**Pre-Analysis Plan for the Baseline and First Follow-up Surveys for**

**"Encouraging hands-on job experimentation among teenagers"**

*Anne Brenøe, Alexia Delfino, Claudio Schilter, Stefan Wolter*

September 22, 2023

## 1. Introduction

This document pre-specifies the analysis we will conduct with data from the baseline and two-month follow-up surveys. 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 schools and 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 of experiencing a specific occupation attends a workshop in school related to their search for an apprenticeship.

The randomized controlled trial (RCT) will be implemented among 8$^{th}$ grade students in the Canton of Lucerne, Switzerland. In this setting, around two-thirds of a birth cohort will do an apprenticeship after compulsory education (typically starting after 9$^{th}$ grade). Although there are various possibilities for further training and education after the completion of an apprenticeship, the initial choice of the occupation of the apprenticeship is a strong predictor of individuals' later occupation on the labor market.

## 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 schools and firms in the Canton of Lucerne, Switzerland.

We collaborate with schools in multiple towns. The exact number is to be defined, as we are recruiting locations on a rolling basis over time. In each town, the school organizes a one-day school event (henceforth *event*) devoted to occupational choice. During this event, students visit 4-5 different firms where they get to experience one occupation in each firm. We refer to each of these visits as a trial apprenticeship (TA). In addition, students also attend 1-2 workshops during the event. Each TA and workshop have a time slot of around 50 minutes and are attended by a group of students (i.e., not just one student at a time). The group is usually supervised by a teacher or volunteering parent during every TA. Each student has an individual schedule with 6 slots. Thus, in most cases, the individual student will not attend multiple TAs or workshops with the same peers.

The workshops take place in a classroom and can include, for example, a current apprentice and/or apprenticeship supervisor talking about apprenticeships and apprenticeship search or practicing making a phone call to inquire about day-long or longer trial apprenticeships (which are commonly done in the months after the event). Students typically walk or cycle from location to location. If the locations are far apart, they are accompanied by a teacher or volunteering parent and use either public transport or the teacher's/volunteer's car. There are many different TAs, some of which are offered more than once, and there are two different workshops that are typically repeated 6 times (as there are 6 timeslots in the day). This depends on the location.

Schools collaborate with local firms to create all TAs and often also for one of the two workshops (e.g., the person talking about apprenticeships will be employed at a local firm). Local firms are asked whether they want to offer any TAs and for how many slots and students per slot they have capacity. Before the event, students get a list of all available TAs in their town and indicate 5-6 TAs they would like to experience during the event (5 is the norm, 6 can be required in larger towns). To coordinate all the visits, schools use a computer program with a specific algorithm to allocate students to firms based on students' preferences and firms' availability. The algorithm does not take into account the ordering of their preferences. The two workshops are typically unrestricted in size and work as a buffer to mitigate logistical difficulties in creating the program.

Before we started collaborating with schools, students would always experience 4 TAs and both workshops. However, we have been collaborating with the IT company that provides the matching algorithm to make it possible to randomize some students to experience a 5$^{th}$ TA.

Students who are randomized into the **control** arm will experience the event as it has been so far, i.e., they get to experience 4 TAs and 2 workshops; these 4 TAs are randomly selected from the list of 5-6 occupations they indicated at signup conditional on some constraints. Thus, control students only get to experience occupations they are interested in. Students who are randomized into the treatment arm will similarly get to experience 4 TAs they signed up for and 1 workshop. On top of this, treated students will experience a 5$^{th}$ TA (instead of a 2$^{nd}$ workshop) which they did not sign up for. In other words, the **treatment** is to experience an occupation they had 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 (female-computers, male-people), so we end up with 7 occupational categories.

