Pre-Analysis Plan: The impact of the Covid-19 pandemic on perceived returns to university education in Switzerland

Laura Hagspiel*

May 10, 2021

Contents

*University of Zurich, Chair of Prof. D. Dorn & supervision by Dr. J. Thibaud

1 Introduction

This paper studies educational plans and perception of pecuniary and non-pecuniary returns to higher education of Swiss pupils in times of the Covid-19 pandemic. I use subjective quantitative measures and a priming treatment to analyze the impact of the Covid-19 pandemic on educational intentions and future earnings expectations in Switzerland.

The rationale behind conducting this study is, that while the literature has recognized the importance of non-pecuniary and pecuniary outcomes for human capital investments, little is known about the role of a real and uncertain labour market shock like the Covid-19 pandemic on educational intentions and perceptions. This is of importance for the theory as well as for policymakers.

At the time of writing, only one paper [1] has looked at how students' beliefs about their current and future (educational) outcomes are impacted by the Covid-19 pandemic. However, there are great differences between this study and the paper mentioned above. There are distinctions in: (i) the underlying sample, (ii) the focus of the study itself, and (iii) the empirical approaches used.

1.1 Research Questions

The main question of this paper is as follows:

- Does the Covid-19 pandemic influence the educational plans of Swiss pupils, moreover the perception of returns to university education?

This research question can be split into two confirmatory sub-hypotheses.

- H1: The Covid-19 pandemic increases the intention of young Swiss to attend a higher educational institute.
- H2: The Covid-19 pandemic increases the perceived pecuniary and non-pecuniary returns to university education.

In addition to knowing the existence of a general change in expectations and intentions, I seek to disentangle potential mechanisms that may drive the observed responses. I have set up the following exploratory hypotheses:

- H3: Students have decreased their hours studying outside of school leading to a perceived negative academic performance during the pandemic.
- H4: Students who are strongly exposed to the pandemic have a higher intention to attend a higher education institution.

As can be seen from the literature, the consumption value is a strong predictor for schooling. I want to see if this still holds in times of a pandemic. I specify the exploratory hypotheses as following:

- H5: Non-pecuniary values have a significant positive effect on educational choices, regardless if exposed to a pandemic.
- H6: The social aspect of a higher educational degree is a significant factor for education intention & the perceived returns to education.

2 Study Design

2.1 Data

To capture the perception of returns to university education and to measure the effect of the Covid-19 pandemic, I conduct an online-survey with pupils in Switzerland from March-April 2021. Over 100 public and private secondary Swiss schools who offer a *Matura* have been asked for permission to distribute the survey. Around 40 schools have responded positively.

Primary sampling units are individuals who are between 16-18 years old who will either, soon face the decision to start working, or to undertake a tertiary degree. Those students visit one of the three different types of secondary school: high school, specialized secondary school, or vocational secondary school.

The sample forms a nationwide data set which includes secondary school pupils' beliefs about future labor market realizations and educational plans. In addition, I collect data on the beliefs about the consequences of the Covid-19 pandemic and on individual exposure to the pandemic.

2.1.1 Missing data from survey questions

Some respondents may not completely finish the survey due to potential misunderstanding and/or time issues. To minimize the first matter I pretested different versions of the survey with various students between the age of 15 and 20. Even the youngest seemed to have understood the questions and so no further considerations have been made.

It is likely that some participants will stop answering midway through the survey. If participants have just opened the survey or only answered the practice question, they are not included in the data set for the analysis.

As participants can only move forward in the survey, and due to the forced-response setup, data that may be missing is on the personal level. Meaning that whole blocks of answers will be missing and not just only a few singular values. In the face of this kind of data attrition, I will first check the dropout rate. If the total number of respondents who did not completely finish the questionnaire is $\leq 25\%$ of the total number of partial + finished surveys, I will consider dropping them. However, I will still have a look at the summary statistics of the uncompleted surveys for any systematic patterns.

In case of a high survey drop out rate ($> 25\%$), I will follow the subsequent steps:

1. First, I will create the dummy variable $Unfinished = 1$ if the individual has not fully finished the survey.

2. I will determine if the dropout is more or less random. Due to only being able to move forwards and the forced-answer setup in the survey, the treatment status is a central factor for potential different stopping patterns. By running a Chi-squared test, I can analyse a potential relationship between the treatment-assignment and the dropping out.

    - If I can not identify a significant pattern in the data, I will use the default option which is deletion.
    - If I identify a systematic pattern in the data, I will impute the missing variables by using the maximum likelihood approach.

2.2 The intervention: Priming

The intervention is a priming treatment in which some participants are more exposed towards the concept of a pandemic. The survey tool I am using - *Qualtrics* - will randomly put those questioned into a control or treatment group. In concrete terms, this means that the two groups face different versions of the questionnaire. One group – the treatment group – has the questions concerning the pandemic first with additionally an image of the nationally distributed "Corona-poster" (see Appendix A) of the Swiss Government shown to them. Also, they have one more question than the control group for the manipulation check, discussed below. The control group instead starts with the questions about the returns to education and educational plans - so that no previous connection to Covid-19 related issues is created.

