README – Replication Files for *"Conscious Coupling: Impact of a Skills Intervention to Address Intrahousehold Constraints for Women Entrepreneurs in Ethiopia"*

Last updated: June 2025

This file describes the information and procedures needed to replicate the results of the working paper "Conscious Coupling: Impact of a Skills Intervention to Address Intrahousehold Constraints for Women Entrepreneurs in Ethiopia", by Sophia Friedson-Ridenour, Adiam Hagos Hailemicheal, Dumebi Ochem and Sreelakshmi Papineni, published in the *World Bank Policy Research Working Paper Series*.

Computational Requirements

- Computer software: STATA
    - Stata version 12.1 or higher required
    - Code was last tested on a computer with version 18.5 installed
- Required user-written commands:
    - ietoolkit
    - outreg2
    - estout
    - swindex
    - winsor
    - ftools
    - elabel
- The replicator should expect the code to run for about 5-7 minutes.
- System specifications: the code was last run on an Intel® Core™ i7-1065G7 CPU @ 1.30GHz 1.50 GHz, 12.0 GB RAM, 64-bit operating system on Windows 11 Home.

Instructions for Replicators

All main tables were created using LaTeX editor and the replicator can run all cleaning, construction and analysis scripts sequentially from start to end without interruption using *0_MasterDofile.do*. The replicator will first need to define a personal directory path for "rep_path" in the master do-file before running any program (line 72, insert in replacement of string "YOUR PATH").

If necessary, the replicator can install the user-written packages used in the analysis by changing the flag from 0 to 1 in the master do-file (line 86). This will install all dependencies locally, if not already installed. This is a one-time set-up per computer.

Should one wish to run do-files individually, it is still necessary to always run lines 46 to 78 in *0_MasterDofile.do* at the beginning of any session, in order to set up all paths globals. It is also important to note that do-files 1 to 4 must be run in order, while do-files 5 to 7 may be run in any order.

The replicator may equally change the flags defining the do-file groups for the cleaning stage and that for the analysis stage (lines 108 and 109 in do master do-file).

Datasets Descriptions

All datasets provided are primary data from individual surveys collected by the authors for the purposes of a randomized intervention evaluating the impact of *ScaleUp!*, a skills-oriented business training in Ethiopia. Data was collected over screening, baseline and follow-up stages using an implementing partner. Surveys were conducted with a primary female respondent in the household; during the follow-up phase, the male spouses were also interviewed separately. The unique identifier for all datasets is an individual respondent. All raw datasets have been de-identified for public use, and can be accessed by any external users. Data sources are all in *.dta* format.

- Raw datasets at screening stage (unique identifier is variable *dot_1_1*):
  - 210328_WB_Couples_Training_Phone_Screening_Final_public.dta
    - Contains data collected at the screening stage.
    - Relevant variables include whether the respondent is interested in the ScaleUp! training and whether she consents for her husband to participate in the training.
    - Date acquired: 28 March 2021
  - 210728_WB_Couples_Training_Phone_Screening_Final_public.dta
    - Contains data collected at the first follow-up of the screening stage.
    - Relevant variables are as above.
    - Date acquired: 28 July 2021
  - 211206_WB_Couples_Training_Phone_Screening_round2_public.dta
    - Contains data collected at the second follow-up of the screening stage.
    - Date acquired: 6 December 2021
  - 220530_WB_Couples_Training_Phone_Screening_round3_public.dta
    - Contains data collected at the third follow-up of the screening stage.
    - Date acquired: 30 May 2022
- Raw datasets at baseline stage (unique identifier is variable *meta_1_1*):
  - 210927_WB_Couples_Training_BL_Clean_public.dta
    - Contains data collected by the authors during the first round of the baseline stage for all female survey respondents, irrespective of whether they are included in the intervention.
    - Date acquired: 27 September 2021
  - 220204_WB_Couples_Training_BL_Clean_r2_public.dta
    - Contains data collected during the second baseline round for all female survey respondents.
    - Date acquired: 4 February 2022
  - DOT_Couples_Training_strat_randomization.dta
    - Contains the treatment allocation and strata for each respondent that was randomized for the intervention in the first survey round.

