# Unwilling to reskill? Evidence from a survey experiment with Italian jobseekers

## Pre-Analysis Plan

Alexia Delfino, Andrea Garnero, Sergio Inferrera, Marco Leonardi, Raffaella Sadun

August 4, 2023

## Contents

# 1 Project Summary

In recent decades, labor markets in all OECD countries have experienced profound changes driven by digitalization, globalization, demographic shifts, and the imperative of a rapid net-zero transition. These transformative forces have led to sectoral shifts, altered consumption patterns, and demanded new approaches to work. As a result, workers are faced with the challenge of swiftly updating their skillsets to remain competitive and avoid being left behind in the modernizing labor landscape.

Despite the widespread acknowledgment that training plays a vital role in supporting workers through these dynamic transformations, the uptake of training programs remains disappointingly low. Across the OECD, a mere 40% of adults, on average, engage in job-related formal and non-formal training in any given year (OECD, 2019). Furthermore, participation rates among low-skilled adults are alarmingly 40 percentage points lower than their high-skilled counterparts, raising concerns about potential group-specific barriers hindering access to training opportunities.

In our work we want to study the specific barriers preventing jobseekers from pursuing training, especially for occupations in rapidly growing sectors. Our research investigates the demand for training and reskilling among jobseekers, exploring heterogeneity across various jobseeker groups and different types of training programs. To accomplish this, we gather new data and employ a discrete choice model to quantify the demand for training and reskilling among Italian jobseekers. Additionally, we simulate alternative training policies, comparing the effects of generic training subsidies to targeted training design interventions on training uptake rates, specifically for reskilling initiatives. By identifying and understanding the barriers faced by our participants, we aim to propose effective measures tailored to their needs. This study aspires to provide valuable insights into the design of targeted policies and training subsidies that can bolster the uptake of effective programs and foster successful reallocation of workers into high-growth sectors.

The study focuses on Italy, where the urgency of reskilling is particularly pronounced. The adult population in Italy exhibits lower levels of numeracy and literacy and lower participation rates in work-related training activities compared to other OECD countries. The country also contends with a significant proportion of NEET (Not in Education, Employment, or Training) individuals and a high share of under-qualified workers. Unfortunately, existing training programs are largely disconnected from the actual demands by companies. Training is offered based on old catalogues that are poorly aligned with local labor demand (OECD, 2023) and the quality assurance of training programs in Italy is often minimal or nonexistent.

By shedding light on the barriers to training participation in Italy and proposing data-driven solutions, this research strives to contribute to the effective allocation of investments in training programs and the formulation of policies that promote a dynamic and adaptable labor force capable of thriving amidst ongoing labor market transformations. Ultimately, the findings from this study can serve as a valuable resource for policymakers, researchers, and stakeholders seeking to enhance workforce development strategies and foster economic resilience in an ever-evolving global labor landscape.

# 2 Experimental Design

## 2.1 Sampling and Recruitment of Participants

We conducted the survey with two different samples of Italian jobseekers.

- **Survey-Company Panel**:

  The first sample was recruited through the survey company CE&Co. The company sent out our survey to their registered users in order to have a sample representative of Italian unemployed people in terms of age, regional location and gender.

  Incentives for this sample consisted in tokens that they could redeem from the survey company, for a value of around 5 euros.

- **Jobseekers registered at jobcentres in the Metropolitan area of Milan (Italy)**:

  We work in partnership with AFOL Metropolitana, a consortium of jobcentres in the metropolitan city of Milan. As part of this collaboration, we sent an invitation to participate in our survey to the jobseekers in their registries. We extracted three different samples from their registries:

  1 Sample of jobseekers who registered in one of the consortium's centres up from January to September 2022

  2 Sample of jobseekers who registered in one of the consortium's centres between the 1st of September 2022 and the 28th of February 2023

  3 Sample of jobseekers who registered in one of the consortium's centres between the 1st of March 2023 and the 7th of July 2023

For all the samples, the IT personnel selected only people with a valid email address in their records. For samples 2 and 3, the IT personnel also selected jobseekers without any active employment contract.

The reason for these different data extractions is that people register in the job-centres on a rolling basis, and thus we proceeded to sample potential participants in waves. Moreover, from Septembre 2022 the job-centres started using a new onboarding procedure for jobseekers (which involves samples 2 and 3). For this reasons, we use two distinct survey links for sample 1 vs samples 2/3 (so that assignment to the survey and its treatments are stratified by sample). We will always control for a dummy for being from sample 1 (vs samples 2/3) in the regressions and analyses specified below.