To test the functioning of the priming, I apply two methods. Firstly, the common approach, is to do a manipulation check. A sentence construction task has been added

in which participants of the treated group will finish the word which starts with the letter "Pa". If the intervention works, then the majority should have answered with the word "Pandemic". Additionally to this specific awareness check, the responses to the questions regarding the different concerns about the Covid-19 pandemic can also be seen as an indirect control. In case of a successful activation of the concept of the Covid-19 pandemic, the level of concern should differ between the treated and the untreated group.

2.3 Subjective Treatment Effect

Another way to measure the effect of the Covid-19 pandemic, is to directly ask participants about the outcome now (during the pandemic) and if there were no pandemic; this is the so called subjective treatment effect. Mathematically, I can write it as:

$$STE_i = Y_i(Covid19 = 1) - Y_i(Covid19 = 0) \tag{1}$$

This approach was the main empirical strategy of the paper mentioned above, [1]. Which also looks at the relationship between Covid-19 and educational outcomes.

2.4 Index for Covid-19 Exposure

Because the pandemic affects everyone, but to differing amounts, I measure the impact of the Covid-19 pandemic by relating educational intentions and beliefs to the level of exposure to the pandemic. As I have a series of questions that measure how exposed and experienced people are in regard to the Covid-19 pandemic, I will additionally to looking at the individual facets of exposure reduce the dimension to a single component by creating an "Exposure Index". The Exposure Index will be introduced in section 3.1.

3 Empirical Analysis

This section is committed to present the empirical specifications used for the statistical analysis.

3.1 Measures

3.1.1 Main (Outcome) Variables

An outcome variable collected is the **expectation about monthly future pre-tax earnings** at age 25 in CHF (Discrete, 0 - 10000). I collect expected earnings for three different educational scenarios, conditional on having a job at that age:

1. Bachelor's degree from a university,
2. Bachelor's degree from a university of applied science,
3. Going into the workforce after school.

The secondary outcome variable is the **student's intention** to follow a tertiary educational path. I elicit for each one of the three educational paths - university, university of applied science, job market entry after school - the probability to choose one of the path, conditional on finishing school (Discrete, 0-100).

The non-pecuniary benefits & costs of higher education are captured by the **consumption value**. The consumption value consists of 5 different dimensions in my survey: (i) Enjoyment (ii) Social life (iii) Learning new things (iv) Financial troubles (v) Stress and/or Anxiety. Each value will be captured for the scenario where one goes to a higher educational institute and when one starts to work, (each dimension is: Discrete, 0-100).

3.1.2 Other Variables

To measure potential uncertainty in employment for different educational attainments, I have measures of the following variable for the three different educational scenarios:

- **Employment probability** (Discrete, 0-100).

I measure the direct exposure to the Covid-19 pandemic through different questions, giving the following variables:

- Active search for information about the Covid-19 pandemic - **Awareness Indicator** (Categorical):
- **Personal exposure** (Dummy):
  - Yes for knowing someone personally that has caught the virus

- **Economic exposure knowledge** (Dummy):
  - Yes for knowing someone who has lost their job/partial work
- **Perceived probability of catching the virus** (Discrete, 0-100)
  - 0 for very unlikely & 100 for very likely plus option *I have/had the coronavirus*

I ask direct questions in regard to changes due to Corona:

- **Negative school performance** (Dummy)
  - Yes for worse school performance
- **Change in tertiary education intention** (Categorical)
  - No
  - Yes, less likely
  - Yes, more likely

I have gathered the reason(s) for the possible change in the intention to visit a tertiary educational institution. The following options are given: (i) Family matters, (ii) Attractiveness of the study, (iii) Financial situation (iv) Social aspect, (v) Corona, (vi) Other reasons.

For the subjective treatment effect, I elicit changes in hours of school studying:

- **Current studying hours** outside of school time (Categorical, less than 5h to more than 17h bins):
- **Current studying hours, were there no Covid-19 pandemic** (Categorical, less than 5h to more than 17h bins)

I also collect variables for the potential impact the pandemic can have/had on various different aspects of the student's life:

- **Overall concern** (Discrete, 0-100)
- **Social life concern** (Discrete, 0-100)
- **Mental state concern** (Discrete, 0-100)

- **Health concern** (Discrete, 0-100)
- **Future academic performance concern** (Discrete, 0-100)
- **Future economic situation concern** (Discrete, 0-100)

To analyse a potential influence of school performance on our main outcome variables, I have a measure of the following variables:

- **Probability of finishing current education** (Discrete, 0-100)
- **Probability of expected performance in tertiary level** (Discrete, 0-100)
  - 0 for just managed to finish the degree and 100 for finishing with excellence

The basic control variables I am collecting are the following:

- Gender (Categorical)
- Age (Discrete, 15 or less - 18+)
- Graduation date (Categorical, graduating this year 2021 - graduating 2023 or later)
- Canton (Categorical)
- Public school (Categorical)
- Swiss nationality (Dummy)
- Relative school performance (Discrete, 0-100)
- Mother's education (Categorical)
- Father's education (Categorical)
- Academic environment (Dummy)
- Type of school (Categorical)
- Focus subject for high school pupils (Categorical)

Finally, I have a dummy variable for the intervention where individuals who are assigned to the priming group =1, or assigned to the control group, =0.

