

Can Soft Skills Training Improve Enterprise and Employment Growth in Tanzania? A Randomized Evaluation of the Strengthening Rural Youth Development through Enterprise (STRYDE) 2.0 Program

Pre-analysis plan¹

July 2019

Margherita Calderone, University of Turin and ODI

Nathan Fiala, University of Connecticut and RWI

Lemayon Melyoki, University of Dar es Salaam

Annekathrin Schoofs, University of Passau and RWI

Rachel Steinacher, Innovations for Poverty Action

¹ This document outlines the pre-analysis plan for a randomized evaluation of the STRYDE 2.0 program – a comprehensive youth employment program implemented by TechnoServe and funded by MasterCard Foundation in urban and rural Tanzania.

1. Description of the sample to be used in this study

The research team partnered with TechnoServe (TNS), a non-governmental organization, to evaluate their Strengthening Rural Youth Development through Enterprise (STRYDE) 2.0 program using a cluster randomized controlled trial. The STRYDE 2.0 program in Tanzania aims at targeting 10,800 youth aged between 18 and 30 from two regions located in the southwest of Tanzania - Mbeya and Songwe region - to be trained in six cohorts between 2015 and 2019.

This study follows cohort number four of STRYDE 2.0 participants. The cohort was mobilized between April and June 2017, trained between July and October 2017, awarded in a Business Plan Competition (BPC) of December 2017, and supported during the aftercare through spring 2018. Our goal is to understand the impact of the STRYDE 2.0 program, as well as the additional impact of receiving capital to buy materials and tools to run a small business through the program's BPC.

In cohort four, mobilization was attempted in 161 clusters: 9 wards in the Mbeya Urban area and 152 villages in the rural areas of the Mbeya and Songwe regions. Mobilization was conducted in villages grouped in pairs based on location: each pair was close enough to allow trainers (hereinafter referred to as Business Counselors, BC) to access both, but far enough to minimize spillovers. Mobilization was considered successfully completed if at least 20 participants per cluster completed the baseline. Once the mobilization was completed, clusters were randomly assigned to treatment or control within each BC strata. The final baseline sample includes 135 clusters and 4,537 observations divided between 72 treatment and 63 control clusters.

For the endline, we plan to follow-up about 18 individuals per cluster for a total of 2,430 observations. These individuals will be randomly selected from the full baseline sample. This selection will be proportional to the percentage of people in the cluster who participated in the BPC and those who did not. In line with these percentages, we will also re-weight all remaining BPC participants included in the baseline and add them to the endline sample. This sample sums up to 3,808 observations. The experimental sample for the evaluation of the BPC includes all participants from 88 clusters² for a total of 1,496 observations, plus 249 standard winners. The BPC sample is partially overlapping with the STRYDE endline as it involves the groups of the

² These include 75 clusters/ training groups from the experimental sample (i.e. 72 clusters comprising 69 rural villages and 3 urban wards with two training groups each) and 13 clusters from the non-experimental one.

BCs that were not part of the impact evaluation as well as the groups, and students, of the other BCs who were part of the evaluation but allowed to re-mobilize participants as needed³.

For tests of randomization balance, we check balance of the following variables: household demographics and composition, involvement in income generating activities and income, locus of control, as well as questions about motivation and obstacles for attending the training.

2. Key data sources

Data for the study come from both individual and household surveys and administrative attendance data provided by TNS.

In June 2017, participants who signed up for the program also hand-filled a brief questionnaire which collected baseline information of respondents. This information includes key socio-economic characteristics of respondents, such as basic demographics, living situation, involvement in different income generating activities, income and its sources, and spending patterns and savings. It also includes a series of questions aimed at measuring locus of control, as well as specific questions about motivation for wanting to enroll into the business training program, expected income post-training, and fears about possible challenges to participate in the training.

At the end of August 2017, baseline respondents were re-interviewed with a 2-minute phone-based survey during which they were asked to confirm their identity and contact details, as well as to identify a family member which could have been contacted for an additional household survey⁴.