The anonymized lists of samples 1, 2 and 3 were shared with the research team, who then excluded individuals i) of foreign nationality and ii) outside the age range 18-55.

In total, sample 1 was made of 33,965 individuals, sample 2 of 5,838 individuals and sample 3 of 4,421 individuals. According to some piloting we did, we expect a response rate between 1% and 2%. Thus we expect to collect between 400 and 800 answers from the Milan jobseekers' sample.

We offered an Amazon voucher of 5 euros for all the first 500 respondents to the survey.

As of today, some exploratory analysis has already been performed by the research team on the first sample extracted from the survey company panel. This pre-registration is therefore meant to be applied to the second sample, which includes samples 1, 2 and 3 from the job-centres' jobseekers.

## 2.2 Survey sections

The survey instrument consisted of five parts:

1. **Screening**: First, we screened participants on the basis of the eligibility criteria and further asked about the highest education level achieved, whether the person has children and gender. To be eligible to participate, a respondent should be i) between 18 and 55 years old, ii) not working, iii) actively looking for a job and iv) not enrolled in any training course.

2. **Working and training history**: Second, we asked respondents about their past training and work experience, whether they felt their past job was part of their identity, the skills needed in

their previous job, expectations of wage, training effectiveness and job finding rates for the main job they are searching for. We also asked some information about their job search (e.g, hours spent searching, number of occupations).

3. **Attitudes towards new occupation**: Third, we randomly assigned every survey respondent to a "treatment occupation" (IT assistant or construction technician). We provided some basic description of the tasks involved in the occupation through a short video of approximately 30 seconds. We then asked expectations of employer's demand, wages, job finding rates, fit with own identity, confidence in skills and interest for this occupation.

4. **Discrete choice experiment**: Fourth, respondents are asked to choose between six training options with randomized characteristics, as described in subsection 2.4.2.

5. **Demographics and unemployment**: Fifth, we collect further demographic information such as whether the person receives unemployment benefits, duration of unemployment, time spent in household chores and personality traits (risk aversion, trust).

## 2.3 Power analysis

Following Stefanelli and Lukac, 2020, we performed a power analysis setting the relevant parameters for our discrete choice experiment. In discrete choice experiments statistical power is influenced positively by the sample size and the number of "tasks" assigned to each respondent; it is influenced negatively by the number of alternatives for each attribute.

We focus our analysis on the Average Component Marginal Effect (ACME), the most commonly used causal measure of interest that is used to analyse conjoint data. The ACME represents the causal effect of shifting one attribute of a conjoint profile on the probability of choosing that profile while averaging over the distribution of the other profile attributes. For instance, in our setting the ACME represents the change in the probability of choosing one training when the training content changes from "horizontal" to "vertical".

We analyze the required sample to be able to estimate comfortably (i.e. with a target statistical power of at least 80%) the median ACME that Stefanelli and Lukac, 2020 found in their literature review. Such ACME equals 0.05, meaning that a change in one training attribute induces at least 5 percentage points increase in the probability of choosing such training, all else being equal. According to this exercise, our target number of respondents is between 450 and 650 people within each sample to achieve an ACME of 0.05.

Such targets have to be considered as lower bounds: for this analysis, we set the number of levels for each training attribute to 4, while Figure 1 shows that only one attribute has 4 levels. Given that the number of levels for each variable impacts negatively power, and that our attributes have on average 2.5 levels, we consider these targets as conservative.

## 2.4 Randomization

### 2.4.1 Between subjects

We introduce two between-subjects randomized treatments in the surveys:

- **Treatment occupation**:

  After screening, the survey software randomly assigned respondents to one of two high-demand occupations: IT Assistant or Construction Technician. Assignment was stratified by gender. We refer to the occupation assigned to each respondent as "treatment occupation".

- **Experimental treatments**:

  In order to study possible frictions to the demand for training, we introduced experimental variation within the survey right before the discrete choice experiment (part 4). Respondents are randomized into the following three groups:

  1. Information (INFO): before asking for their training choices, respondents are provided with information on vacancies and average wage in the treatment occupation.

  2. Role model and growth mindset (MINDSET): before asking for their training choices, respondents are shown a "success story" of a person who succeeded in reinventing herself after a training and learn about the concept of "growth mindset".

