

**Pre-Analysis Plan (1) for the Baseline and Follow-up Surveys for
“Encouraging hands-on job experimentation among teenagers (experiment 2)”**

Anne Brenøe, Alexia Delfino, Claudio Schilter, Stefan Wolter

November 3, 2023

1. Introduction

This document pre-specifies the analysis we will conduct with data from the baseline and two follow-up surveys immediately after treatment and three weeks later. We will submit separate pre-analysis plans for the outcomes from administrative data and the endline survey.

This study aims to understand how experimentation affects job search and career choices. We hypothesize that encouraging teenagers' experimentation of different occupations will affect their job search and ultimately help them make better choices for their career. To test this hypothesis, we designed a field experiment in collaboration with firms where the treatment consists of experiencing the real-life work environment in an occupation that the student had not considered. We compare the treated group to a control group that instead experience an occupation similar (or identical) to their preferred occupation.

We have already registered and submitted a pre-analysis plan (PAP) about a similar experiment we conduct with schools. This PAP is for a second experiment that will be conducted in collaboration with firms.

2. Description of the intervention

To test whether experimentation with the real-life work environment in non-considered occupations affects students' occupational search, we designed the following intervention in collaboration with several large firms in Switzerland.

All our partner firms regularly offer trial apprenticeships (TA), which are days during which students (typically in grade 8) can shadow and/or try-out tasks in a given occupation within the firm. The idea of TAs is to help students better understand their preferences before applying for apprenticeships during grade 9.

We collaborated with our partner firms to co-design a new model of TA which enables students to experiment with occupations which they did not necessarily consider before (called "Vorschnupperlehre"/"Pre-TA").

Students apply for this TA online, with the goal of learning more about one precise occupation. However, they know from the ad that they would experience their targeted occupation for half the time and could experience a different occupation (potentially similar, potentially different) for the other half.

Our experimental sample will consist of all the students who apply and are accepted by the company for this TA. Students who are randomized into the **control** arm will experience the occupation of their choice, and then a similar or identical occupation for the rest of the time.

Students who are randomized into the **treatment** arm will similarly get to experience their targeted occupation for the first half of the TA. However, for the second half, they will experience an occupation which is different than the targeted one. In other words, the **treatment** is to experience an occupation they had - most likely - not considered before.

To make sure to get variation in the types of occupations experienced in the treatment group, we categorized all occupations along different margins. We describe this categorization procedure in the next paragraphs.

Categorizing occupations

We categorize all occupations by their gender composition and type. For *gender composition*, we defined three categories: i) male-dominated (with <33.33 percent females among apprenticeship graduates from 2019 to 2021), ii) female-dominated (with >66.66 percent female graduates) and iii) neutral (between the previous two). We construct the average share of female graduates for the period 2019–2021, using administrative educational data (Längsschnittanalysen im Bildungsbereich; LABB). In terms of *type*, we categorize all occupations as being predominantly working with computers, predominantly working with hands and/or machines/tools, or predominantly working with people like clients and patients. To do so, we have carefully read the descriptions of the apprenticeships on the governmental website for apprenticeship information and career advice (www.berufsberatung.ch) and manually characterized apprenticeships as predominantly involving computers, machines/hands, or in-person contact with people (e.g., patients and clients).

In theory, we would then have 3*3 categories which combine the gender-composition criterion and the type. However, some categories do not exist in practice within a firm.

From each of our partner firms, we got a list of all the occupations available for TAs and classified them into one of our categories. We then let the firms know which pairings of occupations for the control group and for the treatment group are possible by matching occupations belonging to the same category (for the control group) and belonging to different categories (for the treatment group). While the number of pairings is potentially large, in practice this was reduced by each firm's logistical, geographical, and scheduling constraints. We also make sure that the afternoon-occupations are not very different in terms of requirements, to rule out that participants that applied for occupations with low requirements see occupations with high requirements in the afternoon and vice versa. The common idea, however, is that a student in the control group is assigned to a pair of occupations within the same category. For instance, if a student applies for a TA in “logistics” and is assigned to the control group, s/he may be randomly assigned to experience the pair of occupations “logistics” (in the first half of the TA) and “custodian” (in the second half).

A student in the treatment group is assigned to a pair of occupations belonging to different categories. For instance, if a student applies for a TA in “logistics” and is assigned to the treatment group, s/he will be randomly assigned to see the pair of occupations “logistics” (in the first half of the TA) and “commerce” (in the second half).

The exact pair of occupations assigned to a student is random but depends on the date of the TA assigned to him/her and the location of the student. Specifically, firms provide dates to us on which they have the capacity to show our participants a morning occupation, a control occupation, and a treatment occupation. There will be different combinations for different

dates, but each morning occupation of a specific date in a specific company commonly has only one treatment and one control occupation for the afternoon. The companies will only ever mention the morning occupation to potential applicants (all they know about the afternoon is that it will be another occupation in the same company). Students only learn during the morning (usually just before lunch) which occupation they will experience in the afternoon.

