

Closing the Gaps: The Role of Screening Questions and Self-Reporting in Measuring Women's and Young People's Employment and Work

Ivette Contreras¹

Lelys Dinarte-Diaz²

Amparo Palacios-Lopez³

Valentina Costa⁴

Steffanny Romero⁵

¹Corresponding Author. The World Bank. Email: icontreras@worldbank.org.

² The World Bank, IZA, CESifo, HiCN. Email: ldinartediaz@worldbank.org.

³ The World Bank. Email: apalacioslopez@worldbank.org.

⁴ The World Bank. Email: vcosta@worldbank.org.

⁵ The World Bank. Email: sromeroesteban@worldbank.org.

Reproducibility Package README

Overview

The code within this replication package executes the analysis detailed in the paper "Closing the Gaps: The Role of Screening Questions and Self-Reporting in Measuring Women's and Young People's Employment and Work" using Stata. This package includes all data files and codes required to go from the anonymized data to the replication of the tables and figures included in the analysis. The master file manages the execution of all codes, anticipated to run for approximately 5 hours. Our primary objective in this reproducibility package is to ensure both accessibility and credibility in replicating our research methodology and results.

Data Availability and Provenance Statements

Labor Measurement for Women and Youth: Randomized Survey Experiment in El Salvador

To collect the relevant data for our experiment, we surveyed the sample of 1,008 households (2,480 household members) between August and October 2022. The interviews were conducted using a Computer-Assisted Personal Interviewing (CAPI) program, in which we embedded the household assignments to the treatment groups. The experiment was conducted in two regions within the Departments of San Salvador and Usulután. We selected these regions to achieve high variation in the types of activities reported. Region 1, which is located in the Department of San Salvador, includes households located close to the Metropolitan Area of San Salvador, the main economic activities of which are commerce, construction, manufacturing, and sugar cane production. Region 2 is situated in the Department of Usulután, where the main economic activities include coffee production, fishing, and commerce.

Using the 2007 Salvadoran Population Census (i.e., the most recent census available), we identified 276 enumeration areas (EAs): 114 in Region 1, and 162 in Region 2. Each of these EAs had at least 30 households and was classified as either rural or peri-urban. From the 276 EAs, we randomly selected 48 that were evenly distributed across the rural and peri-urban areas and then conducted a full household listing for each EA. We then randomly selected 21 households within each EA for a total of 1,008 households to constitute our study sample. Within each household, we randomly chose a minimum of two and up to a maximum of four working-age household members aged between 15 to 64 years, which yielded a total of 2,480 working-age individuals. Since our main hypothesis consists of the differential impacts of these survey methods by age and sex, we stratified the selection of the working age household members by these two variables, and we defined youth as individuals aged between 15 to 24 years.

We conducted a stratified randomization at the EA-level to assign the 1,008 households with equal probability to one of the following treatment groups:

Group	Type of labor module	Type of respondent
Control group: 336 households	Standard labor module (following 19 th , ICLS)	Proxy responses accepted for other HH members if they were not available. If available each HH member (15-64 years old) provides his/her own information
T1: 336 households	Standard + LOA module	Proxy responses accepted
T2: 336 households	Standard labor module (following 19 th , ICLS)	Self-responses enforced for the eligible household members (15-64 years old).

The survey instrument included up to 18 modules (see [LINK](#) to access the questionnaire). The first module collected information from the household roster to identify all household members. After completing the household roster, the CAPI program randomly selected the working-age household members (aged 15 to 64 years) who would participate in the experiment. Each randomly selected member (or the member's proxy respondent for T1 or C) provided information solicited via the three main modules: education, access to technologies, and labor. Participants in T1 also responded the LOA module. We made sure that the three relevant modules were collected at the beginning of the experiment and from all of the randomly selected self- or proxy respondents in the three treatment groups.

Encuesta de Hogares de Propósitos Múltiples 2022 (EHPM) - El Salvador

The reproducibility package also uses data from the Encuesta de Hogares de Propósitos Múltiples (EHPM) from El Salvador. This data is public but cannot be redistributed in the reproducibility package. We included instructions for downloading the data in the do-file 02_ehpm.do.

Statement about Rights

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

- All data **are** publicly available.
- Some data **cannot be made** publicly available.
- **No data can be made** publicly available.

The data for this project is owned by the World Bank (it came from a survey data procured by the World Bank under the project P168212).

Getting Started

Outline and Folder Structure

The reproducibility package consists of three folders:

- *Data anonymized*: this folder should contain all the raw data files, the intermediate files, and the final data set used for the analysis.
- *Do files*: this folder contains the related codes. The Master File labeled as “**master-dofile.do**” orchestrates the sequential execution of all code files according to the specified order.
- *Output*: this folder contains all the tables and figures reported in the paper in raw form. Output files are called appropriate names (e.g., Table1.xls, Figure1.png) and should be easy to correlate with the manuscript.

How to Reproduce

The initial action within this reproducibility package involves accessing the master file named `master-dofile.do` and execute its contents. Before proceeding, ensure Stata is installed on your system. This file establishes directory paths and executes the necessary Stata code to generate analytical outputs. Here's a concise overview of its contents:

PART 1: Install user-written packages and harmonize settings. This part of the code is deactivated, all package files are included in the reproducibility package and there is no need to download them.