3.1.3 Defining returns

To calculate the corresponding proportional gain in earnings due to different educations, I define expected monetary returns as following:

$$ER_i^U = log(ExpEarn_i^U) - log(ExpEarn_i^W)$$
$$ER_i^{US} = log(ExpEarn_i^{US}) - log(ExpEarn_i^W)$$

where $ExpEarn_i^U$ is expected earnings of individual i for the scenario university U. $ExpEarn_i^W$ is the perceived earnings for the scenario entering the labour market W and $ExpEarn_i^{US}$ is for the scenario university of applied science US.

I will define the non-pecuniary returns as the following:

$$CV_i^D = V(HE)_i^D - V(W)_i^D$$

whereby CV_i^D is the non-monetary return from higher education. $V(HE)_i^D$ is the value of the consumption-value dimension D in case of the scenario higher education HE. $V(W)_i^D$ is the expected value of the same individual i for the dimension D in case of the scenario working W. The consumption value dimensions are the following: (i) Enjoyment (ii) Social life (iii) Learning new things (iv) Financial troubles (v) Stress and/or Anxiety.

3.1.4 The Exposure Index

The Exposure Index will be constructed by applying the variable addition approach with equal weighting. Before creating the index, I will scale the variables first, to render them comparable. Hereby, I will use the traditional z-normalization to avoid an arbitrary choice of scaling. This will process the data such that the mean is equal to 0 and the standard deviation is 1.

$$z_i = \frac{x_i - mean(x_i)}{sd(x_i)}$$

I apply an equal weighting to the variables that make up the Exposure Index. The Exposure Index is defined by:

$$EX_i = \frac{1}{4}\left(AW_i + PE_i + EE_i + Corona_i\right),$$

where AW_i indicates how often individual i actively looks at Covid-19 related media. PE_i equals 1 if the same individual knows someone personally who has caught the coronavirus. $EE_i = 1$ if i knows someone personally who has lost their job or in short-term work due to the pandemic, and $Corona_i$ measures the expectation that the individual i will have had Covid-19 before the summer. I will treat persons who have indicated that they have had the coronaviurs before summer as 100.

3.2 Balancing Checks

To check if the randomization worked between treatment and control groups, I will perform descriptive statistics at the baseline for the two groups. In particular, the statistics will so far include the following variables: gender, age, public school, graduation date, Swiss nationality, relative school performance, parental education, academic environment, type of school. I will run a two sample t-test to check for any significant differences between the two groups.

3.3 Estimation procedures for the pandemic/treatment effect

I have three different measures to determine the effect of the Covid-19 pandemic on behaviour, intentions and beliefs. The different approaches are: (i) Subjective Treatment Effect & Direct Questions (ii) Priming Treatment (iii) Exposure Measures/Index.

3.3.1 Subjective Treatment Effect and Direct Questions

For the Subjective Treatment Effect, I elicit changes in studying hours due to the pandemic by applying equation 1. Then I will use a Wilcoxon Test with the Alternative-Hypothesis that on average the pandemic has reduced the hours studied outside of school.

Moreover, I can go into more detail and see if there are any heterogeneous effect by running following regressions, once with cantonal fixed effects and once without them:

$$\Delta_i = \beta_0 + \beta_1 Gender_i + \epsilon_i$$
$$\Delta_i = \alpha_0 + \alpha_1 Swiss + \epsilon_i$$
$$\Delta_i = \delta_0 + \delta_1 HighSocio_i + \epsilon_i$$

where Δ_i is the difference in hours studied now vs. had there not been a Covid-19 pandemic (Subjective Treatment Effect). $Gender_i$ determines the gender of i. $Swiss_i$ responds to whether i has Swiss nationality. $HighSocio_i$ is equal to one if at least one of the parents has a higher educational degree. Standard errors will be clustered at the canton level.
Alternatively, I can run Wilcoxon Tests (Kruskal Wallis test respectively) for the specific difference.

Additionally to these personal characteristics, I want to check if the school form and the school type matters. Therefore, I will additionally run the following regressions, again once with cantonal fixed effects and again once without them:

$$\Delta_i = \gamma_0 + \gamma_1 School_i + \epsilon_i$$
$$\Delta_i = \nu_0 + \nu_1 Public_i + \epsilon_i$$

where $School_i$ is a vector of the school type individual i is visiting. $Public_i$ is equal to one if individual i visits a public school. Standard errors are clustered at canton level.

In the end I will also run a regression with additional basic control variables:

$$\Delta_i = \beta_0 + \beta_1 Gender_i + \beta_2 School_i + \beta_3 Swiss_i +$$
$$\beta_4 HighSocio_i + \beta_5 Public_i + \beta_6 Rel.Performance_i$$
$$+\beta_7 Graduation_i + \beta_8 Subject_i + \gamma FE_c + \epsilon_i$$

where $Rel.Performance_i$ is an indicator for how well individual i's performance in school relative to the colleagues is. $Graduation_i$ states when i will finish school. $Subject_i$ indicates the focus subject. FE_c are cantonal fixed effects. The standard errors are clustered.
I will also in general run regressions with the variable Age_i instead of $Graduation_i$ separately to avoid potential collinearity.

