

Reproducibility Package – README

Gender and Goals Matter for Youth Employment:Returns to Socio-Emotional Skills Training in Tanzania

This folder contains the full reproducibility package for the Tanzania SES RCT analysis. The package is organised into three main folders. Users only need to run the master do-file after updating the local path to the package on their computer.

Data Availability Statement

All datasets used in this replication package were created as part of the project:

SEL Tanzania – Soft Skills Training Impact Evaluation

Data were collected by the project team and are distributed in de-identified form within this replication package. No datasets were downloaded from external repositories.

Baseline Survey Dataset

Filename: SES_Tan_Baseline_labelled_noPII_25Feb2022.dta

Source: World Bank Africa Gender Innovation Lab & Innovations for Poverty Action

URL: To be provided upon archival.

Access date: Not applicable; project-generated data.

Citation: Cassidy, R.; Das, S.; Delavallade, C.; Kipchumba, E.; Munshi Sulaiman, M. (2022).

SEL Tanzania Baseline Survey (Labelled, No Personally Identifiable Information).

Data set. World Bank Africa Gender Innovation Lab & Innovations for Poverty Action.

Access: Forthcoming in the World Bank Microdata Library (MDL).

License: Confidential research data; de-identified files distributed within this replication package.

Endline Survey Dataset

Filename: SES_Tan_Endline_labelled_noPII_24March2022.dta

Source: World Bank Africa Gender Innovation Lab & Innovations for Poverty Action

URL: To be provided upon archival.

Access date: Not applicable; project-generated data.

Citation: Cassidy, R.; Das,S.; Delavallade, C.; Kipchumba, E.; Munshi Sulaiman, M. (2022).

SEL Tanzania Endline Survey (Labelled, No Personally Identifiable Information).

Data set. World Bank Africa Gender Innovation Lab & Innovations for Poverty Action.

Access: Forthcoming in the World Bank Microdata Library (MDL).

License: Confidential research data; de-identified files distributed within this replication package.

Second Endline Survey Dataset

Filename: SES_Tan_2nd_Endline_labelled_noPII_24Nov2022.dta

Source: World Bank Africa Gender Innovation Lab & Innovations for Poverty Action

URL: To be provided upon archival.

Access date: Not applicable; project-generated data.

Citation: Cassidy, R.; Das, S.; Delavallade, C.; Kipchumba, E.; Munshi Sulaiman, M. (2022). *SEL Tanzania Second Endline Survey (Labelled, No Personally Identifiable Information)*.

Data set. World Bank Africa Gender Innovation Lab & Innovations for Poverty Action.

Access: Forthcoming in the World Bank Microdata Library (MDL).

License: Confidential research data; de-identified files distributed within this replication package.

Training Attendance and Feedback Dataset

Filename: Soft Skills Tanzania Trainee Attendance and Feedback_Clean.dta

Source: World Bank Africa Gender Innovation Lab & Innovations for Poverty Action

URL: To be provided upon archival.

Access date: Not applicable; project-generated data.

Citation

Cassidy, R.; Das, S.; Delavallade, C.; Kipchumba, E.; Munshi Sulaiman, M. (2022).

SEL Tanzania Training Attendance and Feedback Data.

Data set. World Bank Africa Gender Innovation Lab & Innovations for Poverty Action.

Access: Forthcoming in the World Bank Microdata Library (MDL).

License: Confidential research data; de-identified files distributed within this replication package.

Constructed SES Scores Dataset

Filename: SES_scores_Tan_pooled.dta

Source: World Bank Africa Gender Innovation Lab & Innovations for Poverty Action

URL: To be provided upon archival.

Access date: Not applicable; project-generated data.

Citation

Cassidy, R.; Das, S.; Delavallade, C.; Kipchumba, E.; Munshi Sulaiman, M. (2022).

Pooled Socio-Emotional Skills Scores Dataset (Derived Analysis File).

Derived data set. World Bank Africa Gender Innovation Lab & Innovations for Poverty Action.

Access: Forthcoming in the World Bank Microdata Library (MDL).

License: Confidential research data; de-identified files distributed within this replication package.

Cleaned Panel Analysis Dataset

Filename: SES_Tan_Panel_cleaned_12Sep2023.dta

Source: World Bank Africa Gender Innovation Lab & Innovations for Poverty Action

URL: To be provided upon archival.

Access date: Not applicable; project-generated data.