3. Recruitment of participants, randomization, and timeline

The field experiment is implemented through our collaboration with large firms operating in Switzerland. The firms agreed to conduct a baseline survey with the participants in the morning of the TA (before treatment is revealed).

3.1 Sample

We conduct the intervention among 8th grade students. Most of our sample will do an apprenticeship after compulsory schooling (9th grade).

Currently, we have three large companies on board for the years 2023/2024 and 2024/2025. Our current agreements with partner firms allow us to have 1200 TA slots in total, for an expected sample size (if compliance is perfect) of 1200 students. However, the final sample size may differ depending on take-up (i.e., some students may not show up to the TA) and logistical constraints that may arise in the companies (e.g., staff shortages in managing the TA).

We aim to reach a sample of at least 900 students with the current set of partner companies. Should the sample be lower than this, we will engage in actively looking for additional partners or extending by one year the project.

Timeline

After students apply for the TA, we randomize their treatment status in Stata. This will then be communicated to our partner firms.

At the beginning of the day of the TA and as soon as they get to the firm, students fill in the baseline survey. This first survey takes approximately 10 minutes. In the baseline survey, students answer questions related to their occupational preferences, school grades, personality traits, and family background.

At the end of the TA, we give students a short survey to get their immediate reactions to the experience they went through. After 3 weeks, we invite participants to a second follow-up survey. Moreover, we intend to conduct an endline survey at the end of 9th grade and to merge our survey data to administrative data on trial apprenticeship and apprenticeship applications submitted through a popular online apprenticeship platform. Most students only sign up to this platform in the second half of 8th grade and only start applying in the spring of 8th grade. We will also receive data from our partner firms about applications to their trial apprenticeships and apprenticeships. We will submit a separate PAP for these outcomes. Finally, we intend to merge our survey data to administrative educational data from the Swiss Statistical Office, for which we will submit a separate PAP.

3.2 Randomization

Randomization of the treatment will be done at the individual level. We stratify randomization by event (i.e., the combination of firm, location, date, and morning occupation), and gender. We will randomize half the participants into the control arm and the other half into the treatment arm (unless there are specific constraints for an event which require a smaller share of people in the treatment, but we consider these cases as exceptions). We will stratify the randomization by event and gender using lists provided by the partner firm before the event. In case of late signups (i.e., people not present in the initial list), we will randomize them stratifying only by event. Randomization is done in Stata by the research team.

4. Primary outcomes

4.1 Hypotheses

The intervention is hypothesized to:

- a) Increase the breadth of students' occupational search (search breadth).
- b) Increase students' interest in apprenticeships outside their baseline preferences (apprenticeship type).

4.2 Primary Outcomes

We have two families of primary outcomes, which are both measured in the follow-up survey. For all continuous outcome variables, we throughout standardize them to have a mean of zero and a standard deviation of one for the control group.

Search breadth: We ask two questions:

- Since the Pre-TA at *FIRM*, for which occupations did you search for information?
- Since the Pre-TA at *FIRM*, for which professions did you apply or are you planning to apply for a trial apprenticeship, an information event, or an apprenticeship?

For each of these questions, the respondent can select occupations from a list, write occupations in a free text box, as well as report none. Based on these questions, we construct two primary outcomes within this family:

- i. Dummy equal to 1 if the person searches for any occupation different from the occupation they experienced in the morning of the TA (*outside preferences*).
- ii. Number of occupations they search for (*breadth*). To count the number of occupations the person searches for across the two questions, we only count each distinct occupation once even if the person might indicate the same occupation in their answer to both questions.

We hypothesize that treatment increases the value of these two outcomes.

As a robustness of the outside preferences outcome, we will also consider categories rather than the exact occupations, for both outcome variables. So the first outcome will become a dummy equal to 1 if the person searches for any occupation **in categories** outside of the **category of the morning occupation**. The second outcome will become the number of categories the person searches for.

Type index: Using the same questions as for the search breadth outcomes, we merge the selected occupations to data with information about the female share and math requirements. For all occupations they have searched for across the two questions, we construct the average of the female share, math requirements, and the share of searched occupations of different type than the morning occupation (in terms of being mainly computer/hands or machines/people). We construct a co-variance weighted summary index using the following variables:

- i. abs(average female share in morning occupation – average female share searched)
- ii. abs(average math requirements in morning occupation – average math requirements searched)
- iii. the share of searched occupations of different type than morning occupation.

We hypothesize that treatment increases the value of the type index, i.e., that the occupations treated students search for are more different from the morning TA occupation than the occupations control students search for.

For robustness analyses, we will also consider the primary outcomes defined for each of the questions separately as well as for each of the three dimensions in the type index.

4.3 Adjustments for multiple hypotheses testing (MHT)

Following Benjamini, Krieger, and Yekutieli (2006), we will use false discovery rate corrections to account for multiple hypothesis testing across our primary and secondary 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 or secondary outcomes.

We will do FDR corrections separately for primary and secondary outcomes.