In addition, we also take into account how demanding an apprenticeship is and how skilled students are. To measure how demanding an apprenticeship is, we use data on the math requirements from https://www.anforderungsprofile.ch/. We have bought their proprietary data where experts rate these requirements on a continuous scale from 0-100 for most apprenticeships. We classify job requirements as low, medium, and high. We define occupations with high requirements to be those with math requirements at or above the level of the IT apprenticeship (Informatiker*in). We define occupations with low requirements to be those with math requirements strictly below the level of the medical secretary and assistant apprenticeship (Medizinische Praxis Assistant). All occupations that fall in between are classified as medium.

To classify students' skills, we consider students to be highly skilled in math if they take the math class of the highest track (track A) and their grade is strictly higher than 4 (the passing grade – i.e., they are better than just on the verge of passing). We classify students to have low math skills if they take the math class of the lowest track (track C) or they take the medium track (track B) math class and have a grade of 4 or lower. Everybody else is classified to have medium math skills.

To assign the treatment occupation, we first consider the 5 to 6 occupations that the student ranks as their favorites and assign them to the corresponding occupational categories we defined. These are the categories that the student is interested in. To pick a treatment occupation, we then consider all the available TAs[1] within their town from their non-preferred categories that are in their own or an adjacent skill category (i.e., we do not consider "low" occupations for "high" students or "high" occupations for "low" students. However, we can consider "medium" occupations for "high" or "low" students). We then randomly choose 2 of the remaining occupational categories. If there are fewer than four TAs available in these categories, we add another occupational category. In case no occupational categories are left and we have still fewer than four TAs, we add back the category they ranked the lowest among their 5-6 choices (and repeat that for the second lowest in the rare case that there are still less than 4 TAs left). The student can then veto up to two of these TAs. We then randomly select one TA among the non-vetoed TAs in these non-preferred categories to be the 5$^{th}$ TA which treated students experience during the event. We also pick a backup treatment-TA which we could use in case there is an organizational problem to use a specific TA.

[1] For categories that have a large number of TAs (usually this is only the hand/machine-male category), we reduce the number of possible TAs in order not to overcrowed the shortlist that students are presented to veto (see below). To reduce the number of TAs, we drop those with the most extreme gender shares and those that are relatively uncommon apprenticeships. Across all categories, we also considered the opinion of the organizers to drop or not drop specific occupations.

## 3. Recruitment of participants, randomization, and timeline

The field experiment is implemented through schools and the employer association of the Canton of Lucerne (KGL), who owns the software and operates the events. The schools conduct our surveys in class and agree to our design, while KGL opens their software for our changes and creates schedules etc. taking into account our research design.

### *3.1 Sample*

We conduct the intervention among 8th grade students in the school tracks C (low) and AB (medium-high). We do not include students from the academic high school (Gymnasium). Most of our sample will do an apprenticeship after compulsory schooling (9th grade).

Currently, we have two towns on board for the school year 2023/2024, we aim to have around 6 town in 2024/2025, and at least as many in 2025/2026. We might be able to convince more towns to join. With the current number of towns on board and with an assumed participation rate of 90-95 percent of the students, we expect 350 participants in the school year 2023/2024, and around 1,000 in 2024/2025 as well as 2025/2026. Thus, our expected sample size is 2,300 students. Should we be successful in recruiting more schools, we will aim for a maximum sample size of 3,000 students.

### *3.2 Timeline*

We will start with a soft launch of the study in two towns in the fall 2023. In both towns, the event will take place on September 26th, 2023. We will implement the hard launch in the school year 2024/2025 and repeat it in the following school year.

Some weeks before the school event, students fill in the baseline survey, typically during their school subject devoted to occupational choice. This first survey takes approximately 15 minutes. In the baseline survey, students give consent to participate in the study, answer questions related to their occupational preferences, school grades, personality traits, gender, and family background. Importantly for the event, the survey also includes a module where students indicate 5-6 jobs they would like to experience during the event as well as the jobs they veto in case of treatment (see above).

After students have filled in the baseline survey, we randomize their treatment status and allocated additional TA (and backup TA) in Stata. Their 5-6 preferences along with our randomized additional TA enters KGL's IT program that then allocates students such that they see 4 of their preferences and – if in treatment – the additional treatment TA.