- Also contains all the survey data as in the baseline datasets above, but only the treatment status and randomization strata are extracted for analysis.
- Date acquired: 5 October 2021
    - DOT_Couples_Training_strat_randomization_R2BL.dta
        - Contains the treatment allocation and strata for each respondent that was randomized for the intervention in the second survey round.
        - Also contains all the survey data as explained above.
        - Date acquired: 5 January 2022
    - WB_Couples_Training_Implementation_Monitoring.dta
        - Contains data collected by the author's implementation partner tracking the respondents' attendance to the ScaleUp! training.
        - Date acquired: 23 December 2022
- Raw datasets at endline stage (unique identifier is variable *meta_1_1*):
    - WB_Couples_Training_EL_main_public.dta
        - Contains data collected by the authors during the endline stage for all tracked female survey respondents, irrespective of whether they participated in the training.
        - Date acquired: 31 October 2023
    - WB_Couples_Training_EL_spouse_public.dta
        - Contains data collected at the endline stage for all tracked male spouses, irrespective of whether the female respondents participated in the training.
        - Date acquired: 31 October 2023
- Intermediate datasets:
    - Screening_full_nodup.dta
        - Contains the data from all four survey datasets devoid of any duplicates, ready to be merged to the final datasets.
    - Implementation Monitoring_Consolidated_Report.dta
        - Contains the data from the monitoring dataset devoid of any duplicates, ready to be merged to the final datasets.
    - DOT_Baseline_data.dta
        - Contains all of the constructed outcomes from the baseline survey data.
        - Relevant outcomes include business performance, socioemotional skills, as well as intimate relationship with one's spouse.
    - DOT_treatment_status.dta
        - Contains the respondents' treatment status and randomization strata extracted from *DOT_Baseline_data*, to be merged to *DOT_Endline_data* for the construction of composite indices.
    - DOT_Endline_data.dta
        - Contains all of the constructed outcomes from the endline survey data.
        - Relevant outcomes as above.
    - DOT_data_merged.dta

- Contains all of the raw survey data as well as the constructed outcomes, merging the baseline and endline datasets in wide format by the respondent ID.
- Final datasets:
  - IFWE_DOT_Data.dta
    - This is the final dataset that is used in the analysis, containing only the constructed and labelled outcomes at both baseline and endline.

Disclaimer: During the data collection process at the endline stage, the authors identified 14 respondents that had duplicate entries in the baseline surveys, and thus needed to be removed from the analysis. This led to a drop in the total number of observations from 987 to 973. The datasets reporting these issues are listed below:
- 230811_WBC_initialduplicatescheck.dta
  - Contains the list of respondents that were identified to have duplicate entries in the baseline surveys. This dataset was compiled and shared by the authors' monitoring partner.
- DOT_duplicates_check.dta
  - Contains a revised list of the 14 duplicate respondents, which was worked on manually by the authors to identify which of the duplicates should be removed from the analysis (specifically the variable *Keepdrop*). The code systematically keeps the respondents' first interview in the analysis.

Additionally, in the final baseline dataset *DOT_Baseline_data.dta*, there are three respondents that were randomized for the intervention but had all critical survey entries empty, due to their business being temporarily closed at the baseline stage. These observations were removed from the final analysis, leading to a drop in observations from 973 to 970.

