

Selection or More Capital? Experimental Evidence on High-Growth Entrepreneurs in Kenya

Authors: Francisco Campos, Abba Safir, Celine Koffka, David McKenzie and Bilal Zia

Contents

1. Overview
2. Data Availability
3. Instructions for Replicators
4. List of Exhibits
5. Requirements
6. Code Description
7. Folder Structure

Overview

This replication package reproduces the results presented in "Who Should We Fund, and How Much Should We Give Them? Experimental Evidence from a Kenyan Business Plan Competition". The study embeds two experiments within a Kenyan business plan competition to test how selection intensity and grant size determine entrepreneurial outcomes. Applicants received either a US\$9,000 grant after a streamlined process, or went through multiple stages and were randomly assigned US\$9,000 or US\$36,000. All grants generated jobs, but impacts on firm performance only persist over three years under multi-stage selection. The larger grants did not yield significantly greater long-term impacts than the multi-stage US\$9,000 grant, suggesting diminishing returns and limited lumpy investment opportunities. Selection, rather than grant size, determines long-run firm growth.

The package reproduces the paper results from raw data stored in the Data/Original directory. All analyses are implemented in Stata 19.5 and can be executed through the master script:

master file.do

Replicators should begin by ensuring that all required datasets are available in the appropriate folders and that the required Stata packages listed in the Requirements section are installed.

The master file.do script executes the cleaning, analysis, and output-generation procedures required to reproduce the paper's results.

The replication package was developed and tested on macOS.

Some procedures are computationally intensive. In particular, the generation of Table T1 and Table C2 may require substantial runtime.

Runtime: 2h20min

Data Availability

All data is available under licensed access in the World Bank Microdata Library (MDL).

Data name: baseline_stream.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Baseline survey data for applicants assigned to the streamlined selection process. Collected in February 2020 prior to grant assignment.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: baseline_multi.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Baseline survey data for finalists in the multi-stage selection process. Collected January–February 2021 prior to grant assignment.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: treatment_status.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Final treatment assignment file identifying selection pathway, treatment status, control status, and grant amount assigned to study participants.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya

Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: 1_year_fu.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: First follow-up survey conducted between May and August 2022, measuring short-term impacts approximately one year after grant disbursement.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: 2_year_fu.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Second follow-up survey conducted between February and May 2023, measuring outcomes approximately two years after grant disbursement.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: 3_year_fu.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Third follow-up survey conducted between July 2024 and March 2025, measuring medium-term outcomes approximately three to four years after grant disbursement.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: intake_info.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Application-stage information collected from applicants between June and September 2019, including applicant characteristics, business information, and proposal details.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: eligible_youth.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Dataset containing all applicants deemed eligible after screening for program eligibility and application completeness.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: scores.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Initial application scores from the first-round screening process used to evaluate eligible applicants and determine progression through the competition.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: judges_score.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Scores assigned by independent judges during the business plan evaluation

stage of the multi-stage selection process.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data name: Treatment_status.dta

Source: Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025)

URL: Forthcoming on the World Bank Microdata Library

Note: Final treatment status to which a respondent was assigned.

Access Date: [Insert access date after publication]

Access Type: Licensed Access

License: Licensed Access (World Bank Microdata Library)

License URL: To be assigned upon publication

Citation: Campos, F., Safir, A., Koffka, C., McKenzie, D., & Zia, B. (2026). Kenya Youth Employment and Opportunities Project (KYEOP) – MbeleNaBiz Business Plan Competition Impact Evaluation (2019–2025) [dataset]. World Bank Microdata Library. Forthcoming.

Data Sources

The original source datasets used for this study contain personally identifiable information (PII) and therefore cannot be distributed through the replication package or publicly released.

Prior to preparing the replication package and World Bank Microdata Library (MDL) deposit, the original datasets were processed using the script: cleaning for replication package.do

This script performs the disclosure review and de-identification process required for public distribution. Specifically, it removes personally identifiable information and other confidential variables that are not required to reproduce the analysis presented in the paper. The script then saves de-identified versions of the datasets for use in the replication package and MDL archive.

Because this script is a one-time data release and confidentiality procedure, it is not called by master file.do and is not required to reproduce the paper results. All analysis and replication code included in the package begins from the de-identified datasets generated by this script.

The version of the data that will be made available through the World Bank Microdata Library (MDL) under Licensed Access is the de-identified version created by cleaning for replication package.do. The original datasets containing personally identifiable information will not be distributed.

The following de-identified datasets are included in the replication package and correspond to the files that will be deposited in the MDL archive:

- 1_year_fu.dta
One-year follow-up survey data collected from participating entrepreneurs.
- 2_year_fu.dta
Two-year follow-up survey data collected from participating entrepreneurs.
- 3_year_fu.dta
Three-year follow-up survey data collected from participating entrepreneurs.
- baseline_multi.dta
Baseline survey data for applicants participating in the multi-stage competition process.
- baseline_stream.dta
Baseline survey data for applicants participating in the streamlined competition process.
- Eligible_youths.dta
Dataset containing information on respondents eligible for the multi-stage competition.
- intake_info.dta
Intake information for all competition applicants.
- scores.dta
Competition scoring data used during the evaluation and selection process.
- Treatment_status.dta
Treatment assignment dataset identifying the randomized allocation of winners across treatment groups.

