

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

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
LOA: 336 households	Standard + LOA module	Proxy responses accepted
ESR: 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.

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.

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 contains all the raw data files.
- *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.

PART 2: Prepare folder paths and define programs.

PART 3: Run do-files.

PART 3.1: 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 3.2: This part runs `03_main_tables_figures.do` do-file that will generate all the main tables and figures. Note that the code requires downloading the dataset *master_data.dta* follow the instructions at the next section.

PART 3.3: 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
master_data.dta	Contains individual-level information from all	No

	survey modules. Follow the instructions to download it.	
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 02_ehpm.do file to download it.	No
visit_id	Contains visit-level information	Yes

Obtaining the microdata

Dataset: Methodological Innovations for Measuring Informality – A Survey Experiment on Labor Measurement 2022 (SELM 2022)

Reference ID: SLV_2022_SELM_v01_M

Source: World Bank Microdata Library, LSMS collection

Catalog page: <https://microdata.worldbank.org/catalog/8575>

1. **Create an account.** Go to <https://microdata.worldbank.org/auth/register> and sign up for a free account. Confirm your email address using the link the Microdata Library sends you. Skip this step if you already have an account.
2. **Log in.** Click the "Get Microdata" tab. The site will ask you to log in first. Use your email and password, and you will be sent back to the download page.
3. **Accept the terms of use.** Before you can download, you must agree to the access conditions.
4. **Download the files.** After you accept, the page shows the "Data Files" section with three options: CSV, SPSS, and Stata 14. Click "Download" next to "Data in Stata 14 format" (ZIP, 1.43 MB). This downloads the file SLV_2022_SELM_v01_M_STATA14.zip. Open the zip and extract **master_data.dta** into the subfolder \Reproducibility package\Data anonymized\Clean data. Do not rename or edit the file, since the code reads it from this location with its original name.

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/ 03_main_tables_figures.do	29	Table1.xlsx
Table 2	Do files/ 03_main_tables_figures.do	104	Table2.xlsx
Table 3	Do files/ 03_main_tables_figures.do	226	Table3.xlsx
Table 4	Do files/ 03_main_tables_figures.do	316	Table4.xlsx
Table 5	Do files/ 03_main_tables_figures.do	412	Table5.xlsx
Table S3.1.	Do files/ 04_appendix_tables.do	29	TableS3_1.xlsx

Table S3.2.	Do files/ 04_appendix_tables.do	49	TableS3_2.xlsx
Table S3.3.	Do files/ 04_appendix_tables.do	71	TableS3_3.xlsx
Table S3.4.	Do files/ 04_appendix_tables.do	124	TableS3_4.xlsx
Table S3.5.	Do files/ 04_appendix_tables.do	187	TableS3_5.xlsx
Table S3.6.	Do files/ 04_appendix_tables.do	240	TableS3_6.xlsx
Table S4.1.	Do files/ 04_appendix_tables.do	277	TableS4_1.xlsx

Computational requirements

Software Requirements

Stata 18 should be used to run the codes. Below are the versions of each command used. To ensure you are using the same version of the command, please copy the ritest ado file provided in the reproducibility package's ado files folder to your personal Stata ado path.

```
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.