  3. Control group (CONTROL): no additional information is provided before training choices.

  Both the MINDSET and INFO treatments were administered through videos, which had a length of approximately two minutes. Both videos consisted in voice over slides, and had the same voice narrating the text. Viewers could skip the video, but were asked some comprehension questions afterwards. In case of a wrong answer, the survey would recommend them to watch the video again. We recorded the time spent watching the video, as well as the answers to comprehension questions.

These treatments are used to study whether respondents change their training demand and preferences when provided with additional information. We randomized assignment to the three treatments within each gender-occupation cell (e.g., women-IT). This strategy leads to a total of six treatment groups, within each gender.

### 2.4.2 Within subjects: Discrete choice experiment

In the main part of the survey, respondents were asked to choose between training options with randomized characteristics. In particular, each respondent was presented with a series of choices among different training programs that differ in the following features: 1) content, 2) price, 3) subsidy, 4) effectiveness (as the job finding rate of participants in the six months after the training), 5) length, 6) delivery (virtual or in person), 7) structure (modular or intensive). Figure 1 shows all the levels of the different factors used in the discrete choice experiment.

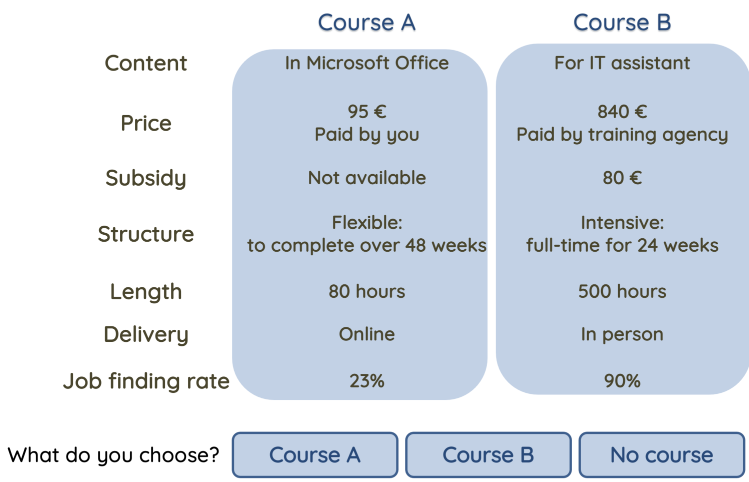
Figure 1: Factor levels

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| **Content** | **Price** | **Subsidy** | **Job Finding Rate** | **Length** | **Delivery** | **Structure** |
| Horizontal | 1.67/h for jobseeker | 1.72/h | 23 | 80 | In-Person | Intensive (Monday-Friday) |
| | 1.17/h for jobseeker | 3.44/h | 50 | 200 | Online | Flexible (within 48 weeks) |
| | 1.17/h for training agency | | 70 | | | |
| | 1.67/h for training agency | | | | | |
| Vertical | 1.67/h for jobseeker | 1.72/h | 50 | 200 | In-Person | Intensive (Monday-Friday) |
| | 1.17/h for jobseeker | 3.44/h | 70 | 500 | Online | Flexible (within 48 weeks) |
| | 1.17/h for training agency | | 90 | | | |
| | 1.67/h for training agency | | | | | |

Each training option is hypothetical, but has been created based on real courses available in Italy through a benchmarking exercise. Importantly, we allow respondents to choose an opt-out option, that is we allow them not to choose any training. After a respondent choose her favourite option, we ask a follow up question: "If you favourite option was not available, which of the remaining two options would you choose?". This allows us to know the full ranking of the three options for each respondent.

Figure 2 shows an example of a choice scenario.

Figure 2: Example of choice scenario




We decided which factors to use in the scenarios through preliminary qualitative work with job-seekers in Italy. Once the features and their levels had been determined, we built choice scenarios using an efficient design (i.e. a design that, compared to other options, minimises the generalized variance of the parameter estimates), through the stata code 'd-create". This method maximises the D-efficiency of the design based on the variance-covariance matrix of the parameter estimates of the conditional logit model (Cook and Nachtsheim, 1980; Zwerina et al., 1996; Carlsson and Martinsson, 2003). This method is employed to avoid using a full factorial design, i.e. a design in which each combination of the proposed factors is presented as an alternative to the respondent. The D-efficient design allows a precise identification of the parameters of interest, without imposing huge cognitive burden to the respondents. We also checked empirically that the design leads to a balanced presentation of levels within and across blocks, as well as no correlation between factor.

We picked 18 binary choice scenarios in total, divided in 3 blocks. Every respondent was randomly assigned to one block only, and thus had to make six choices. We also randomized respondents to see one of three variants of every block, which differed only in the order of presentation of the attributes.