Furthermore, I have directly asked for changes in schooling performance and in tertiary education intentions due to the pandemic. As I have collected reasons for why one has changed their schooling intentions, I will not run further regressions on it. However, for the dummy "Negative impact on schooling performance", I will perform an exact Binomial Test to see if the data follows a Bernoulli(p=0.5) distribution.
Then I will run following Probit-regressions to see any heterogeneity in the responses

- one time with cantonal fixed effects and otherwise just as below:

$$NegPerform_i = \beta_0 + \beta_1 Gender_i + \epsilon_i$$
$$NegPerform_i = \gamma_0 + \gamma_1 School_i + \epsilon_i$$
$$NegPerform_i = \alpha_0 + \alpha_1 Swiss + \epsilon_i$$
$$NegPerform_i = \delta_0 + \delta_1 HighSocio_i + \epsilon_i$$
$$NegPerform_i = \nu_0 + \nu_1 Public_i + \epsilon_i$$

where $NegPerform_i$ is equal to one if the pandemic had a negative impact on individual i's academic performance. $Gender_i$ indicates the gender. $School_i$ is the type of school individual i visits. $Swiss_i$ captures if i is Swiss. $HighSocio_i$ is equal to one if one of the parents of individual i has a tertiary degree. $Public_i$ indicates if i visits a public school. Standard errors will be clustered at the canton level.

Similar to above, I will also run a Probit-regression with additional controls:

$$NegPerfom_i = \beta_0 + \beta_1 Gender_i + \beta_2 School_i + \beta_3 Swiss_i +$$
$$\beta_4 HighSocio_i + \beta_5 Public_i + \beta_6 Rel.Performance_i$$
$$+\beta_7 Graduation_i + \beta_8 Subject_i + \gamma FE_c + \epsilon_i$$

I aim to graphically depict the results of the two heterogeneity analysis.

3.3.2 Basic statistical model for the priming intervention

I will estimate the effect of the intervention on the key outcome variables with a treatment analysis. This means that I will regress expected earnings and educational intentions on the treatment dummy and control variables. For the perception of monthly earnings for the three different scenarios, I will run the following specification:

$$log(ExpEarn_i^S) = \beta_0 + \beta_1 Treatment_i + \mu X_i + \gamma FE_c + \epsilon_{Si}$$

where $log(ExpEarn_i^S)$ is the logarithm of the expected earnings of individual i for the scenario S where S can be one of the 3 scenarios: university, university of applied science, job market entry. The logarithm can only be applied in case of when no participants expect zero income; otherwise I will consider adding a constant as part of the pre-processing stage. X_i is a vector including the baseline variables. I propose standard demographic and proxy variables: age, gender, graduation date, public school indicator, Swiss nationality, relative school performance indicator,

mother's & father's education, academic environment measurement, focus subject and indicator for type of school. However, I will need to analyse how the graduation date and age interact with each other to potentially avoid collinearity. FE_c are the cantonal fixed effects. Standard errors will be clustered at canton level.

I will run the same regression again but this time I will use the pecuniary returns as outcome variables -once with control variables and fixed effects and once without:

$$ER_i^S = \beta_0 + \beta_1 Treatment_i + \mu X_i + \gamma FE_c + \epsilon_{Si}$$

where ER_i^S are the expected monetary returns of individual i for the education S where S can be university or university of applied science. Again with clustered standard errors.

Furthermore, I will add the consumption value variable to increase the precision of the regression above such that in the end I will run the following specification:

$$ER_i^S = \beta_0 + \beta_1 Treatment_i + \beta_2 CV_i + \mu X_i + \gamma FE_c + \epsilon_{Si}$$

CV_i is a vector of the non-pecuniary returns to higher education. The consumption value consists of 5 different dimensions: (i) Enjoyment (ii) Social life (iii) Learn new things (iv) Financial troubles (v) Stress and/or Anxiety. Clustered standard errors.

To check for a potential impact of the intervention on educational intentions, I will run the following specification:

$$ProbPath_i^P = \beta_0 + \beta_1 Treatment_i + \mu X_i + \gamma FE_c + \epsilon_{Pi}$$

where $ProbPath_i^P$ is the probability of individual i to choose the educational path P. P is one of the three different paths: (i) University (ii) University of applied science (iii) Direct job market entry. Clustered standard errors.

To accurately measure the impact I can add various covariates, one by one to the regression above. So that in the end I will run the following regression:

$$\begin{aligned} ProbPath_i^P = \beta_0 + \beta_1 Treatment_i + \beta_2 ProbEmploy_i^P + \beta_3 CV_i \\ + \beta_4 ER_i^U + \beta_5 ER_i^{US} + \mu X_i + \gamma FE_c + \epsilon_{Pi} \end{aligned}$$

where $ProbEmploy_i^P$ is the perceived probability to have a paid job at age 25 conditional on the given educational scenario P. P is one of the following set ups: (i) Bachelor's degree from university (ii) Bachelor's degree from university of applied science (iii) Job market entry after school. ER_i^U is the expected return from a university of individual i. ER_i^{US} is the expected return from a university of applied science. Clustered standard errors.

Supplementary, I will in general run the longer regressions for expected returns and probability of path with $ExpPerform_i$ as a further independent variable. However, I will need to look at how $ExpPerform_i$ and relative academic performance in the control vector interact with each other to not run into the problem of collinearity. A potential solution to this - if needed - might be to take the difference between $ExpPerform_i$ and $RelPerfomance_i$.

