Savings and Job-Search - An Impact Evaluation of a Financial Education Intervention in Vietnam: Pre-Analysis Plan

Lukas Hensel, Stefano Caria[*]

June 18, 2017

1   Introduction

We evaluate a randomized controlled trial to test the effects of a financial education intervention and an additional social media intervention to encourage savings in two cities in Vietnam. Participants are recruited amongst vocational school students and are randomly divided into three groups:

- **Control group**, whose members only receive a youth employment intervention (Skills to Succeed). The Skills to Succeed training includes employability skills training, which includes entrepreneurship, on-the-job training, career counselling, and job linkage services.
- **Treatment group 1**, whose members receive a financial education intervention on top of the youth employment intervention. The financial education intervention is a half-day training program emphasizing the importance of saving and teaching different saving strategies, including goal setting and spending strategies.
- **Treatment group 2**, whose members receive social media encouragement on top of the financial education intervention and the youth employment intervention. Participants can sign-up for a savings challenge wherein students can sign up to save a given amount for a given goal over 6 months, and win progressively bigger prizes for achieving their 1-

---

[*]University of Oxford

month, 3-month, and 6-month goals. They are also are encouraged to join a Facebook group through which students can:

- Win prizes upon completing quizzes and contests soliciting user-generated content that illustrate financial education themes
- Receive tips, in multi-media form, on selecting and using financial services and overcoming other common financial management issues
- Receive peer recognition upon reaching milestones in the savings challenge
- Be exposed to other mechanisms and messaging that reinforce the enhanced financial education content and foster sustained enthusiasm and motivation around saving.

The data collection took place in three stages. First, the baseline data was collected from March to August 2016 while participants were still part of the youth employment intervention. Second, a subset of candidates answered a financial knowledge test right before and after the financial education intervention. Finally, the endline data was collected using a phone survey conducted in May 2017.

2   Research Design and Question

The main question this trial attempts to answer is whether a financial education intervention can help credit constraint youth to improve their job-search behaviour through increased savings. The savings should enable youths to extend their job-search duration which should, on average, result in better placements. Similarly, savings could also help credit-constraint youths to start businesses using the accumulated capital. Furthermore, the trial is intended to analyse psychological effects associated with building a saving habit, such as self-control, planning, and future orientation.

This experiment is designed to estimate the marginal impact of i) the financial education intervention and ii) the additional social media encouragement on several outcome families. First, we will test for the direct impact of the interventions on knowledge about saving and on time-preferences on a monetary domain. In a next step we will test whether either of the interventions affected savings behaviour, job-search, and employment outcomes. Lastly, we will conduct a heterogeneity analysis to identify sub-groups that benefited more than

others. Beyond analysing the pure impact of the program on outcomes, we will also conduct a formal mediation analysis to understand channels through which the intervention affected outcomes.

2.1 Definition of Outcomes

We specify three families of outcome variables. The first family only consists of an incentivized time-preference measure. The second family contains outcome variables related to savings. The third family of outcomes is related to search behaviour and the the fourth family of outcomes is related to employment outcomes. All outcome variables are defined below

- Saving outcomes (family 1):
  - Currently saving, a dummy variable equal to one if the participant is currently saving money.
  - Saving frequency, a dummy variable equal to one if the participant saves more frequently than the control group median.
  - Saved amount, a continuous variable measuring the total amount of savings.
  - Saving purpose (1-2), two dummy variables indicating whether a participant saved for two different purposes (pay for job-search expenses, starting a business)[1]

- Search behaviour (family 2):
  - Number of applications per month of job-search (graduates with zero reported month of job-search will be coded as 0.25 months)
  - Number of months of job search before finding the first job after graduation (we set as missing observations on individuals who did not start a job)
  - Number of applications submitted before finding the first job after graduation (we set as missing observations on individuals who did not start a job))

- Employment outcomes (family 3):[2]
  - Currently wage-employed or self-employed (dummy variable)

---

[1]We will also report the results for whether participants saved to pay for further education or to wait for a better job. However, these results are only secondary hypothesis and will not be included in the multiple error correction.

[2]In addition to all variables mentioned below, we will also report the impact on whether employed participants are still in probation. However, this is only a secondary outcome and will not be included in the multiple hypothesis test correction.

- – Currently wage-employed (dummy variable)
- – Currently self-employed (dummy variable)
- – Monthly earnings (we set as zero observations on individuals who did not start a job)
- – Written contract (dummy variable)[3]
    - * Non-working coded as missing
    - * Non-working coded as zero

- – Using skills learned in vocational training conditional on being employed (self-assessed dummy variable)

We also consider two separate mediating outcomes.
- Above median time preference as measured by an incentivized measure on a monetary domain in the endline survey
- Knowledge on about saving as measured by six knowledge questions administered before and directly after the training.