This core survey module allows us to estimate jobseekers' demand for training, as well as their willingness to pay for different features of the courses. For instance, we will be able to understand how much money people are willing to pay for a course which gives better chances of finding a job after participation. Moreover, within each choice one of the training courses is specialized towards a growing occupation (e.g., IT assistant or construction technician). That is, one option is for training, and one is for **reskilling**. This will allow us to explore how to best encourage jobseekers into reskilling towards growing jobs.

### 2.5 Data Collection

For the sample from the survey company panel, surveys were completed between the 6th and 21st fo December 2022. We received a total of 597 responses.

For the sample from AFOL, invitations to the survey were sent on the 6th of May 2023 for subsample 1, on the 6th of April 2023 for subsample 2, and on the 19th of July 2023 for subsample 3. After approximately 2 to 3 weeks from the first invitation email, a reminder was sent for each subsample. While the collection is still ongoing, we collected so far a total of 434 answers.

We will check for completion time of the respondents and quality of answers. We will exclude respondents with completion time less than 20% of the median time and with 75% or more of the questions with an "I don't know" reply (out of the questions that allow to say "I don't know"). For the main analysis, we may also exclude other types of problematic respondents that we can detect (e.g., people with clearly inconsistent answers across parts of the survey). We will show in appendix robustness checks with the inclusion of this last category.

## 3 Measurement of Primary Outcomes

Our primary outcomes are meant to capture interest in the two high-demand occupations we selected, and in undertaking specialized training program to access these occupations (i.e., reskilling).

1) **Interest in treatment occupation**

The survey asked respondents to rate on a scale from 0 to 10 their interest in the treatment occupation. This is our main measure of interest for the treatment job.

**Variables construction**:

1 We will use the raw version of the survey variable, as well as a dummy for having interest above median in the sample.

2 We will also construct six additional versions of the variable for interest in the treatment job by residualizing the survey response to six different set of controls. In particular, we will take the

residuals from the following regressions:

$$interest_i = \beta X_i + \delta_T + \epsilon_i$$

where $\delta_T$ are controls for the randomized treatments of occupation type and information/mindset treatment and $X_i$ is a set which includes the following controls (added progressively):

1 Demographics: gender, age, region.

  * Gender is a dummy equal to 1 for being a woman.
  * Age will be used in continuous form, as well as in a categorical version with four dummies for different age ranges (18-25, 26-35, 36-45, 46+).
  * Region consists in three indicators for North, Centre and South of Italy.

2 Work history: past job, jobs searched, unemployment length, subsidies.

  * An open-ended question asked about the main past job that the respondent had in the past. We manually coded the answers into standard 2-digit SOC categories. We will use indicators for each 2-digit occupation in the regressions.
  * Job searched is a count of how many different occupations the person is searching.
  * Unemployment length is measured in two ways: i) the length - in months - of the current unemployment spell and ii) the number of months that the person has been without a job since finishing her studies. In case of inconsistencies between these two variables, we will use the one whose distribution is more closely aligned with similar populations of jobseekers in Italy, according to academic papers and policy reports. If available, we will use the information from the records of the partner jobcentres.
  * Subsidies is a dummy equal to one if the respondent is the recipient of any unemployment government support. We may distinguish between different types of schemes (with one dummy for each) if there is enough variation in the sample.

3 Skills: actual and perceived skills

  * For perceived skills, we asked the following question: "To what extent do you think you'd be able to do this job with your current skills? (0-10)"

    For actual skills, we presented respondents with a list of skills that are required in the treatment occupation (from $O^*$Net) and asked to select all of those that they have.

As measure of actual skills, we compute the share of skills selected out of the total displayed.

4 Expectations on treatment job: expected wage, hiring likelihood, demand by employers

* For expectations of being hired, we asked: "Regardless of how interested you are, if you were to apply for this job, how likely are you to find it in the next three months?"
* For wage expectations, we asked: "Now imagine being hired for this job. What net[1] monthly salary do you think you would be offered?"
* For expectations of employers' demand, we asked: "What do you think is employers' demand for employees in this occupation?". Respondents could select Very High, High, Low, Very Low, I don't know.

5 Self-identification: fit with treatment job and past occupation

* To what extent do you think that this job corresponds to the view you have of yourself? (0-10)
* We also asked a four-item scale to measure the extent to which the respondents' past jobs contributed to shaping their self-perception. The statements were: i) I frequently talk about this job with family and friends, ii) I am proud of doing this job, iii) This job is part of my identity, iv) My family and friends identify myself with this job. For each statement, the respondent had to answer whether they Strongly Agree, Agree, Disagree or Strongly Disagree. We sum up all the responses across the four statements to build a single index.