To elicit possible changes in the consumption values due to the intervention, I will run the following regression:

$$CV_i^D = \beta_0 + \beta_1 Treatment_i + \epsilon_{Di}$$

where CV_i^D is individual's i consumption value of the dimension D. $Treatment_i$ is the treatment dummy. Standard errors are clustered at the canton level.

I will also run the regression above with the baseline control variables being here gender, graduation date, parental education, public school, school type, Swiss nationality, relative school performance, academic environment and cantonal fixed effects:

$$CV_i^D = \beta_0 + \beta_1 Treatment_i + \mu X_i + \gamma FE_c + \epsilon_{Di}$$

Standard errors are clustered at the canton level.

3.3.3 Fixed effect model for the Covid-19 Exposure Index

Another way to look at the impact of the pandemic on the outcome variables is to consider how exposed someone is to the pandemic as a whole. Therefore, compared to the regressions above, I will use the Exposure Index as the explanatory variable. For the expected earnings I will first run the simple relationship:

$$ER_i^S = \beta_0 + \beta_1 EX_i + \epsilon_{Si}$$

with standard errors clustered at the canton level.

Then I will control for the omitted variable bias by adding several control variables and cantonal fixed effects. I will start off with the basic controls, and add on other potential influential variables:

$$ER_i^S = \alpha_0 + \alpha_1 EX_i + \mu X_i + \gamma FE_c + \epsilon_{Si}$$
$$ER_i^S = \delta_0 + \delta_1 EX_i + \delta_2 CV_i + \mu X_i + \gamma FE_c + \epsilon_{Si}$$

where ER_i^S is the expected monetary returns of individual i for the scenario S. EX_i is the Exposure Index. FE_c are the cantonal fixed effects. CV_i is the vector containing the different dimensions of the consumption value. X_i is a vector containing the baseline variables as defined by 3.3.2. With clustered standard errors.

Now, I will also run the different specifications for the educational intentions as the dependent variable:

$$ProbPath_i^P = \beta_0 + \beta_1 EX_i + \epsilon_{Pi}$$
$$ProbPath_i^P = \alpha_0 + \alpha_1 EX_i + \mu X_i + \gamma FE_c + \epsilon_{Pi}$$
$$ProbPath_i^P = \delta_0 + \delta_1 EX_i + \delta_2 ER_i^U + \delta_3 ER_i^{US} + \delta_4 ProbEmploy_i^P + \delta_5 CV_i + \mu X_i + \gamma FE_c + \epsilon_{Pi}$$

where $ProbPath_i^P$ is the probability to choose one out of the three educational paths P. $ProbEmploy_i^P$ is the perceived probability to have a paid job at age 25 in the scenario P. Clustered standard errors.

I will also see if the level of exposure to the pandemic has an influence on the consumption values:

$$CV_i^D = \beta_0 + \beta_1 EX_i + \epsilon_{Di}$$
$$CV_i^D = \alpha_0 + \alpha_1 EX_i + \mu X_i + \gamma FE_c + \epsilon_{Di}$$

where CV_i^D is the non-pecuniary return to higher education of individual i for the dimension D. Standard errors are clustered at canton level.

Above, I have only tested for a potential relationship between the Exposure Index and the key outcome variables. However, I am also interested which aspect of pandemic exposure can lead to potential changes in beliefs and intentions. Therefore, in the mechanism section I will look at how the individual exposure aspects play in explaining the pandemic effect.

3.4 Mechanisms

Above I looked at various ways to determine if the pandemic has an effect on the main outcome variables. This section however, aims at eliciting how the pandemic influences educational intentions and beliefs. To see which aspects of the Covid-19 pandemic will drive the results, I will look at potential individual mechanisms one by one.

3.4.1 Drivers of the Exposure Index

I will start off with looking at the correlation-matrix including all variables to generally identify potential drivers of the pandemic. I have collected so far, the following mechanisms relating how the Covid-19 pandemic could potentially drive the effect of the Exposure Index on the main outcome variables:

- More exposed individuals have seen a negative change in hours studied,
- More exposed individuals are more worried overall about the pandemic,
- More exposed individuals are more worried about their health and mental state,
- More exposed individuals worry a lot about their future economic situation and future academic performance,
- More exposed individuals worry more about their social life,
- More exposed individuals are less likely to finish current education,
- More exposed individuals experienced a negative impact on their school performance due to Covid-19.

To understand what lies behind potential changes in the key outcome variables, I will apply the following general steps, based on the assumption that there is a significant relationship between the Exposure Index and the main outcome variables [1]:

1. I check if the potential driver is related to the Exposure Index:

$$Mechanism_i = \beta_0 + \beta_1 EX_i + \mu X_i + \gamma FE_c + \epsilon_i$$

where $Mechanism_i$ is one of the following variables: differences in studying hours - Δ_i, level of overall Covid19 worry - OW_i, level of health concern - HW_i,

[1] In case of non significance, I will still test - see below 3.4.3 - if individual exposure measures have predictive power and analyse their possible mechanisms then.

mental state worry - MW_i, the level of worries concerning the social life - SW_i, the level of future economic concerns - EW_i, probability to finish current school - $ProbSchool_i$, negative school performance due to Covid - $NegPerfom_i$, and APW_i is the level of concerns towards future academic performance. With clustered standard errors at the canton level.

