Entrepreneurship Education and Teacher Training in Rwanda
PIs: Moussa Blimpo, Todd Pugatch
Tracer Survey 2021 Analysis Plan
March 28, 2022

1   Overview

This document presents a pre-analysis plan for the Educate! Exchange Program 2021 tracer survey of program alumni. The program ran in 103 randomly assigned Rwandan secondary schools from 2016-2018. Data were also collected from a randomly assigned control group of 104 schools. The 2021 tracer survey was in the field July-August 2021. The survey targeted participants from the 2016 baseline survey by randomly sampling 14 of the 15 students per school (budget constraints prevented surveying all students). All surveys were conducted over the phone using contact information provided in previous rounds. The Kigali office of Innovations for Poverty Action (IPA) conducted the survey.

Previous survey rounds occurred in 2016 (baseline), 2018 (endline), and 2019 (2019 tracer). We completed this pre-analysis plan before analyzing any data from the 2021 tracer survey, but after analyzing data from all previous rounds.

2   Methodology

2.1   Main effects

Our analysis will build on the analysis plans for the 2018 endline and 2019 surveys, which were posted to the AEA Trial Registry (AEARCTR-0001030). Additionally, the endline analysis plan was peer-reviewed and published as a Stage 1 Registered Report in *Journal of Development Economics* (Blimpo and Pugatch 2019). The main statistical method we use is ordinary least squares linear regression (OLS). This is the appropriate method because randomized control trials solve (in theory) the selection problem for estimation of the mean outcome difference due to assignment to treatment. Moreover, OLS allows us to adjust easily for the stratification of treatment, ensuring that we rely on experimental variation. We will cluster standard errors by school to account for correlated outcomes among students within a school, the unit of treatment.

The main results will come from the regression:

$$y_{isg} = \alpha + \beta T_{sg} + \delta X_{0isg} + \gamma_g + \varepsilon_{isg} \tag{1}$$

where i indexes students; s indexes schools; and g indexes strata. The strata are district-school type cells, where school types are public and non-public. In equation (1), y is an outcome; T is an indicator for assignment to treatment; X is a vector of predetermined characteristics; γ is a stratum fixed effect; and ε is an error term. The vector X will include, at minimum, the baseline outcome (if available) and baseline outcomes found to be imbalanced between treatment and control groups in previous analysis (female, an employment dummy, and grit; Blimpo and Pugatch 2021). We will also include any baseline characteristics significantly correlated with differential survey attrition between treatment and control, following equation (3) below. Because randomization occurred

within strata, the strata fixed effects ensure that treatment assignment T is unrelated to the error term. The coefficient of interest is β, which measures the intent to treat (ITT), or the effect of the offer of teacher training T on the mean outcome.

Due to the ongoing COVID-19 pandemic at the time of the survey, attrition may be high and correlated with treatment. We will analyze attrition using the regressions:

$$A_{isg} = \alpha_0 + \alpha_1 T_{sg} + \gamma_g + \varepsilon_{isg} \qquad (2)$$

$$A_{isg} = \delta_0 + \delta_1 T_{sg} + \delta_2 (T_{sg} \times X_{0isg}) + \delta_3 X_{0isg} + \gamma_g + \varepsilon_{isg} \qquad (3)$$

where A is an indicator for attrition, X is the vector of baseline characteristics reported in Table 2 (baseline balance) of our endline analysis (Blimpo and Pugatch 2021), and all else is as in equation (1). In equation (2), the coefficient α_1 measures any differential attrition rate between treatment and control. In equation (3), δ_2 measures any differences in observable characteristics between students who attrit in the treatment and control groups. If we reject either of the null hypotheses $H_0: \alpha_1 = 0$ or $H_0: \delta_2 = 0$ at the 5% level, then our main analysis will re-weight equation (1) by the inverse of the attrition probability, calculated using a logistic regression of equation (2). Table 1 lists the baseline characteristics in X_0.

Table 2 lists the outcomes we plan to test using equation (1).

We will not impute missing values for any dependent variables. For covariates (e.g., baseline outcomes), we will replace missing values with the control group mean and include a dummy for missing values in the regression (Haushofer and Shapiro 2015). To deal with outliers, we will winsorize all financial variables (e.g., income, savings) at the 99th percentile. We will not impute values for outliers in other variables.