6 Training expectations: perceived increase in wage and hiring with training

* For expectations of hiring, we asked two additional questions: "Suppose you take part, for the next three months, in a training course that prepares you for the [treatment occupation]. In your opinion, upon completion of the training course, what would be the probability of finding this job?" and another question in which the course also provides a certification recognized by firms. We compute two variables from these questions: the difference in the likelihood of finding the job with the training course or without, and the difference in the likelihood of finding the job with the course which provides a certification recognized by firms or with a simple course.

[1] In Italy, wages are taxed at the source and, therefore, people tend to know better their net wage which is reported at the bottom of their payslip rather the gross one from which taxes and social security contributions are deducted.

* We repeat the same set of questions as described above also for wage expectations. We compute two variables from these questions: the difference in the expected wage with the course or without, and the difference in the expected wage with the certified course or with a simple course.

For the regression, we will code as 0 responses which are missing (e.g., people replied "I don't know") and add a dummy for when the variable is missing to the regression.

2) **Interest in reskilling**

We will define interest in reskilling by using respondents' answers to the six different discrete-choice scenarios they saw. Reskilling for us means preferring a vertical training (i.e., specific training preparing for an in-demand occupation) compared to a generalist training, or no training at all.

**Variables construction**:

- First, for every respondent we will compute two variables of interest in reskilling: i) the share of scenarios (out of six) in which s/he picked a vertical course as first choice and ii) whether s/he ever picked a vertical course as first choice in any of the given scenarios. We will also compute the analogous variables for the horizontal training option and for the opt-out.
- Second, we will also explore interest in reskilling by computing the average ranking given to the vertical training, horizontal training and opt-out by a respondent across scenarios.
- Third, we will quantify the increase in marginal utility and the willingness to pay for a vertical training course in our discrete-choice model described below.

# 4 Empirical strategy

We will divide our empirical analysis in five main parts: i) Determining which factors account for the interest in a treatment occupations, ii) Determining which factors account for interest in reskilling, iii) Estimating the preferences for specific types of training programs and their features through the discrete choice analysis, iv) Building policy counterfactuals, v) Examining the relationship between choices made in the discrete choice analysis exercise and real labor and training outcomes. We describe the analysis and main hypotheses in the following subsections.

## 4.1 What drives the interest in a treatment occupation?

We will explore different factors that correlate with interest in the treatment occupation.

In particular, we will look at the following factors:

- Demographics: gender, age, region.
- Work history: Past occupation, past job, jobs searched, unemployment length, subsidies.
- Perceived and actual level of skill: we will combine these variables in an "index of skills"
- Expectations of being hired, wage and employers' demand: we will combine these variables in an "index of expectations"
- Identification with the occupation: we will combine these variables in an "index of job identification"
- Training expectations: we will combine these variables in an "index of training expectations"

The corresponding variables have been defined in section 3. All indices indicated above will be created following Kling et al. (2007).

## 4.2 What drives the interest in reskilling?

### 4.2.1 Correlates

We will explore different correlates of interest in reskilling.

- First, we will look at the correlation between reskilling outcomes and the different factors that explain interest in the treatment occupation.
- Second, we will explore differences in reskilling outcomes between people with above median or below median interest in the treatment occupation. We will consider all our different definitions of interest in the treatment occupation for this exercise.

### 4.2.2 Determinants

To explore determinants, we will look at the effects of our randomized treatments: INFO and MINDSET.

The main regression of interest is:

$$reskilling_i = \beta_1 INFO + \beta_2 MINDSET + \delta X_i + \epsilon_i$$

where $reskilling_i$ is one of our primary measures of reskilling, and $X_i$ includes demographics, work and training history, and treatment occupation (variables as defined in section 3). We will also present a model where we use Post-Double Selection LASSO (Belloni et al., 2014) to choose controls out of the full set of variables asked before treatment (all those listed in section 3, including interest in treatment occupation). We will also present a regression with repeated observations ($j = 6$) for each person, where the outcome $reskilling_{ij}$ is whether the person picked the vertical training in each $j^{th}$ choice. The specification will be the same as before, but with clustered standard errors at the individual level.