2. Then I will regress the potential mechanisms separately on the dependent variables to verify potential explanatory power:

$$y_i = \beta_0 + \beta_1 Mechanism_i + \mu X_i + \gamma FE_c + \epsilon_i$$

Where y_i takes the following variables: ER_i^S is the expected returns to the school scenario S, where S is either university or university of applied science. $ProbPath_i^P$ is the educational intentions to follow one of the following paths P: university, university of applied science, direct job market entry. X_i is a vector containing the baseline controls as defined in the previous section. Standard errors clustered at canton level.

I will in general run the two regressions which check for a potential mechanism once with the fixed effects and control variables and once without them.

3.4.2 Drivers of the priming intervention

The following mechanisms, related to the Covid-19 pandemic, could potentially drive the effect of the intervention:

- High overall future worries due to the pandemic,
- High social worries,
- High future economic and academic performance worries,
- High mental and health state worries,
- High perceived probability to catch the coronavirus,
- Negative perception about the current schooling performance due to the pandemic,
- Large reduction in current studying hours.

I will apply the same general steps as above to determine a potential explanation of why the increasing salience might have affected perceptions and intentions, assuming that there is a significant relationship between the y_i's and the $Treatment_i$ dummy:

1. I will run the following regression:

$$Mechanism_i = \beta_0 + \beta_1 Treatment_i + \mu X_i + \gamma FE_c + \epsilon_i$$

   where $Mechanism_i$ is one of the following variables: Future overall worries due to the pandemic - OW_i, level of current social worries - SW_i, future economical worries due to the pandemic - EW_i, future academic performance worries - APW_i, how negatively the pandemic has influenced health worries - HW_i, the level mental worry - MW_i, measurement for the level of active search for Covid-19 pandemic information - AW_i, the perceived probability of catching the virus - $ProbofCatch_i$, a perceived negative impact on schooling performance due to Covid-19 - MW_i, and Δ_i - the change in studying hours. Clustered standard errors.

2. I will check for an actual significant relationship between the mechanism and the outcome variables:

$$y_i = \beta_0 + \beta_1 Mechanism_i + \gamma FE_c + \mu X_i + \epsilon_i$$

   where y_i takes the following variables: ER_i^S is the expected returns to the school scenario S where S is university returns or university of applied science returns. $ProbPath_i^P$ is the educational intentions to follow one of the following paths P: university, university of applied science, direct job market entry. Standard errors are clustered at canton level.

3.4.3 The individual facets of the Exposure Index

As the mechanism approach used above can not be applied or is inappropriate for most of the individual exposure measures, I will run separately versions with and without controls & cantonal fixed effects. I will run specifications of the following

regressions to determine the role of the individual exposure factors:

$$ProbPath_i^P = \beta_0 + \beta_1 AW_i + \beta_2 PE_i + \beta_3 EE_i + \beta_4 ProbofCatch_i + \beta_5 ER_i^U$$
$$+\beta_6 ER_i^{US} + \beta_7 ProbEmploy_i^P + \beta_8 CV_i + \mu X_i + \gamma FE_c + \epsilon_{Pi}$$
$$ER_i^S = \delta_0 + \delta_1 AW_i + \delta_2 PE_i + \delta_3 EE_i + \delta_4 ProbofCatch_i + \delta_5 CV_i +$$
$$\mu X_i + \gamma FE_c + \epsilon_{Si}$$
$$CV_i^D = \alpha_0 + \alpha_1 AW_i + \alpha_2 PE_i + \alpha_3 EE_i + \alpha_4 ProbofCatch_i +$$
$$\mu X_i + \gamma FE_c + \epsilon_{Di}$$

where AW_i denotes the amount of active search for Covid-19 related media. PE_i is equal to one if the individual i know someone personally who has caught the virus. $EE_i = 1$ if the same i knows someone who has lost their job/is in short-time work due to the pandemic. $ProbofCatch_i$ is individual i's belief that they have caught the virus before this summer. Standard errors are clustered at the canton level.

3.4.4 Mechanism of the Subjective Treatment Effect/Direct Questions

I will test if a possible change in study hours in times of the pandemic, more students see a reduction in their schooling performance. First I will run the simple Probit-regression without any controls:

$$NegPerform_i = \beta_0 + \beta_1 \Delta_i + \epsilon_i$$

Standard errors will be clustered at canton level.

Finally, I will add control variables:

$$NegPerform_i = \beta_0 + \beta_1 \Delta_i + \beta_2 Gender_i + \beta_3 School_i + \beta_4 Swiss_i +$$
$$\beta_5 HighSocio_i + \beta_6 Public_i + \beta_7 Rel.Performance_i$$
$$+\beta_8 Graduation_i + \gamma FE_c + \epsilon_i$$

Clustered standard errors.

3.4.5 Heterogeneity in priming treatment effect

To check for possible heterogeneity in the priming treatment effect, I can use interaction terms. For example if one is overall more worried regarding the pandemic, the treatment might have worked more likely and one might be more aware of potential

changes in returns:

$$ER_i^S = \beta_0 + \beta_1 Treatment_i + \beta_2 OW_i + \beta_3 Treatment_i \times OW_i + \gamma FE_c + \epsilon_{Si}$$

Standard errors are clustered at the canton level.