All files are provided in Stata.dta format and are sufficient to reproduce the analysis presented in the manuscript.

Statement about Rights

I certify that the authors of the paper have legitimate access to and permission to use the data used in this paper.

Instructions for Replicators

To reproduce all results in the manuscript, users should follow the steps below.

Download and unzip the full replication package while preserving the folder structure.

Open the file:
master file.do

update the root directory path in the master file to match the local location of the replication package. The following section of code must be modified:

```
global directory "/Users/celine/Library/Mobile  
Documents/com~apple~CloudDocs/WB/KYEOP/BPC/Data Analysis/Replication package/"
```

```
* Set directory  
cd "${directory}"  
* PATHS  
global raw_data      "Data/Original/"  
global modified_data "Data/Modified/"  
global output        "Output/"  
global do_files       "Do-files/"
```

Users should replace the value assigned to global directory with the path to their local copy of the replication package.

Ensure that all required Stata packages listed in the Requirements section are installed.

Run:

master file.do

The master script automatically executes all required cleaning, construction, analysis, and output-generation scripts in sequence. The package reproduces the analysis directly from the de-identified raw datasets provided in Data/Original.

For reproducibility, the code sets random seeds where required.

Some procedures are computationally intensive. In particular, the generation of Table T1 and Table C2 may require substantial runtime.

List of Exhibits

The provided code reproduces:

All numbers provided in text in the paper

All tables and figures in the paper

All exhibits are generated automatically through execution of master file.do.

Output files are stored in the following directories:

Output/Main/Tables

Output/Main/Figures

Output/Annex/Tables

Output/Annex/Figures

Tables are exported primarily in .xlsx format and figures in .png format. In some cases, a single manuscript table is constructed from multiple exported output files. For example, separate outputs corresponding to different panels or outcome categories (e.g., t5_profits.xlsx and t5_sales.xlsx) may need to be combined to reproduce the final formatted manuscript table. No manual editing of statistical results is performed.

Requirements

Computational Requirements

The replication package was developed and tested on macOS using Stata 19.5. Replicators should ensure that all required user-written Stata packages are installed prior to running the code.

The package reproduces all results through execution of:

master file.do

No additional software beyond Stata is required.

Software Requirements

- **Stata version 19.5**
 - estout
 - tabmiss
 - ritest (net install ritest)
 - fre

The required packages are not installed automatically by the replication code and must be installed manually by the user prior to execution.

Memory and Runtime and Storage Requirements

The replication package includes all required datasets and output directories.

Runtime: 2h20min

Some procedures are computationally intensive. In particular, the generation of Table T1 and Table C2 may require substantial runtime.

Code Description

The replication package is organized around a central master script:

master file.do

The master script:

- sets the working directory and global paths,
- opens and stores log files,
- initializes timers for runtime tracking,
- sets reproducibility parameters including random seeds where necessary,
- and sequentially executes all cleaning and analysis scripts required to reproduce the paper results.

The replication workflow proceeds in two stages: data cleaning and data analysis.

Data Cleaning:

The cleaning scripts construct harmonized analysis datasets from the original de-identified survey and administrative data sources. These scripts generate intermediate and final datasets stored in the Data/Modified directory.

The main cleaning scripts include:

- Prep baseline stream.do Cleans baseline data from the streamlined BPC process
- Prep baseline multi.do Cleans baseline data from the multi-stage BPC process
- Prep 1 year follow-up.do Cleans one-year follow-up survey data
- Prep 2 year follow-up.do Cleans two-year follow-up survey data
- Prep 3 year follow-up.do Cleans three-year follow-up survey data
- Final datasets pooled.do Creates pooled and final analysis datasets

Data Analysis:

The analysis scripts generate all tables and figures included in the manuscript and appendix. Output files are written to the Output/Main and Output/Annex directories.

Main manuscript exhibits are generated through dedicated scripts, including:

- table 1.do – table 9.do Generate main manuscript tables
- figure 2.do and figure 3.do Generate main manuscript figures
- tables 5, 6, A5, A6.do Generates Tables 5–6 and Appendix Tables A5–A6 jointly due to shared sharpened q-value computations

Appendix exhibits are generated through additional scripts:

- table A1.do – table A4.do Generate Appendix A tables
- table B1.do Generate Appendix B table
- table C1.do and table C2.do Generate Appendix C tables
- table E1, E2.do Generate Appendix E tables
- figure F1.do Generate Appendix figure
- table F1.do – table F3.do Generate Appendix F tables

The package is designed for one-click execution through master file.do. Users are not expected to run individual scripts manually unless reproducing selected outputs only.

Intermediate datasets are stored in:

Data/Modified

Final tables and figures are exported primarily as:

.xlsx
.png

Some manuscript tables are assembled from multiple exported output files corresponding to different panels or outcome groups. No manual editing of statistical results is performed.

Folder Structure

The folder structure looks as follows and needs to be replicated in order to run the entire code on one-click using the "Master file.do":

```
Data
├── Original
└── Modified

Do-files
├── Master file.do
├── 1. cleaning.do
└── 2. analysis.do

Output
├── Main
│   ├── Tables
│   └── Figures
└── Annex
    ├── Tables
    └── Figures
```