The household survey took approximately 20 minutes and was conducted via phone⁵. The

³ The 40 TNS BCs involved in the impact evaluation committed to work as hard as possible on keeping IPA-registered youth engaged. However, they were also allowed to re-mobilize participants in whatever strategy they saw fit in order to reach their target of two training groups of minimum 32 students. In addition, 4 BCs worked outside of the experimental sample either because they were assigned to areas in which the mobilization failed or because, for personal reasons, their work location was not flexible and, hence, were spared from being part of the randomized evaluation.

⁴ Married people were asked to provide contact details of their spouse, while unmarried individuals were asked to provide contact details of their household head –if different from self.

⁵ Calls started at the end of August 2017 and stopped in October, when most of the participants were about to graduate from training and the information that would have been collected could no longer be regarded as baseline data. In such a short period, the data collection team managed to attempt calling 87% of the overall sample: 3,948 respondents, 2,150 from the treatment group and 1,798 from the control. Within each cluster, respondents were randomly sorted to be reached during the follow-up. Everyone from the sorted list was called three times, with calls being made at least one day apart, before being replaced for not responding by the next person on the list. Interviews were considered completed for that cluster once 25 (50 in urban areas) respondents picked-up the call, consented to confirm their identity, and provided the details of a family member to contact for the household survey. Out of 3,948 individuals,

questionnaire focuses on measuring how individuals interact with their families in order to better understand how families affect the investment decisions of participants. The household survey elicits data regarding key household characteristics, training decisions by youth and their families, expectations about earnings after training, and aspiration levels.

Furthermore, we use administrative data by TNS to determine attendance rates by matching our baseline data and TNS data.

The endline survey will be implemented in the middle of 2019, two years after the end of the STRYDE 2.0 program. It will be collected using face-to-face interviews and will take approximately 1 hour to administer.

3. Hypotheses to be tested through the causal chain

We present our hypotheses to test in this section. We refer to the treatment arm of receiving the STRYDE 2.0 program as treatment A and to the treatment arm of receiving funds from the BPC as treatment B.

- a. H_0/H_a : No impact (positive impact) of receiving Treatment A mainly on participants'
 - i Employment,
 - ii Economic status,
 - iii Entrepreneurial and career skills,
 - iv Psychological characteristics, and
 - v Social outcomes.
- b. H_0/H_a : No impact (positive impact) of receiving Treatment B mainly on participants'
 - i Employment,
 - ii Economic status,
 - iii Entrepreneurial and career skills, and
 - iv Psychological characteristics, and
 - v Social outcomes.

the team reached 2,811 respondents -71% of the sample. Out of the 2,811 respondents that picked-up the phone, 2,548 (i.e. 91%) consented to participate in the survey and confirm their identity and 2,192 (78%) provided contact details of a family member. In the short time remaining, the team managed to attempt calling 1,842 family members -84% of the sample of respondents that provided a contact. Out of a sample of 1,842 observations, 1,507 (82%) respondents were reached and 1,425 (77%) consented to participate in the additional household survey.

- c. Heterogeneity analysis according to the following dimensions
 - i Gender of the participant, and
 - ii The role of the family unit in the ability of youth to take-up the training program, attend classes, graduate, and benefit from it.

4. Measurement of key variables

The primary outcomes are the following:

- Employment. The relevant measures will include:
 - Economic activities (formal or informal employment status, income generating activities)
 - Quantity and quality of work (employed days per month, daily employment hours per month, job satisfaction)
 - Duration of work (permanent, seasonal, or short-term)
 - Vulnerability of work (good and bad month earnings, months the business was not operational)

- Measures on economic welfare, personal finance and working capital. The relevant measures will include:
 - Respondent is main income contributor
 - Cash earnings
 - Business performance (employees, profits, collaboration, location)
 - Savings behavior (type and amount of savings, reasons to save)
 - Borrowing behavior (type and amount of loan, business investments)
 - Financial resilience
 - Income hiding
 - Household assets

The secondary outcomes are the following:

- Entrepreneurial and career skills. The relevant measures will include:
 - Self-confidence in job applications (ability to find information about job opportunities in the community and CV writing)

- Self-confidence in entrepreneurial activities (ability to run an own business, obtain credit and bargain cheap prices for business)
- Business practices and business plan competition
- Financial knowledge (numeracy, attitudes, and awareness)
- Psychological characteristics. The relevant measures will include:
 - Motivation to start a new business
 - Self-regulation (e.g. self-control, risk-taking)
 - Well-being (current status of life satisfaction, future life optimism)
 - Business and personal aspirations
 - Locus of control
 - Grit
- Social outcomes. The relevant measures will include:
 - Decision making power (e.g. consumption, investments, savings, and non-financial)
 - Social capital (number of memberships in community groups, leadership position(s), and trust)
 - Gender empowerment / pro-gender-equality beliefs
 - Partner selection (preferences, current relationship status, fertility preferences, sexual behavior)
 - Intimate partner violence (attitudes towards and acceptability of gender-based violence, experiences with intimate partner violence)
 - Violent behavior in the community

5. Sample size and power calculations

The baseline sample for the impact evaluation of the STRYDE 2.0 program includes 135 clusters, 72 treatment and 63 control clusters, with a mean of 33.61 individuals per cluster (and a variance of 7.474) and a total of 4,537 observations.

We consider that program effects of 20-25% will be necessary to justify the costs of the STRYDE 2.0 program. Given the expected take-up rate, the adjusted effect size will be around 11%.

The experimental sample for the evaluation of the STRYDE 2.0 program targets a total of 3,808 completed surveys. This sample includes about 18 individuals per cluster for a total of around 2,430 observations and all BPC participants who were also interviewed during baseline and regardless of whether they were already randomly selected in the 18 individuals per cluster or not. Power calculations, conducted in Optimal Design using baseline data, suggest that such a sample would be well powered (attaining a power level of 80%) to detect effect sizes of 11% in an indicator of monthly income.

The experimental sample for the evaluation of the BPC includes 88 clusters for a total of 1,496 observations, plus 249 standard winners. We plan to include all of them in the endline survey. Power calculations suggest that this sample would be well powered to detect effect sizes on income of about 11% for the BPC alone.

6. Treatment effect equation to be estimated

To test hypotheses outlined in section 3a, we will employ an ANCOVA specification (McKenzie, 2012; Bruhn and McKenzie, 2009). We will be estimating the following simple model using Ordinary Least Squares, which regresses the outcome indicator on the treatment status of individual i controlling for the baseline value of the indicator:

$$(1) \quad Y_{i\ Post} = \alpha + \beta T_i + \eta Y_{i\ Pre} + \delta X_{i\ Pre} + \varepsilon_{i\ Post}$$

where $Y_{i\ Post}$ represents the different outcomes of interest as outlined above (3a.i, 3a.ii, 3a.iii, 3a.iv, 3a.v) for individual i , measured after the intervention. T_i is a dummy variable equal to 1 if the individual was registered in a cluster randomly selected to receive the STRYDE 2.0 program and 0 if not. $X_{i\ Pre}$ represents control variables unbalanced at baseline. The estimation includes BC fixed effects because the randomization was implemented within BC strata. Finally, $\varepsilon_{i\ Post}$ represents the unobserved individual-specific residual. Standard errors will be adjusted for clustering at the ward / village level. The Intent-To-Treat effect of the program will be estimated by β .

We then assess the effects of the BPC cash intervention (see 3b) by modifying equation (1) as follows:

$$(2) \quad Y_{i\text{ Post}} = \alpha + \beta C_i + \eta Y_{i\text{ Pre}} + \delta X_{i\text{ Pre}} + \varepsilon_{i\text{ Post}}$$

where C_i is a binary variable equal to 1 if the individual from the treatment group was randomly assigned to receive the cash of the BPC and 0 if not, while other variables are the same as those defined in equation (1).