Citation

Cassidy, R.; Das, S.; Delavallade, C.; Kipchumba, E.; Munshi Sulaiman, M. (2023).

SEL Tanzania Panel Dataset (Cleaned Analysis File).

Derived data set. World Bank Africa Gender Innovation Lab & Innovations for Poverty Action.

Access: Forthcoming in the World Bank Microdata Library (MDL).

License: Confidential research data; de-identified files distributed within this replication package.

Folder structure

1_Data

This folder contains all the data used in the analysis.

- 2_Intermediate: Partially processed datasets (de-identified and labelled) for each survey round.
- 3_Final: Fully cleaned, anonymised, and analysis-ready dataset used by all regressions and figures.

No manual changes are required in this folder when reproducing the results.

2_Do

This folder contains all Stata do-files used in the analysis. They are designed to be run sequentially through a single master do-file.

List of do-files:

- 0_SES_Tan_Master.do (or run_all.do)
 - Master script that runs all do-files in the correct order
 - The only file the user needs to run to reproduce all results (except table A3, see below). **Update file path before running.**
- 2_SES_Tan_Variable_Creation_02Feb2026.do
 - Creates standardized variables and indices
 - Constructs aggregate measures (self-reported, behavioral, gaps)
 - Generates interaction terms and analysis-ready variables
- 3_SES_Tan_Impact_Evaluation_02Feb2026.do
 - Produces descriptive statistics and balance tables
 - Runs all regression specifications
 - Generates regression tables and figures for the paper and appendix

3_Output

This folder contains all outputs generated by the analysis.

All empirical tables and figures reported in the manuscript are reproducible from the replication package using the provided code and data. Conceptual diagrams and literature-review exhibits are not generated by statistical code and are therefore not included in the replication outputs.

Tables

Exhibit	Description	Replication Output File
Table 1	Effect of SES Training on Aggregate SES	SES_Aggregate_20240712.tex

Table 2	Effect of SES Training on Labour Market Outcomes	Secondary_Outcomes_20240712.tex
Table 3	Effect of SES Training on Gender Outcomes and Psychological Well-Being	Gender_Psychological_Outcomes_20240712.tex
Table 4	Labour Market Outcomes Heterogeneity by Job-Search Status	Secondary_Outcomes_ht_jobsearch_20240712.tex
Table A1	Baseline Characteristics Balance	balance.tex
Table A2	Training Uptake	traininguptake.tex
Table A3	Overconfidence and Job-Seeking Behaviour at Baseline	Overconfidence_baseline.tex
Table A4	Aggregate SES Effects by Sex	SES_Aggregate_female_20240712.tex
Table A5	Behavioural SES Heterogeneity – Awareness	SES_Aggregate_Beh_med_s36_1_20240712.tex
Table A6	Behavioural SES Heterogeneity – Management	SES_Aggregate_Beh_med_s37_1_20240712.tex
Table A7	Behavioural SES Heterogeneity – Cognitive Ability	SES_Aggregate_Beh_med_cogability_20240712.tex
Table A8	Behavioural SES Heterogeneity – Job Search	SES_Aggregate_Beh_ht_jobsearch_20240712.tex
Table A9	Self-Reported SES Heterogeneity – Social Desirability	SES_Aggregate_SR_med_sdi_20240712.tex
Table A10	Self-Reported SES Heterogeneity – Job Search	SES_Aggregate_SR_ht_jobsearch_20240712.tex

Figures

Exhibit	Description	Replication Output File(s)
Figure 1	Effects of SES Training on Overconfidence	SES_Overconfidence_20240712.pdf
Figure A3	Labour Market Outcomes by Sex	Secondary_Outcomes_female_20240712.pdf
Figure A4	CATE on Overconfidence (Midline/Endline × Gender)	Mid_CATE_Men_d_s38_gnew_score_20240712.pdf; Mid_CATE_Women_d_s38_gnew_score_20240712.pdf; End_CATE_Men_d_s38_gnew_score_20240712.pdf; End_CATE_Women_d_s38_gnew_score_20240712.pdf
Figure A5	Effects on Disaggregated SES Skills	Self_Awareness_20240712.pdf; Self_Management_20240712.pdf; Social_Awareness_20240712.pdf; Relationship_Management_20240712.pdf

How to reproduce the results

1. Download or clone the reproducibility package.
2. Update the project path in the master do-file and to match the location on your computer.
3. Open Stata and run the master do-file.

No other scripts need to be run manually.