PP10B/C))/3$
<i>earnings aspirations</i>		
<i>earning_asp</i>	Monthly earnings to which the respondent aspires in ten years	AS12
<i>anticipated discrimination</i>		
<i>woman_earn_less_meca</i>	A dummy variable for whether the respondent believes that women working in auto-mechanics will earn less than men	PP11C==3

<i>woman_earn_less_welder</i>	A dummy variable for whether the respondent believes that women working in welding will earn less than men	PP12C==3
<i>woman_earn_less_hairdresser</i>	A dummy variable for whether the respondent believes that women working in hair-dressing will earn less than men	PP13C==3
<i>woman_earn_less_tailor</i>	A dummy variable for whether the respondent believes that women working in tailoring will earn less than men	PP14C==3
<i>social_sanctions</i>	A score capturing the extent to which the respondent expects discriminations/social sanctions to affect women in male dominated trades	A score 0-3 with one point for each of the following: NG5=0 NG5=1 NG6=2 NG6=3 NG7=2 NG7=3
<i>networks</i>		
<i>netw_male_trade</i>	A dummy variable for whether the respondent knows someone who works in a male dominated trade	AS101A=1 AS102A=1
<i>netw_freq_male_trade</i>	A dummy variable for whether the respondent knows and frequently sees someone who works in a male dominated trade	AS101D=1 AS101D=2 AS102D=1 AS102D=2
<i>netw_fam_male_trade</i>	A dummy variable for whether the respondent has someone in their family working in a male trade	<i>check</i> AS101C=1 AS102C=2
<i>Before coding the next two vars, make sure to replace all the AS105* & AS106* vars for individuals who selected auto-mechanics, welding or tailoring</i>		

<i>insector_mentor</i>	A dummy variable for whether the respondent knows somebody within her community who practices the trade she selected.	AS105A AS106A (depending on whether individual was assigned to first or second choice)
<i>freq_insector_mentor</i>	A dummy variable for whether the respondent knows somebody within her community who practices the trade she selected AND talks to them at least once a month	AS105D=1 AS105D=2 AS106D=1 AS106D=2 (depending on whether individual was assigned to first or second choice)
<i>technical skills</i>		
<i>knowledge_elec</i>	A dummy variable for whether the respondent answered correctly the knowledge question related to electricity	AT1=1
<i>experience_elec</i>	A dummy variable for whether the respondent answered positively the experience question related to electricity	AT2=1
<i>knowledge_clim</i>	A dummy variable for whether the respondent answered correctly the knowledge question related to air conditioning	AT3=1
<i>experience_clim</i>	A dummy variable for whether the respondent answered positively the experience question related to air conditioning	AT4=1
<i>knowledge_automech</i>	A dummy variable for whether the respondent answered correctly the knowledge question related to auto mechanics	AT5=1

<i>experience_automech</i>	A dummy variable for whether the respondent answered positively the experience question related to auto mechanics	AT6=1
<i>knowledge_weld</i>	A dummy variable for whether the respondent answered correctly the knowledge question related to welding	AT7=1
<i>experience_weld</i>	A dummy variable for whether the respondent answered positively the experience question related to welding	AT8=1 & AT9=1
<i>knowledge_carp</i>	A dummy variable for whether the respondent answered correctly the knowledge question related to carpentry	AT10=1
<i>experience_carp</i>	A dummy variable for whether the respondent answered positively the experience question related to carpentry	AT11=1
<i>tech_knowledge_score</i>	Score taking values between 0 and 1 capturing the respondent's level of technical knowledge	$(AT1 + AT3 + AT5 + AT7 + AT10)/10$
<i>tech_exp_score</i>	Score taking values between 0 and 1 capturing the respondent's level of technical experience	$(AT2 + AT4 + AT6 + AT8 + AT9 + AT11)/6$
<i>cognitive skills</i>		
<i>digit_score</i>	A score from 0 to 8 capturing the number of digit sequence that the respondent was able to repeat correctly	CC1 to CC8

<i>raven_score</i>	A score from 0 to 12 capturing the number of raven puzzles for which the respondents was able to provide the correct answer	CC13 to CC24
<i>non cognitive skills</i>		
<i>self_esteem_score</i>	A score from 0 to XX capturing the respondent's degree of self-esteem	CN1 - CN3
<i>job priorities</i>		
<i>top_earnings1</i>	A dummy variable for whether the respondent ranks <i>earnings</i> first among job selection criteria	EM10_1=1
<i>top_earnings2</i>	A dummy variable for whether the respondent ranks <i>earnings</i> first OR second among job selection criteria	EM10_1=1 EM10_1=2
<i>bot_earnings1</i>	A dummy variable for whether the respondent ranks <i>earnings</i> last among job selection criteria	EM10_1=4
<i>bot_earnings2</i>	A dummy variable for whether the respondent ranks <i>earnings</i> last OR second to last among job selection criteria	EM10_1=3 EM10_1=4
<i>top_multitasking1</i>	A dummy variable for whether the respondent ranks <i>multitasking</i> first among job selection criteria	EM10_6=1
<i>top_multitasking2</i>	A dummy variable for whether the respondent ranks <i>multitasking</i> first OR second among job selection criteria	EM10_6=1 EM10_6=2
<i>bot_multitasking1</i>	A dummy variable for whether the respondent ranks <i>multitasking</i> last among job selection criteria	EM10_6=4

<i>bot_multitasking2</i>	A dummy variable for whether the respondent ranks <i>multitasking</i> last OR second to last among job selection criteria	EM10_6=3 EM10_6=4
<i>agency</i>		
<i>empw_score</i>	A score from 0 to 10 capturing the respondent's empowerment level/ability to make his or her own decisions.	CA2a to CA2k
<i>women_empw_score</i>	A score from 0 to 10 capturing the respondent's perception of women's ability to make their own decisions	CA3a to CA3k
<i>Age at first sexual relation</i>		
<i>age_1st_relation</i>	A numerical variable capturing the age at which the respondent had his or her first sexual relation	AS36

ADMINISTRATIVE OR M&E DATA - TRAINING CONDITIONS

Variable	Definition	Source/Construction
<i>choices</i>		
<i>two_choices</i>	A dummy variable for whether the respondent selected two trades	Application Form
<i>assigned_1st_choice</i>	A dummy variable for whether the respondent was assigned to his or her first choice	Application Form
<i>class and training center</i>		
<i>proportion_men</i>	A variable taking values between 0 and 1 capturing the proportion of men in the respondent's class	M&E data
<i>trav_training_center</i>	A variable capturing the travel time between the respondent's home and his/her assigned training center	M&E data (or Midline data)