To test hypotheses outlined in 3c, we next conduct a heterogeneity analyses that allows estimation of the impact according to gender and family background. Heterogeneous treatment effects will be obtained by estimating (1) and (2) with an additional interaction effect that interacts treatment status with the variable of interest, as illustrated in (3) and (4), respectively:

$$(3) \quad Y_{i\text{ Post}} = \alpha + \beta_\tau \tau_i + \beta_{1 \times \tau} (T_i \times \tau_i) + \beta_1 T_i + \eta Y_{i\text{ Pre}} + \delta X_{i\text{ Pre}} + \varepsilon_{i\text{ Post}}$$

$$(4) \quad Y_{i\text{ Post}} = \alpha + \beta_\tau \tau_i + \beta_{1 \times \tau} (C_i \times \tau_i) + \beta_1 C_i + \eta Y_{i\text{ Pre}} + \delta X_{i\text{ Pre}} + \varepsilon_{i\text{ Post}}$$

where the variable τ_i either indicates gender or the role of the family unit.

7. Plan for how to deal with multiple outcomes and multiple hypothesis testing

We have a relatively rich set of outcome measures to explore treatment effects along various interesting dimensions. To deal with multiple hypothesis testing, we will employ two different approaches.

First, we will group our outcome measures into domains where items within a domain are measuring an underlying common factor. The five domains we are interested in are: employment, economic status, entrepreneurial and career skills, psychological characteristics, and social outcomes.

Second, on top of the standard model described in the section above, within each domain and across domain indexes, we will also calculate the Family-Wise Error Rate (FWER) adjusted p-values using the Westfall and Young step-down resampling method. The FWER represents the probability that at least one hypothesis out of a family of hypotheses is falsely rejected (type-1 error). Hence, the FWER results will be used to account for the multiple inference problem which increases the likelihood that some of the outcomes are statistically significant by chance even if there is no treatment effect.

8. Procedures to be used for addressing survey attrition, outliers and missing data

8.1. Attrition

We plan to use an extensive tracking exercise after the conclusion of the standard survey. We will perform a probability reweighting of the data and calculate the effective survey follow-up rate. To understand the drivers of attrition, we will estimate the following equation:

$$(5) \quad AT_{i\ Post} = \alpha + \beta T_i + \eta Y_{i\ Pre} + \delta X_{i\ Pre} + \varepsilon_{i\ Post}$$

where AT_i is an attrition indicator and the other variables are the same as those defined in equation (1). If the treatment status variable is significant at the 5% level, we will also perform a bounding exercise as suggested in Lee (2009). This semi-parametric approach relies on relatively weak assumptions about how a randomly assigned treatment influences outcomes of interest to obtain intervals on the estimated size of the treatment effect in the presence of non-random attrition.

8.2. Procedures to be used for addressing outliers

At endline, monetary values will be top-censored at the 99th percentile to contain outliers.

8.3. Procedures to be used for addressing missing covariate values

We will follow Lin and Green (2016) in treating missing covariates. If no more than 10 percent of the covariate's values are missing, we will recode the missing values to the overall mean (again testing sensitivity of estimates to these approaches by comparing results with those obtained from the sample with non-missing covariates). If more than 10 percent of the covariate's values are missing, we will include a missingness dummy as an additional covariate and recode missing values to 0.

8.4. Procedures to be used for addressing missing dependent variables

To deal with missing values on our outcome measures, we will adopt the approach described in Kling et al. (2007) and impute missing values by setting them equal to the mean of the respective outcome variable for the relevant treatment group, and testing sensitivity of main coefficient

estimates to this approach by comparing results with those obtained from the sample with non-missing outcome variables.

References

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4), 200-232.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1), 83-119.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies*, 76 (3), 1071-1102.

Lin, W., & Green, D. P. (2016). Standard operating procedures: A safety net for pre-analysis plans. *PS: Political Science & Politics*, 49(3), 495-500.

McKenzie, D. (2012). Beyond Baseline and Follow-up: The Case for More T in Experiments. *Journal of Development Economics*, 99: 210-221.