2.2 Treatment effect trajectories

To understand the trajectory of treatment effects over time, we will also estimate equation (1) using outcomes from the endline and 2019 tracer surveys. Separate analysis plans detail the outcomes relevant to these surveys. For purposes of the present analysis plan, we will report treatment effect trajectories across all survey waves for repeated outcomes with consistent measurements. These treatment effect trajectories will also adjust for selective attrition using equation (2) if this adjustment is found to be necessary for the 2021 tracer.

2.3 Heterogeneous treatment effects

The effect of the program may differ among students or teachers. Testing for such effects is important as it may point to key policy nuances. We will allow treatment effects to vary according to observable characteristics of a student or school by modifying equation (1) as:

$$y_{isg} = \alpha + \beta_1 T_{sg} + \beta_2 (T_{sg} \times X_{0isg}) + \beta_3 X_{0isg} + \delta y_{0isg} + \gamma_g + \varepsilon_{isg} \qquad (4)$$

where X_0 is some characteristic determined prior to the treatment. A non-zero value of β_2 indicates that the effect of treatment differs according to X_0.

To keep estimation tractable, we will limit estimation of equation (4) to the following student outcomes:

- university enrollment (U1˙univenroll)
- any business involvement [earn˙receive˙401e]
- employment [earn˙receive˙401e]
- total income from business and employment [howmuch˙401ae]

The characteristics X_0 we plan to test are the following:

- gender (BSQ301)
- past academic performance (S3 exam score, BSQ404)
- household socioeconomic status (SES): SES will be measured as an indicator for being above the median of the first principal component of the following variables:
    - parents' education (BSQ310)
    - household assets (BSQ303-308)
    - parents' occupation (BSQ309, indicator for business/professional)

3 References

Blimpo, Moussa P., and Todd Pugatch. 2019. "Entrepreneurship Education and Teacher Training in Rwanda." *Journal of Development Economics* 140 (September): 186–202. https://doi.org/10.1016/j.jdeveco.2019.05.006.

———. 2021. "Entrepreneurship Education and Teacher Training in Rwanda." *Journal of Development Economics* 149 (March): 102583. https://doi.org/10.1016/j.jdeveco.2020.102583.

DellaVigna, Stefano, Nicholas Otis, and Eva Vivalt. 2020. "Forecasting the Results of Experiments: Piloting an Elicitation Strategy." *AEA Papers and Proceedings* 110 (May): 75–79. https://doi.org/10.1257/pandp.20201080.

Haushofer, Johannes, and Jeremy Shapiro. 2015. "Household Response to Income Changes: Pre-Analysis Plan." https://www.socialscienceregistry.org/docs/analysisplan/384/document.

4 Tables

Table 1: Baseline characteristics

characteristic	variable
School characteristics	
boarding school	BHQ109
S4 enrollment, male	BHQ209
S4 enrollment, female	BHQ209
teachers, upper secondary	BHQ210
teacher absence (%), last 3 Tuesdays	BHQ213
currently has electricity	BH216
head teacher knows Skills Lab definition	BHQ611
head teacher considers interactive pedagogies as effective	BHQ614
Teacher characteristics	
female	BTQ200
age	BTQ201
qualified	BTQ202
showed written entrepreneurship lesson plan	BTQ224
comfortable with interactive pedagogies	BTQ300
can calculate business profit	BTQ402
knows definition of business profit	BTQ405
holds another job	BTQ600
Student characteristics	
female	BSQ301
household assets	BSQ306
mother completed primary school	BSQ310
repeating S4	BSQ402
S3 exam score (aggregate)	BSQ404
employed during school holiday	BSQ700
understands compound interest	BSQ803
has savings	BSQ804
can calculate business profit	BS1002
wants to enroll in post-secondary	BSQ1100
plans to start a business	BSQ1102
grit index	BSQ1300-1303

Entrepreneurship Education and Teacher Training in Rwanda
PIs: Moussa Blimpo, Todd Pugatch
Tracer Survey 2021 Analysis Plan
March 28, 2022