The school event commonly takes place in late September, but schools can choose to hold it a few weeks earlier or later.

In the first half of November, there is a regional occupational fair (called Zebi) lasting several days where around 80 percent of school classes will attend. Approximately a week before the fair, we will send a message (via email/sms) to all students to remind them about their experiences at the school event.

Approximately a week after the fair, i.e., about two months after the school event, we invite students to fill in a follow up survey. They fill in the survey in school, but we will also allow them to fill it in at home in case they do not attend school the day the rest of the class fills it in.

At the end of 9th grade, we invite them to a brief endline survey, for which we will submit a separate PAP. Moreover, we intend 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, reason for which 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.

### *3.3 Randomization*

Randomization of the treatment will be done at the individual level. We stratify randomization by school, classroom, and gender. We randomize half the participants into the control arm and the other half into the treatment arm. 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 in November of 8th grade. 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 three questions: "Which occupations did you look into during Zebi?", "Please think about the time between the school event and Zebi. For which occupations did you search for information? *Searching for information includes activities, such as spending some time on the internet to know more about the occupation and talking with someone who is working in that occupation.*", and "For which professions are you planning to apply for a trial apprenticeship, an information event, or an apprenticeship?". For each of the three questions, the respondent can select occupations from a list, write occupations in a free text box, as well as report none. Based on these three questions, we construct two primary outcomes within this family:

i. Dummy equal to 1 if the person searches for any occupation outside of the occupations they listed in their preferences for the event (*outside preferences*).

ii. Number of occupations they search for (*breadth*). To count the number of occupations the person searches for across the three questions, we only count each distinct occupation once even if the person might indicate the same occupation in their answer to each of the three 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 **categories** of the top 3 occupations they listed in their preferences. 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 three questions, we construct the average of the female share, math requirements, and the share of searched occupations of different type than occupations they listed in their preferences for the event 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 occupations they listed in their preferences for the event – average female share searched)
ii. abs(average math requirements in occupations they listed in their preferences for the event – average math requirements searched)
iii. the share of searched occupations of different type than occupations they listed in their preferences for the event

We hypothesize that treatment increases the value of the type index, i.e., that the occupations treated students search for are more different from their preferred occupations at signup to the event than the occupations control students search for.

For robustness analyses, we will also consider the primary outcomes defined for each of the three 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:

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

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 strata fixed effects, and ε_i is an error term (we use robust standard errors). The baseline characteristics that can be included in X_i are:

a) indicators for each of the categories they have in their preferences for the event,
b) indicators for having done 1, 2, or 3+ TAs (with 0 being the omitted category),
c) indicators for planning to do 1-2, 3-5, 6+ TAs in 8th grade (with 0 being the omitted category),
d) dummy for vetoing any occupation for the event,
e) indicators for having low or medium math skills (with high math being the omitted category; with the definition of skill level as above in Section 2),
f) indicators for having low or medium language skills (with high language being the omitted category; with the definition of skill level as for math skills),
g) indicators for parental gender norm attitudes as terciles (sum of two statements),
h) dummy for stating that parents are strongly or very strongly involved in occupational choice,
i) dummy for not having any parent with a university degree,
j) indicators for having one immigrant parent, two immigrant parents, both parents are from non-German speaking countries,
k) dummy for being old for grade,
l) dummy for being on time for grade and having done 2 years of kindergarten (in contrast to 1 year),
m) dummies for being above the median for the personality traits: extrovert (sum of two statements), social conformity (sum of two statements), risk preference, and competitiveness.