Do-File Structures

- 0_MasterDofile.do
  - Calls all of the do-files in the directory in sequence to construct outcomes and produce all tables in the paper. It will run the entire replication from start to finish, install all packages, and set up all user defined globals and programs needed during the cleaning and the output stages.
- 1_DuplicatesCleaningDofile.do
  - Performs and initial cleaning of the screening and training monitoring datasets, by identifying duplicate or incorrect respondent ID codes.
- 2_BaselineConstructionDofile.do
  - Constructs all the main variables from data obtained from the baseline and screening surveys.
- 3_EndlineConstructionDofile.do
  - Constructs all the main variables from data obtained from the endline and spousal surveys.

- 4_FinalCleaningDofile.do
  - Assembles the intermediary baseline and endline datasets, labels all outcomes and prepares the final dataset for analysis.
- 5_DescriptivesDofile.do
  - Produces Tables 1 and 2 in the main paper, and Tables A1 and A2 in the appendix (outputs listed in table below).
  - Please note that, unlike all other tables, Table A1 was compiled manually in the LaTeX editor using data inputs from *IFWE_DOT_Data.dta* for easier formatting, rather than using the do-file to produce a *.tex* file. This code exports an Excel file displaying all inputs exactly as they appear in the appendix table, but without any of the formatting.
- 6_MainAnalysisDofile.do
  - Produces Tables 3 to 10 in the main paper (outputs listed in table below). It also extracts p-values from all main regressions and calculates sharpened two-stage q-values (reported in Tables 3 and 4 only).
  - This code produces a *.tex* file version of each table that can be pasted directly into a LaTeX editor. Some manual formatting was carried out only LaTeX in order to ensure that the regressions' column titles are wrapped and the table fits within a portrait-oriented page. The code also produces a .csv file of the tables for easier viewing, but does not include the necessary formatting or table notes.
- 7_AppendixAnalysisDofile.do
  - Produces Tables A3 to A9 in the appendix (outputs listed in table below).
  - This code produces a *.tex* file version of each table that can be pasted directly into a LaTeX editor. Some manual formatting was carried out only LaTeX in order to ensure that the regressions' column titles are wrapped and the table fits within a portrait-oriented page. The code also produces a .csv file of the tables for easier viewing, but does not include the necessary formatting or table notes.
- AnonymizationDofile.do (not provided here)
  - De-identifies all raw datasets that are provided and used in the analysis. This script cannot be run in this replication package due to the presence of personally identifiable information (PII), but it is listed here for transparency.
  - Some of the steps involved in the de-identification process include:
    - Generating encoded versions of respondent's names and phone numbers, which are used in the cleaning stage for the identification of duplicate ID codes.
    - Removal of variables containing respondents' names, phone numbers, spouses' names (or other contacts), enumerators' names, business names, precise locations of businesses, GSP coordinates of interview location, etc.

List of Tables and Programs

The provided code reproduces all tables and figures, both in the main body of the paper and in the appendix.

Table	Program	Lines	Output file(s)
Table1	5_DescriptivesDoFile.do	87-104	Tab1_balance.tex, Tab1_balance.xlsx
Table2	5_DescriptivesDoFile.do	107-124	Tab2_profits.tex, Tab2_profits.xlsx
Table3	6_MainAnalysisDoFile.do	366-414	Tab3_business.tex, Tab3_business.csv
Table4	6_MainAnalysisDoFile.do	417-465	Tab4_invest.tex, Tab4_invest.csv
Table5	6_MainAnalysisDoFile.do	468-516	Tab5_buspract.tex, Tab5_buspract.csv
Table6	6_MainAnalysisDoFile.do	519-567	Tab6_finance.tex, Tab6_finance.csv
Table7	6_MainAnalysisDoFile.do	570-624	Tab7_ses.tex, Tab7_ses.csv
Table8	6_MainAnalysisDoFile.do	628-676	Tab8_decisions.tex, Tab8_decisions.csv
Table9	6_MainAnalysisDoFile.do	679-727	Tab9_marriage.tex, Tab9_marriage.csv
Table10	6_MainAnalysisDoFile.do	730-778	Tab10_time.tex, Tab10_time.csv
TableA1	5_DescriptivesDoFile.do	127-223	TabA1_takeup.xlsx
TableA2	5_DescriptivesDoFile.do	226-249	TabA2_takeup_corr.tex, TabA2_takeup_corr.xlsx
TableA3	7_AppendixAnalysisDoFile.do	349-389	TabA3_attrition.tex, TabA3_attrition.csv
TableA4	7_AppendixAnalysisDoFile.do	392-434	TabA4_husband.tex, TabA4_husband.csv
TableA5	7_AppendixAnalysisDoFile.do	437-479	TabA5_business.tex, TabA5_business.csv