Hence, I will interact the treatment dummy with different concerns and exposure measurements. Concretely that means I will run the further following specifications:

$$ER_i^S = \alpha_0 + \alpha_1 Treatment_i + \alpha_2 EW_i + \alpha_3 Treatment_i \times EW_i + \gamma FE_c + \epsilon_{Si},$$
$$ER_i^S = \iota_0 + \iota_1 Treatment_i + \iota_2 PE_i + \iota_3 Treatment_i \times PE_i + \gamma FE_c + \epsilon_{Si},$$
$$ER_i^S = \gamma_0 + \gamma_1 Treatment_i + \gamma_2 APW_i + \gamma_3 Treatment_i \times APW_i + \gamma FE_c + \epsilon_{Si},$$
$$ER_i^S = \delta_0 + \delta_1 Treatment_i + \delta_2 AW_i + \delta_3 Treatment_i \times AW_i + \gamma FE_c + \epsilon_{Si},$$
$$ER_i^S = \rho_0 + \rho_1 Treatment_i + \rho_2 EE_i + \rho_3 Treatment_i \times EE_i + \gamma FE_c + \epsilon_{Si},$$
$$ER_i^S = \tau_0 + \tau_1 Treatment_i + \tau_2 NegPerform_i +$$
$$\iota_3 Treatment_i \times NegPerform_i + \gamma FE_c + \epsilon_{Si},$$

where EW_i catches the effect of the pandemic on the future personal economic situation. PE_i equals 1 if the individual i knows someone who has caught the virus. $NegPerform_i$ equals 1 if the individual i perceives a negative impact on schooling performance due to the pandemic. APW_i indicates how strongly the pandemic has influenced future academic performance concerns. AW_i measures the active search for Covid-19 pandemic information. EE_i equals 1 if individual i knows someone who has lost their job/short-time work. Clustered standard errors.

The same method will be applied, but now with the educational intention as the outcome variable. For educational intentions, I will run the following regression

specifications:

$$ProbPath_i^P = \beta_0 + \beta_1 Treatment_i + \beta_2 OW_i + \beta_3 Treatment_i \times OW_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \kappa_0 + \kappa_1 Treatment_i + \kappa_2 PE_i + \kappa_3 Treatment_i \times PE_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \alpha_0 + \alpha_1 Treatment_i + \alpha_2 EW_i + \alpha_3 Treatment_i \times EW_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \gamma_0 + \gamma_1 Treatment_i + \gamma_2 APW_i + \gamma_3 Treatment_i \times APW_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \delta_0 + \delta_1 Treatment_i + \delta_2 AW_i + \delta_3 Treatment_i \times AW_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \rho_0 + \rho_1 Treatment_i + \rho_2 EE_i + \rho_3 Treatment_i \times EE_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \Theta_0 + \Theta_1 Treatment_i + \Theta_2 SW_i + \Theta_2 Treatment_i \times SW_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \lambda_0 + \lambda_1 Treatment_i + \lambda_2 MW_i + \lambda_2 Treatment_i \times MW_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \iota_0 + \iota_1 Treatment_i + \iota_2 HW_i + \iota_2 Treatment_i \times HW_i + \gamma FE_c + \epsilon_{Pi},$$
$$ProbPath_i^P = \nu_0 + \nu_1 Treatment_i + \nu_2 NegPerform_i +$$
$$\nu_3 Treatment_i \times NegPerform_i + \gamma FE_c + \epsilon_{Pi},$$

where SW_i indicates how strongly the individual i is concerned about their social life due to the pandemic. MW_i measures i's concerns about their mental state due to Covid-19. HW_i stands for health concerns of i. Clustered standard errors at the canton level.

References

[1] Esteban M. Aucejo et al. "The impact of COVID-19 on student experiences and expectations: Evidence from a survey". en. In: *Journal of Public Economics* 191 (Nov. 2020), p. 104271. ISSN: 0047-2727. DOI: 10.1016/j.jpubeco.2020.104271. URL: https://www.sciencedirect.com/science/article/pii/S0047272720301353 (visited on 04/07/2021).

A Questionnaire

Priming Poster

[If in the Treatment Group] Please have a close look at the picture shown below

Block 1: Corona Exposure + Subjective Treatment

Q1 Please answer the following questions: How often do you read a text, listen to a podcast, watch a video about the Corona-pandemic? [Several times a day, Once a day, Not daily but several times a week, Once a week, Less than once a week]

Q2 Do you personally know someone who has caught the Coronavirus? [Yes, No]

Q3 Do you personally know someone who lost their job or is on short time work this or last year due to Corona related issues? [Yes, No]

Q4 Has the Corona-pandemic had a negative impact on your school performance (grades)? [Yes, No]

Q5 Outside of school hours, how many hours a week on average do you spend studying? [Less than 5h, 6-8h., 9-11h., 12-14h., 15-17h., More than 17h.]