To deal with inattentive participants, we will perform four exercises:

- We will limit the sample of the regression to respondents between the $10^{th}$ and $90^{th}$ percentile of time spent on the video screen
- We will limit the sample of the regression to respondents who answered correctly to post-video understanding checks
- We will add analytical weights for the time spent on the video
- We will limit the sample to respondents who watched entirely the video

### 4.3 Discrete choice analysis

As outlined in subsection 2.4.2, we ask respondents to choose between different types of training programs, and randomly vary the characteristics of each scenario program. This allows us to estimate a discrete choice model as outlined in Train, 2009. We will be exploring the same observable factors described in subsections 4.1 and 4.2, since we hypothesize that they ultimately influence choice.

The objective of the empirical analysis is to estimate the parameters (marginal utilities) of the discrete choice model through different logit specifications (Train, 2009), where the parameters of interest will be the features of each training (see section 2.4.2 for a complete list). In particular, we will estimate the following models:

1. Simple conditional logit: we will use these estimates to generate baseline results on the average demand for training and training features.

2. Mixed logit: we will use this model to introduce taste heterogeneity between respondents.

3. Other extensions: We will use other related approaches to test alternative models for the choice process. For example, we will use the nested logit specification, which assumes that respondents first choose whether to opt-in or to opt-out, and then decide which training to undertake conditional on having opted-in.

The ratio of the estimated parameters to the parameter associated with the cost of the training program measures the Willingness to Pay (WTP) for each feature of the training program.

In these estimates, we will always be controlling for an opt-out specific constant as well as a dummy for the block of scenarios. We will also control for the stratification variables and the main controls outlined in section 3. We will also perform other robustness checks, for example controlling for the ordering of displayed factors that the respondents were assigned to.

Finally, we will exploit the richness of our data to explore the heterogeneity of willingness to pay for training's features with respect to observable characteristics of the respondents. For instance, we will analyze the gender and age differences in willingness to pay for reskilling.

### 4.4 Policy counterfactuals

We will explore policy counterfactuals using the results of the discrete choice model. In particular, we will estimate predicted market shares (Rao, 2014) of different types of training courses, using existing courses in Italy, to build alternative types of feature combinations. The objective of this empirical investigation is to estimate how the market shares of vertical and horizontal training, as well as the market share of the opt-out option, may be affected by different kinds of policy interventions. For example, we will compare the effects of generic training subsidies to targeted training design interventions on training uptake rates, specifically for reskilling initiatives.

### 4.5 Real outcomes

We will explore the relationship between our primary outcomes and real-life behavior in two ways.

1. At the end of the survey with the AFOL sample, we asked two questions which are meant to capture real-life interests in training. First, we asked respondents whether they were interested in training course within Lombardy (the region where Milan is located). If they replied "Yes" or "Maybe", we followed up by showing them a list of real course offered in Lombardy. We asked

them to assign 10 points across these courses, to express their preferences for the different types of courses. To encourage truthful answers, we added that "We could communicate your choices to your AFOL caseworker, in order to help your choice of courses in the Region." We are going to construct three main variables from these questions:

- A dummy for whether the respondent is interested in courses in Lombardy (Yes or Maybe);
- The average number of points (out of 10) assigned to vertical courses for in-demand occupations;
- Whether the person assigns any point to a vertical course for in-demand occupations

We are going to correlate our primary outcomes (i.e., interest in the treatment occupation and reskilling) with each of these variables.

2. We will match our survey sample with administrative data on training history, training choices and occupational outcomes. The details of this match are yet to be defined, as they depend on data availability from the partner organization and related legal agreements. We will upload an addendum to this PAP once the variables available become clearer with the field partner.

### 4.5.1 Multiple hypothesis testing correction

If multiple measures of real behavior will be available to us, we will use false discovery rate corrections (Benjamini, 2006) to account for multiple hypothesis testing across our outcome variables. Therefore, for each hypothesis test, we will report two values:

1. The usual $p$-value from a Wald test;
2. False discovery Rate $q$-values, taken across primary outcomes.

We will do FDR corrections separately for training and occupational outcomes, reflecting out belief that results in each domain are of separate interest.

## 5 Heterogeneity

Depending on sample size, we will explore heterogeneity along the lines of gender, age, education, interest in treatment job and treatment occupation.

## 5.1 Analysis by sample

The analyses specified in this PAP are meant to be applied to the AFOL sample, which we suspect that will differ from the survey company sample of respondents along many dimensions (e.g., occupational history, job search, education). We may, however, explore the possibility of merging together both samples for a unique set of analyses, showing the split between the two and their differences in Appendix tables.