Table	Program	Lines	Output file(s)
TableA6	7_AppendixAnalysisDoFile.do	482-530	TabA6_ses.tex, TabA6_ses.csv
TableA7	7_AppendixAnalysisDoFile.do	533-582	TabA7_norms.tex, TabA7_norms.csv
TableA8	7_AppendixAnalysisDoFile.do	585-628	TabA8_business_late.tex, TabA8_business_late.csv
TableA9	7_AppendixAnalysisDoFile.do	631-675	TabA9_finance_late.tex, TabA9_finance_late.csv

Folder Directory

Please note that the code in this replication package assumes that no changes to the folder structure have been made. In order to ensure that the scripts run smoothly, it is advised to follow the folder structure as follows:

- "Data" folder
  - o "Raw" folder – all the original datasets (anonymized versions) used in the construction of the final dataset are stored here. This folder must contain all of the following datasets:
    - WB_Couples_Training_Implementation_Monitoring.dta
    - 210328_WB_Couples_Training_Phone_Screening_Final_public.dta
    - 210728_WB_Couples_Training_Phone_Screening_Final_public.dta
    - 210927_WB_Couples_Training_BL_Clean_public.dta
    - 211206_WB_Couples_Training_Phone_Screening_round2_public.dta
    - 220204_WB_Couples_Training_BL_Clean_r2_public.dta
    - 220530_WB_Couples_Training_Phone_Screening_round3_public.dta
    - 230811_WBC_initialduplicatescheck.dta
    - DOT_Couples_Training_strat_randomization.dta
    - DOT_Couples_Training_strat_randomization_R2BL.dta
    - DOT_duplicates_check.dta
    - WB_Couples_Training_EL_main_public.dta
    - WB_Couples_Training_EL_spouse_public.dta
  - o "Intermediate" folder – all modified datasets used in the cleaning and construction stage will be stored here (produced by the code).
  - o "IFWE_DOT_Data.dta" file: final dataset ready for analysis is stored here in the Data folder
- "Do Files" folder
  - o Must contain all 8 scripts listed above (0_MasterDofile.do runs all of them in sequence)
- "Outputs" folder

- o "Main Tables" folder – all tables for the main section of the paper will stored here, both in *.tex* format for LaTeX, and *.csv* format for easier viewing.
  - o "Appendix Tables" folder – all tables that appear in the Appendix section of the paper will stored here, both in *.tex* format for LaTeX, and *.csv* format for easier viewing (with the exception of Table A1 that was produced manually in LaTeX using data inputs).
- "Ado Files" folder
  - o Contains all user-written commands and packages required by Stata to run the scripts and produce all outputs. This folder should not be modified.

Data Availability Statements

Statement about Rights

The data from an impact evaluation contained in this replication package was collected by the authors using a survey firm. I certify that the authors have legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

The authors have documented permission to redistribute and publish the data contained within this replication package. All sensitive information that cannot be made public has been removed without jeopardizing the reliability and reproducibility of the results. The datasets will be published in the public World Bank Microdata library once the paper has been accepted for publication in a journal.

Contact Information

For questions and/or comments, please contact the corresponding author for the paper Sreelakshmi Papineni (spapineni@worldbank.org) and/or the creator of this data replication package, Dumebi Ochem (dochem@worldbank.org).