- In this school semester (Currently)
- In this school semester (Currently), if the Corona-pandemic would not have happened

Block 2: Corona Worries (Priming Check)

Q6 Overall, to what extent do you think will the Corona-pandemic have a negative impact on your future? [0: No negative consequences - 100: Life-changing negative consequences]

Q7 How likely do you think it is that you will get the Coronavirus before the summer? [0-100; I already have/had it]

Q8 How concerned are you that the Corona-pandemic nowadays negatively impacts your ... [0-100]

- Social life (Friendships, Partnership...)
- Mental state (Mood, Wellbeing...)
- Health

Q9 How concerned are you that the Corona-pandemic will in the upcoming years negatively impact your ... [0-100]

- Future academic performance
- Economic situation (earnings, employment,)

Q10 [If in the Treatment Group] Which word is the first word that comes to your mind that begins with the following letters: **Pa**?

Block 3: Student's beliefs about returns to education

In most of the following questions, you will be asked about how likely an event is to happen. The scale will range between 0 and 100 where 0 means there is no chance that the event will happen and 100 means that the event will certainly take place. Set the number somewhere between 0 and 100 how likely you think the event is.
Q11 Let's try a practice question first: How likely do you think that it will snow tomorrow? [0-100]

The next questions will ask you to think about **hypothetical** scenarios. Please read the questions carefully and try to place yourself in the hypothetical situation when answering the questions. Please try to answer all question, even if you are unsure about it (there is no right or wrong answer)

Scenario 1: Imagine your life in the future when you are 25 years old. How does your life look like when you have taken the university-path after school and graduated from **university with a Bachelor's degree**? Consider the different kinds of jobs that are available to you.

- Q12 How likely do you think it is that you have a paid full-time job at the age of 25? [0-100]
- Q13 Now imagine additionally, that you already do have a full-time job. How much do you think you are earning on average per month, pre-tax (at this age (25))? [CHF, 0-10000]

Scenario 2: Imagine your life in the future when you are 25 years old. How does your life look like when you have taken the university-of-applied-science-path after school and graduated from **university of applied science with a Bachelor's degree**? Consider the different kinds of jobs that are available to you.

- Q14 How likely do you think it is that you have a paid full-time job at the age of 25? [0-100]
- Q15 Now imagine additionally, that you already do have a full-time job. How much do you think you are earning on average per month, pre-tax (at this age (25))? [CHF, 0-10000]

Scenario 3: Imagine your life in the future when you are 25 years old. How does your life look like when you have decided NOT to continue education at a higher educational institute after school but instead to **go into the workforce**. Consider the different kinds of jobs that are available to you.

- Q16 How likely do you think it is that you have a paid full-time job at the age of 25? [0-100]
- Q17 Now imagine additionally, that you already do have a full-time job. How much do you think you are earning on average per month, pre-tax (at this age (25))? [CHF, 0-10000]

Q18 Scenario 4: Imagine that you go to a **higher educational institute** (e.g. university, university of applied science) With what probability do you agree to the following statements: [0-100]

- I will enjoy my studies
- I will meet many new people and enjoy social activities
- I will learn many new things
- I will have financial troubles
- I will have stress and/or anxiety

Q19 Scenario 5: Imagine that you do not go to a higher educational institute but instead you **start working**. With what probability do you agree to the following statements: [0-100]

- I will enjoy my studies
- I will meet many new people and enjoy social activities
- I will learn many new things
- I will have financial troubles
- I will have stress and/or anxiety

Block 4: Educational plans

Q20 How likely do you think it is that you will finish school? [0-100]

Q21 Assuming you finish school, how likely do you think it is you will choose one of the following education paths: [0-100]

- University

- University of applied science
- Direct job entry

Q22 If you were to go to a higher educational institute (e.g. university, university of applied science), how well do you expect to complete your degree? [0-100 Finish with excellence]

Q23 Has your decision to attend a higher educational institution (e.g. university, university of applied science) changed since March last year when the Covid-19 pandemic began? [Yes I am now more likely to visit a higher educational institution, Yes I am now less likely to visit a higher educational institution, No]

Q24 [If "Yes" selected] If yes, for what reason(s) (several options possible): [Family matters, Attractiveness of the study, Financial situation, Social aspect, Corona, Other reasons]

Block 5: Basic questions

Q25 What is your gender? [Female, Male, Other]

Q26 What is your age? [15 or less, 16, 17, 18, 19, 20 or older]

Q27 In which canton do you go to school?

Q28 Were you born in Switzerland? [Yes, No]

Q29 How many years do you have left until you will graduate from school? [This year, Next year, More than 2 years away]

Q30 Do you go to a public school? [Yes, No, I don't know]

Q31 How good do you think your performance in school is (compared to your colleagues)? [0-100]

Q32 What level of education has your mother achieved? [No formal degree, High school degree, Vocational degree, University degree, University of applied science degree, Doctorate or higher, Apprenticeship training, Don't know, Other]

Q33 What level of education has your father achieved? [No formal degree, High school degree, Vocational degree, University degree, University of applied science degree, Doctorate or higher, Apprenticeship training, Don't know, Other]

Q34 Do you have any close friends or relatives that are going or went to a higher educational institute (e.g. university, university of applied science) ? [Yes, No]

Q35 Which type of school do you attend? [High school, Vocational school (IT or Commercial), Specialized secondary school]

Q36 [If "High school"] In which field is your focus subject? [Ancient Language, Modern Language, Physics and the applications of mathematics, Biology and Chemistry, Business and Law, Philosophy/Pedagogy/Psychology, Art, Music, Other]