A further set of outcomes are a series of self-reported psychological outcomes, each based on three four-point Likert-scale questions. We will test whether the treatment affects the following outcomes. For each categories we will report effects on dummies of candidates being above the baseline median. The categories are[4]

- Positive self-concept
- Self-control
- Higher order thinking skills

The top 1% of values of continuous and unbounded variables will be winsorized. Observations at the median will be part of whichever group of below and above median has fewer observations, where applicable.

---

[3]We will also explore impacts on the duration of contracts, but we will treat this only as a secondary outcome.

[4]In addition to testing three variables we will also report impacts on the secondary outcomes social skills, communication skills, and job-search skills. We will not include these variables for multiple hypothesis testing as they were not directly targeted by the intervention and impacts are speculative.

2.2   Control Variables

We estimate all regression models with and without a list of controls which we think might affect savings. The vector of control variables will always include the baseline value of the dependent variable where available in the data.[5] Furthermore, we include the following control variables:

- Dummies for each sector of vocational training (sea food processing, hospitality, garments, mechanics and electronics, and sales and marketing)
- Work experience, a dummy equal to one if an individuals has ever worked before baseline survey
- Above median age
- Gender, a dummy equal to one if the respondent is female
- Migrant, a dummy equal to one if a participant is a migrant from outside the city.
- School fixed effect

We control for the sector of training as candidates face substantially different job-markets across sectors. We also control for school fixed effects (the level of stratification). We further control for previous work experience as this might affect both the job-search prospects and the ability to save. We also control for gender and whether the participant is a migrant as these might affect job-search outcomes.

2.3   Definition of Dimensions of Heterogeneity

We will analyse heterogeneity of treatment effects across the following dimensions:

- Above median age
- Gender
- Whether a participant is a migrant from outside the city.
- Whether a participant saved at baseline
- Whether a participant saved formally at baseline (i.e. using a bank)
- Whether a participant had ever worked at baseline
- Whether a participant preferred wage work at baseline
- Whether a participant mentioned job-search as reason for saving at baseline

---

[5]The questionnaire changed slightly between the survey rounds so that we will control for outcomes as measured at baseline, even if the question changed between survey rounds.

- Whether a participant preferred running a business at baseline
- Whether a participant mentioned starting a business as reason for saving at baseline

Observations at the median will be included in the group with the lower number of observations. We will test for heterogeneous treatment effects by basic demographics (age, gender) and whether participants are migrants as all of the might structurally affect the effectiveness of the treatment. Heterogeneity by variables related to baseline savings seems plausible given the focus of savings of the study. Similarly, preference over workplaces and reasons for savings are likely to affect the impact of the interventions.

3 Model Specification and Estimation

We estimate average treatment effects on all outcomes (including the main families, mediators and secondary outcomes) using the regression model

$$Y = \alpha_0 + \text{Treatment}_1 \cdot \alpha_1 + \text{Treatment}_2 \cdot \alpha_2 + \mathbf{S} \cdot \delta + \mathbf{X} \cdot \mu + \epsilon \tag{1}$$

where Treatment_1 and Treatment_2 are treatment group indicators, $\mathbf{S}$ are school fixed effects (the level of stratification), and $\mathbf{X}$ is a vector of containing the control variables as defined in section 2.2.

To analyse heterogeneous treatment effects, we estimate the following regression model

$$\begin{aligned} Y =& \text{Treatment}_1 \cdot Z_0 \cdot \beta_1 + \text{Treatment}_2 \cdot Z_0 \cdot \beta_2 \\ &+ \text{Treatment}_1 \cdot Z_1 \cdot \beta_3 + \text{Treatment}_2 \cdot Z_1 \cdot \beta_4 \\ &+ Z_0 \cdot \delta_0 + Z_1 \cdot \delta_1 + \mathbf{S} \cdot \delta + \mathbf{X} \cdot \mu + \epsilon \end{aligned} \tag{2}$$

where Z_0 and Z_1 are dummies for the dimension of heterogeneity. Z_0 indicates if the heterogeneity dummy is equal to zero. Similarly, Z_1 indicates if the heterogeneity dummy is equal to one. We will estimate both equations with and without the vector of control variables $\mathbf{X}$.

To further investigate the effects of treatment two, we will separately instrument
- Membership of facebook group
- Start of savings challenge
- Successful completion of savings challenge

with Treatment$_2$ in equation 3:

$$Y = \gamma_0 + \text{Treatment}_1 \cdot \gamma_1 + \mathbf{F} \cdot \gamma_2 + \mathbf{S} \cdot \delta + \mathbf{X} \cdot \mu + \epsilon \tag{3}$$

where $\mathbf{F}$ is one of the intermediate outcomes. We will estimate this specification for all outcome families. We will use heteroscedasticity robust standard errors without clustering as the randomisation was on the individual level.