Empirical specification

Our objective is to measure the effects of the intervention on search behaviors. For each of the primary outcomes, we estimate the following specification at the individual level:

$$Y_i = \beta_0 + \beta_1 Treat_i + \beta_2 X_i + \theta S_i + \alpha Female_i + \varepsilon_i,$$

where Y_i is the outcome of interest for student i . $Treat_i$ is an indicator for treatment status, X_i is a vector of baseline covariates, S_i are event fixed effects (dummies for the exact morning event), and ε_i is an error term (we use robust standard errors). Our main specification will only include event fixed effects and a female dummy.

As robustness, we will use post-double selection LASSO to determine the variables in X_i for the specifications (Belloni, Chernozhukov, and Hansen, 2014). For the LASSO, we will use the full set of baseline variables available and include interactions as well as quadratic terms.

5. Secondary analysis

6.1 Secondary outcomes

As secondary outcomes, we will examine potential mechanisms of the treatment effects.

To study potential channels through which the treatment affects our primary outcomes, we study the following outcomes:

- a) *Beliefs about own skills fit:* We ask, “How much do you think your skills would fit to an apprenticeship as...?” with answer options: not at all (1)/little (2)/moderately (3)/much (4)/very much (5). Each respondent answers this question to the treatment occupation (T occ). We standardize this measure, such that a higher value indicates stronger beliefs in own skills fit.
- b) *Beliefs about work tasks:* We ask, “How much do you think you would like the work tasks if you did an apprenticeship in...?” with answer options: not at all (1)/little (2)/moderately (3)/much (4)/very much (5). Each respondent answers this question to the treatment occupation (T occ). We standardize this measure, such that a higher value indicates stronger beliefs in enjoyment of work tasks.
- c) *Beliefs about work environment:* We ask, “Do you think you would get along well with your colleagues if you did an apprenticeship in...?” with answer options: not at all (1)/little (2)/moderately (3)/much (4)/very much (5). Each respondent answers this question to the treatment occupation (T occ). We standardize this measure, such that a higher value indicates stronger beliefs in getting along well with work colleagues.
- d) *Beliefs about employer demand:* We ask, “If you were to apply for the following apprenticeships in FIRM, how likely do you think you would get an offer?” with answer options: Very unlikely (1) /unlikely (2) / moderately (3)/ likely (4)/ very likely (5). Each respondent answers this question to the treatment occupation (T occ). We standardize this measure, such that a higher value indicates stronger beliefs in employer demand.

Questions a) to c) are also asked in the follow-up survey at the end of the TA day, and we will combine the answers in the two surveys to study these outcomes. As exploratory analysis, we also construct the same measures about control (i.e., afternoon) occupations of their TA event.

6.2 Secondary heterogeneity analysis

We will examine whether there is heterogeneity in the treatment effects by the following dimensions:

- a) Gender: male vs female
- b) Preference for gender composition: based on the morning occupation, we define students to prefer predominantly male, neutral, and predominantly female.
- c) Preference for type of occupation: based on the morning occupation, we define students to prefer computers, working with people, or working with hands and/or machines/tools.
- d) Math skills

6.3 Explorative analysis

Additional heterogeneity

As explorative analysis, we are also planning to study whether there is heterogeneity in the treatment effects by immigrant background. For this purpose, we do not count Germans and Austrians as migrants as they neither face a language barrier, nor are they unfamiliar with a relatively common apprenticeship system. So we define students to have an immigration background if both parents are not born in Switzerland, Germany, or Austria vs at least one parent being born in these countries. In addition, we will analyze the heterogeneity by socio-economic status (defined as none of the parents having a college or university degree vs at least one parent having such degree).

We will explore heterogeneity by the student's baseline personality, considering characteristics such as extraversion and social conformity.

We will further explore whether features of the TA event have heterogeneous effects on students' outcomes. In particular, we will explore the effect of the average TA group size and the gender composition of the student group.

Additional outcomes

We plan to explore the answers given by students in the follow-up administered at the end of the TA day, which include questions such as "Would you recommend the two half-day trial apprenticeships to a friend?", "Do you think your experiences today will make you change your mind about the occupation you target for your future apprenticeship?", "Overall, how useful do you find the information you received and your experiences?", and "How much would you like an apprenticeship in the following occupations?".

We will also consider the following outcome: whether students change their beliefs on what people similar to them should do in terms of job search. We elicit this outcome through a vignette, where we randomize whether the vignette student is called Lara (female) or Lukas (male).

We also plan to study the effects on an alternative definition of the type index, where we instead of considering the deviation from the morning occupation consider the deviation from the treatment occupation (T occ):

- i. $\text{abs}(\text{female share in T occ} - \text{average female share searched})$
- ii. $\text{abs}(\text{math requirements in T occ} - \text{average math requirements searched})$
- iii. $1 - (\text{share of searched occupations of same type as T occ})$

We will also check whether students become more accurate in their knowledge of the occupations they experience by comparing the answers they give in the surveys about the occupation with the answers given by employees about the same occupational features.

6. Power Calculations

We target the recruitment of 1200 students and plan to assign half to treatment and half to control. This implies that with 80 percent power and 5 percent statistical significance, the minimum detectable effect of the treatment is 0.16 standard deviations.