As a robustness check, we will also use post-double selection LASSO to determine the variables in X_i. We will also use, following Belloni, Chernozkuhov, and Hansen (ReStud 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 T occ and

one randomly selected occupation from the same category as the T occ. To create an index, we average the answers to the two apprenticeships and then standardize it. A higher value of the index 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 T occ and the same randomly selected occupation as for beliefs about own skills fit. To create an index, we average the answers to the two apprenticeships and then standardize it. A higher value of the index 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 T occ and the same randomly selected occupation as for beliefs about own skills fit. To create an index, we average the answers to the two apprenticeships and then standardize it. A higher value of the index indicates stronger beliefs in getting along well with work colleagues.

d) *Beliefs about employer demand*: We ask, "Thinking about your skills, how much do you think employers are looking for someone like you to do an apprenticeship in the following occupations?" with answer options: not at all (1)/little (2)/moderately (3)/much (4)/very much (5). Each respondent answers this question to the T occ and the same randomly selected occupation as for beliefs about own skills fit. To create an index, we average the answers to the two apprenticeships and then standardize it. A higher value of the index indicates stronger beliefs in employer demand.

As exploratory analysis, we also construct the same measure about control occupations.

### *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 their first priority for the occupations to experience during the school event, we define students to have a preference for predominantly male, neutral, and predominantly female.
c) Preference for type of occupation: based on their first priority for the occupations to experience during the school event, we define students to have a preference for computers, working with people, or working with hands and/or machines/tools.
d) Math skills (low, medium, high – as defined above)

### *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 (defined as both parents originally not from Switzerland vs at least one parent originally from Switzerland) and 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:

a) Extraversion: We ask the respondents at baseline about their agreement to the statements "I see myself as someone who is reserved" and "I see myself as someone who is outgoing, sociable" on a scale from 1-5. We reverse the scale for the former statement and take the sum of the answers to the two statements. We define students as extravert if the sum is above the median in the sample and as introvert if it is at or below the median.
b) Social conformity: We ask the respondents at baseline about their agreement to the statements "I see myself as someone who does what is normal" and "I see myself as someone who cares about what others think about me" on a scale from 1-5. We take the sum of the answers to the two statements. We define students as conforming if the sum is above the median in the sample and as non-conforming if it is at or below the median. We expect the impact of the treatment to be stronger for non-conforming students.

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

After students' baseline survey, teachers fill in a survey in which they answer for each student which apprenticeship they think the student would fit best to and which apprenticeship the student would benefit the most from experimenting during the event as well as the teacher's attitudes towards the value of experimentation. We will explore heterogeneity based on these answers.

*Additional outcomes*

We will also explore treatment effects on their preferences for occupational type and their motivation factors behind their occupational choice.

For the potential mechanisms described above for the secondary analysis, we will also explore the effect on beliefs about own skills fit/work tasks/work environment/employer demand for these outcomes separately for the T occ and the other randomly selected occupation from the same category as the T occ.

We will also consider the following outcome: *Beliefs about others' fit*: we elicit this outcome through a vignette, where we randomize whether the vignette student is called Lara (female) or Lukas (male). The question is: "Lara is very similar to you in terms of school type and grades. Ever since 6$^{th}$ grade, she has been inclined to do an apprenticeship in OCC-C [top favorite they experienced at event]. Do you think Lara should search for information about

other occupations before her final choice?" The answer options are "No, I think she should only focus on search about OCC-C" and "Yes, I think she should search for information about other occupations". If they choose yes, we ask them to indicate which occupations they would advise her to search information about. If they choose no, we set the average characteristics of the advised occupations to the value for OCC-C. Based on these answers, we construct two advice outcomes similar to the primary outcomes *search breadth* and *type index*.

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

i. abs(female share in T occ – average female share searched)
ii. abs(math requirements in T occ – average math requirements searched)
iii. 1-(share of searched occupations of same type as T occ)

*Potential spillovers*

Exploratorily, we also do the analysis by the treatment status of their best friend in class (which we ask about in the baseline survey).

We will also collect data from some schools that also hold the event but without randomization (all students are "pure control"). So by comparing the outcomes of the control group students in our experiment to these pure control students, we will be able to shed light on possible treatment-control spillovers.

## 6. Power Calculations

We target the recruitment of 2,300 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 0.12 standard deviations.