4 Mediation Analysis

To analyse through which channels the financial education intervention affects jobs-search or employment, we will conduct at formal mediation analysis following Acharya et al. (2016). We hypothesize that the financial education intervention could change participants' time preferences and/or saving behaviour through increased savings knowledge. Furthermore, we hypothesize that job-search could be affected by both changes in savings and changes in time preference. We will conduct this analysis if we find significant treatment effects for any outcomes in families 1-3 in specification 1 (for either treatment 1 or the pooled treatment effect).[6] We distinguish several cases:

- If we find impacts on savings knowledge, time-preferences, and on saving outcomes (all variables in family 1 except savings purpose) we will test the following hypothesis
    1. Is the treatment effect on saving outcomes fully mediated by knowledge and time preferences?
    2. Is the treatment effect on saving outcomes fully mediated by time preferences exclusively?
    3. Is the treatment effect on saving outcomes fully mediated knowledge exclusively?

- If we find impacts on saving outcomes, time-preferences, and on search outcomes (applications per month of search and search duration in month) we will test the following hypothesis
    1. Is the treatment effect on search outcomes fully mediated by saving outcomes and time preferences?

---

[6]We do not include treatment 2, as its theory of change is based on external nudges.

2. Is the treatment effect on search outcomes fully mediated by time preferences exclusively?

3. Is the treatment effect on search outcomes fully mediated by saving outcomes exclusively?

We will conduct a similar analysis for employment outcomes with savings, search, and time preferences as mediators if we find significant impacts. For employment, we will conduct a more nuanced analysis. We will analyse to what extend the purpose of savings (intensified job-search, saving to start a business, and investments in education or longer job-search) matters for the impact of savings on different employment outcomes.

If we find effects on the mediators and outcomes, we will also estimate the indirect treatment effect following section 3.5 of Acharya et al. (2016).

5   Hypothesis Testing

To analyse whether the social media treatment had an effect, we will test for the equality of treatment effects across treatment groups ($\alpha_1 = \alpha_2$) in equation 1 for all families of outcomes. We will also estimate equation 1 with pooled treatment groups if we find no major difference between treatment groups. Similarly, we will also analyse differential treatment effects in equation 2.3 by testing $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$. We will also estimate equation 2 with pooled treatment effects and test for heterogeneity accordingly. For heterogeneous treatment effects we will test $\beta_1 = \beta_3$ and $\beta_2 = \beta_4$. The tests for the instrumental variable estimation will be conducted analogue to equation 1.

We hypothesize that both treatments increase savings and that the social media treatment has a positive effect ($\alpha_2 > \alpha_1 > 0$ for the main outcomes in family 1).

The theoretical predictions for job-search are ambiguous. If participants face credit constraints increased savings could either decrease or increase average search duration: savings could enable job-seekers to search longer for good jobs and/or it could enable them to increase costly search effort (as measured by number of applications) which decreases expected search duration. We will estimate equation 1 for the outcomes in family 2 to test these competing explanations. If credit constraints are not important in this context, search outcomes

might not be affected despite positive savings effects.

Similar to the search outcomes, the predictions for employment are ambiguous. If participants indeed search longer, the treatment groups might exhibit lower employment rates at endline (and perhaps increased rates of probation). On the other hand, if participants increase search intensity in response to the treatment, we would expect higher employment rates at endline. However, increased savings should lead to higher quality jobs as measured by earnings and/or job-characteristics for those who are employed.

5.1 Inference Methods

We construct standard errors using a heteroscedasticity-robust variance matrix estimator. We then construct test statistics using this variance matrix and obtain p-values based on individual tests. To adjust for multiple hypothesis testing, we will obtain q-values that control the false discovery rate across outcomes within each of the outcome families. We will implement the correction for each treatment indicator separately. Following Benjamini et al. (2006), we report the minimum q-value at which we reject that the parameter of interest is zero, rather than pre-specifying a rejection level. We do not include supporting indicators in this adjustment. We do not adjust for multiple testing across outcome families, as outcomes in different families are measuring different concepts that will not necessarily be affected by treatment in the same way.

Given the long time-span between baseline and follow-up survey we expect relative high rates of attrition. We will first test if attrition rates are differential across treatment groups. We will also test if attritted individuals in different treatment groups vary by baseline characteristics by estimating a logistic regression to predict the individual probability of attritting. We will then test for joint insignificance of all coefficients. If we reject this hypothesis, we will re-estimate all specification weighting observations with the inverse predicted probability of attritting. In case of differential attrition, we will also calculate Manski (1990) bounds for the treatment effects at various severity of assumptions and comment at which level potential treatment effects cease to exist (following Blattman et al., 2014).

References

Acharya, A., M. Blackwell, and M. Sen (2016). Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects. *American Political Science Review 110*(03), 512–529.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika 93*(3), 491–507.

Blattman, C., N. Fiala, and S. Martinez (2014). Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda. *Quarterly Journal of Economics*, 697–752.

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *American Economic Review 80*(2), 319–323.