PART 2: Prepare folder paths and define programs.

PART 3: Run do-files.

PART 3.1: This part runs `01_variables_creation.do` do-file that constructs variables and indicators used in the analysis.

PART 3.2: This part runs `02_ehpm.do` do-file that creates a dataset using EHPM for comparison purposes. Note that the code requires to download of the dataset EHPM 2022.sav, follow the instructions at the beginning of the code.

PART 3.3: This part runs `03_main_tables_figures.do` do-file that will generate all the main tables and figures.

PART 3.4: This part runs `04_appendix_tables.do` will generate the appendix tables.

Kindly designate the user root-path in line 30 within the Master do file (`master-dofile.do`). You have the option to execute the entire do file or select specific sections. It's crucial, however, to run the do file in sequence. If opting to run only one section, ensure that preceding sections have been executed beforehand.

Dataset list

The table below lists all datasets in the raw data file required for running the do-files.

Datafile	Note	Provided
all_sections.dta	Contains individual-level information from all survey modules	Yes
EHPM 2022.sav	The 2022 Household and Multipurpose Survey (EHPM) provided the data for individuals in El Salvador that we compared to data that we measured. Follow the instructions on <code>02_ehpm.do</code> file to download it.	No

List of tables and programs

Selected tables and figures in the paper, as explained and justified below.

Table/Figure	Program	Line Number	Output file
Table 1	Do files/ <code>03_main_tables_figures.do</code>	33	Table1.xlsx
Table 2	Do files/ <code>03_main_tables_figures.do</code>	33	Table2.xlsx
Table 3	Do files/ <code>03_main_tables_figures.do</code>	103	Table3.xlsx
Table 4	Do files/ <code>03_main_tables_figures.do</code>	136	Table4.xlsx
Table 5	Do files/ <code>03_main_tables_figures.do</code>	221	Table5.xlsx
Table 6	Do files/ <code>03_main_tables_figures.do</code>	305	Table6.xlsx

Table 7	Do files/ 03_main_tables_figures.do	402	Table7.xlsx
Figure 2	Do files/ 03_main_tables_figures.do	493	Figure2.png
Figure 3	Do files/ 03_main_tables_figures.do	568	Figure3.png
Table A3	Do files/ 04_appendix_tables.do	29	TableA3.xlsx
Table A4	Do files/ 04_appendix_tables.do	70	TableA4.xlsx
Table A6	Do files/ 04_appendix_tables.do	87	TableA6.xlsx
Table A7	Do files/ 04_appendix_tables.do	117	TableA7.xlsx
Table A8	Do files/ 04_appendix_tables.do	207	TableA8.xlsx
Table A9	Do files/ 04_appendix_tables.do	361	TableA9.xlsx
Table A10	Do files/ 04_appendix_tables.do	516	TableA10.xlsx
Table A11	Do files/ 04_appendix_tables.do	671	TableA11.xlsx
Table A12	Do files/ 04_appendix_tables.do	699	TableA12.xlsx
Table A13	Do files/ 04_appendix_tables.do	780	TableA13.xlsx
Table A14	Do files/ 04_appendix_tables.do	861	TableA14.xlsx
Table A15	Do files/ 04_appendix_tables.do	893	TableA15.xlsx
Table A16	Do files/ 04_appendix_tables.do	978	TableA16.xlsx
Table A17	Do files/ 04_appendix_tables.do	1064	TableA17.xlsx
Table A18	Do files/ 04_appendix_tables.do	1101	TableA18.xlsx

Computational requirements

Software Requirements

Stata 18 should be used to run the codes. Below are the versions of each command used.

c:\ado\plus\r\ritest.ado
*! version 1.21 mar2024.
c:\ado\plus\r\reghdfe.ado
*! version 6.12.3 08aug2023
c:\ado\plus\o\outreg2.ado
*! outreg2 2.3.2 17aug2014 by roywada@hotmail.com
*! based on outreg 3.0.6/4.0.0 by john_gallup@alum.swarthmore.edu
c:\ado\plus\e\ereplace.ado
*! Chris Larkin, version 1.0.3
*! Chris Larkin, version 1.0.2
*! Nick Cox, version 1.0.1
c:\ado\plus\b\balancetable.ado
*! version 5.0 19nov2018 Mattia Chiapello
c:\ado\plus\s\swindex.ado
c:\ado\plus\c\coefplot.ado
*! version 1.8.6 22feb2023 Ben Jann
c:\ado\plus\f\ftools.ado
*! version 2.49.1 08aug2023

References

Lars Vilhuber, Connolly, M., Koren, M., Llull, J., & Morrow, P. (2022). A template README for social science replication packages (v1.1). Zenodo. <https://doi.org/10.5281/zenodo.7293838>

Acknowledgment

We acknowledge the support of the International Fund for Agricultural Development and the World Bank Research Support Budget. The Institutional Review Board (IRB) of Universidad Francisco Gavidia in El Salvador reviewed and approved this research project's protocol, which is registered under ID #011-2021.