Table 2: Outcomes

Outcome	variable name				notes
	2021 tracer	2019 tracer	end-line	baseline	
attrition					
Calendar activity					
secondary enrollment	ac0_educsecond	ac0_educsecond			separate outcomes by month
university enrollment	ac1_educuniv	ac1_educuniv			separate outcomes by month
training enrollment	ac2_eductrain	ac2_eductrain			separate outcomes by month
business/management	ac3_bus_mgt	ac3_bus_mgt			separate outcomes by month
employment (paid)	ac4_paid_employ	ac4_paid_employ			separate outcomes by month
Academic activity					
completed secondary school	S1_complt_status	S1_complt_status			
completed TVET course	S5_complt_vstytvet				
enrolled in university	U1_univenroll	U1_univenroll			
enrolled in TVET	U1_univenroll	U1_univenroll			
Household formation					
married	marital_st				
any children	no_children				
number of children	no_children				
age at first birth	age_child, respo_age				Condition on having any children.
desired fertility	ideal_children				
homeowner	house_201				
partner schooling	schooling_level				if married or in committed relationship

Entrepreneurship Education and Teacher Training in Rwanda
PIs: Moussa Blimpo, Todd Pugatch
Tracer Survey 2021 Analysis Plan
March 28, 2022

female empowerment	hhhead, financial_decision, fin_opinion, decision_work, ideal children, no_children				female sample only. Defined as mean(HH head, makes most financial decisions, decides whether to work outside home by self or jointly with spouse/partner; thinks most financial decisions should be made by self or jointly; 1(desired fertility>=number of children)
progressive HH attitude	fin_opinion, decision_partner_work				male sample only. Defined as mean(financial decisions, decision to work outside home should be made by female or jointly)

Skills

grit	g1-g12	g1-g12	ESQ1300-1303	BSQ1300-1303	Outcome is 12-item scale, while baseline outcome is 4-item scale.
belief in future	bf1-bf7	bf1-bf7		BSQ1200/1202/1204/1206/1208	Baseline outcome is locus of control
entrepreneurial spirit	e1-e8	e1-e8			Based on Bruhn et al, *Journal of Political Economy* 2018
monthly discount rate<100%	trust_offered_801	moneyoffer_801	ESQ801	BSQ801	in-person survey only for 2019 tracer
work skills	ss1-ss10	ss1-ss8			From World Bank STEP Skills survey. 2021 tracer version adds 2 questions on internet use.
creativity	a59_totaluses	use_pole			number of uses of a pole

Economic activity

business participation: all	earn_receive_401e	earn_receive_401e	ESQ401	BSQ600	will test against prediction reported in DellaVigna, Otis, and Vivalt (2020), in addition to null hypothesis of zero
business participation: student business club	stud_busi_club_412	type_busi_402	ESQ402	BSQ604	
non-agricultural business	type_busi_407	type_busi_407	ESQ407	BSQ601	
business has paid employees	paid_worker_406	paid_worker_406	ESQ406	BSQ600	
business revenue	lasttime_sales_410				
business profit	often_earn_409, lasttime_profits_410	often_earn_409, lasttime_profits_410	ESQ409-410	BSQ608	

Entrepreneurship Education and Teacher Training in Rwanda
PIs: Moussa Blimpo, Todd Pugatch
Tracer Survey 2021 Analysis Plan
March 28, 2022

business asset value	asset_value				
borrowed money for business	borrowed_800	borrow_800			in-person survey only for 2019 tracer
formal business	busreg_507	busregistered_stp19			
business survived since 2019	same business, business_reported				if had business in 2019 tracer survey
expect business growth	expectation_422				any response other than "likely to remain the same" or "likely to shrink"
employed	earn_receive_401e	earn_receive_401e	ESQ401	BSQ500	
wages	often_pay_617, howmuch_paid_618				
Economic security					
not in employment, education, or training (NEET)	earn_receive_401e, S1_complt_status, U1_univenroll	neet_stp19	neet_el		
income (total)	howmuch_401ae	howmuch_401ae	ESQ401a	BSQ503	
assets	livstc/motorcycle/car/bicycle/boat/computer/land/house				1st principal component of assets
savings (any)	save_700	savedmon_804e	ESQ804	BSQ804	
savings (amount, conditional on any)	howmuch_saved_701	savedamt_805e	ESQ805	BSQ805	
economic security index	cpq* series				mean of coping strategy variables. Source: Chioda & Gertler, Educate! SEED evaluation.

Variable names for 2021 tracer survey from tracer2_bench.xlsx. For HH formation variables, we will report results separately by male/female.