



Basic Information Document

**Methodological Innovations for Measuring Informality –
A Survey Experiment on Labor Measurement in El Salvador**

September 2026

Table of Contents

1.	Introduction	3
2.	Sample and Survey Instrument.....	4
2.1.	Survey Sample	4
2.2.	The instrument	4
2.3.	Description of Datasets	6
3.	Organization of the survey	6
3.1.	Focus Groups	6
3.1.1.	Focus Group Discussions	6
3.1.2.	Approach.....	7
3.1.3.	Focus Group Questions	7
3.2.	Training of Interviewers for the Survey	8
3.3.	Composition of field teams	9
3.3.1.	Supervisors	9
3.3.2.	Enumerators	9
3.4.	Data Monitoring and Evaluation	9
4.	Data Management.....	10
4.1.	Computer Assisted Personal Interview (CAPI)	10
4.2.	Data Communication System	10
4.3.	Data Cleaning	11
4.4.	Confidential Information	11
5.	Using the Data	11
5.1.	File Structure	11
5.2.	Merging Datasets	12

1. Introduction

The purpose of this document is to provide detailed information on the study “Methodological Innovations for Measuring Informality – A Survey Experiment on Labor Measurement in El Salvador”. The general objective of this methodological experiment is to understand how to better measure the informal work of women and youths 15-24 years old in rural and peri-urban areas in two regions of El Salvador (departments of Usulután and San Salvador).

From June to September 2022, the World Bank’s Living Standard Measurement Study (LSMS) has successfully designed and implemented this methodological experiment with the support from FUSADES a Salvadoran firm that was chosen after a competitive process. This dataset includes information of 1,008 households in 48 enumeration areas within two rural or peri-urban regions in El Salvador (in the provinces of San Salvador and Usulután). Within each EA, 21 households were selected at random following a full household listing and in each household a minimum of one male and one female working-age adult household member was selected at random. To avoid contamination across the groups, all households in a particular EA were assigned to only one treatment status, as described in Table 1:

Table 1. Treatment description

Group	Type of labor module	Type of respondent
Control group 336 households	Standard (LSMS labor module following 19 th , ICLS)	Proxy respondent: Main household respondent or if another household member (15-64 years old) is present, he/she provides his/her own information
Group 1: 336 households	New labor module that includes a new list of work activities	Proxy respondent: Main household respondent or if another household member (15-64 years old) is present, he/she provides his/her own information
Group 2: 336 households	Standard (LSMS labor module following 19 th , ICLS)	Only self-respondents: Selected household members (15-64 years old) must provide their own information

Specifically, members of the households assigned to T1 were exposed to an additional survey module (“List of Activities” Module) before collecting their labor information. This additional module consists of a list of the most common labor activities that people in their region do. The list aims to provide examples of different income-generating activities that can be considered as work for statistical purposes. This list has been designed using results from 7 focus groups implemented in rural and peri-urban communities in El Salvador during July 2022. Household members assigned to the T2 should provide their labor information by themselves (as opposed of using a proxy respondent as it is common in household surveys). Households in the C group

were not exposed to the additional survey module nor to the self-respondent option. All respondents completed similar modules on demographic and socio-economic data at the individual- and household-level. The data will be useful in exploring the heterogeneity of survey treatment effects.

The design and implementation of this experiment includes three steps. First, the experiment identified relevant elements for the employment decision in Salvadoran urban and peri urban communities. Focus group discussions helped to understand participants' perceptions of the main barriers to work in the formal sector, such as requirements of work experience, childcare services, availability of formal contract, security, among others. Secondly, using results from these focus group discussions, the experiment identified the most important job attributes and values for non-urban settings (5 for men and 6 for women).

2. Sample and Survey Instrument

2.1. Survey Sample

The survey was conducted in two regions within the Departments of San Salvador and Usulután. We selected these regions in order 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 between the ages of 15 to 64 years, which yielded a total of 2,480 working-age individuals.

2.2. The instrument

The survey was administered in August-October 2022. A total of 18 interviewers and 6 supervisors conducted the survey.

The survey consists of one questionnaire. The **Household Questionnaire** was administered to all households in the sample.

Household Questionnaire: The Household Questionnaire provides information on demographics; access to education and technology; employment and non-farm enterprise; housing; time-use; social norms; social desirability; discrete choice experiment (DCE); and risk aversion.

The contents of questionnaire are outlined below.

Table 2: Questionnaire

Section	Topic	Description
C	Cover	Household identifiers and interviewer identifiers
S	Interview Information and Phone Number Roster	Roster of call attempts, result and respondent of call attempt, interview consent, date and time of call back, roster of phone numbers, the information of the person that the listed phone number belongs to
1	Household Roster	Roster of members of the household, relationship to the household head, gender, age, marital status
2	Education	Level of education, education history, why currently not studying, training program
3	Technology	Access to internet, mobile ownership
4T	List of activities	Region-specific examples of activities
4	Labor	Status in employment, why currently not working, job search, actual job
4A	Skills and readiness to work	Ability to read, write and speak in English, frequency, to write short lists, notes or instructions, government forms, application forms
5	Aspirations	Work history (i.e., age when started working, work experience gained previously, sources of income), future plans (i.e., intended activities after completing education/training program) and aspirations (i.e., desired work, role models, social norms, migration interest)
6C	Time use	Record of the activities for the individual in the last complete 24 hours (starting yesterday morning at 4 am, finishing 3:59 am of the current day)
7	Social norms	Personal behaviour, personal normative beliefs,

Section	Topic	Description
		Social empirical expectations (behaviour) and Social normative expectations
8	Attitudes	Personal attitudes for assessing social desirability bias
9A	Discrete choice experiment	Labor preferences
9B	Risk aversion	Risk aversion attitudes
11	Housing	Conditions of the house, i.e., construction materials of the roof, walls, floor, type of toilets and garbage disposal
5	Non-Farm Enterprise	Family business/non-farm enterprise, sector of family business, challenges faced by the business
12	Food security	Food security status of the household during the last 30 days
12C	Agriculture	Agriculture activities of the household, including crops, livestock and fisheries (capture and aquaculture)
14	Household assets	Ownership of assets and value
6	Other income in the last 12 months	Miscellaneous income received by the household
16	Contact	Contact details

2.3. Description of Datasets

The dataset is presented at the individual level including information for all the modules detailed in Table 2.

3. Organization of the survey

3.1. Focus Groups

3.1.1. Focus Group Discussions

The main goal of this qualitative study is twofold: (1) to create a list of activities considered work in rural and peri-urban communities in El Salvador, and (2) to understand the barriers that women and young people face in entering the labor market and obtaining formal employment. We conducted eight focus groups between June 13 and July 8, 2022 in six Salvadoran municipalities located in the Department of San Salvador (Nejapa and Guazapa) and the Department of Usulután (Jiquilisco, Puerto El Triunfo, San Dionisio, and California). The focus

groups were stratified by age and gender. Four of the focus groups included youth between the ages of 18 and 29 years (two with men and two with women), and the four other focus groups included adults aged 29 years and older (two with women and two mixed). A total of 58 people attended the focus groups: 40 women and 18 men within their respective municipalities.

Table 4. Focus Group Participants by Gender and Municipality

Municipality	Men	Women	Total
Nejapa (young men)	7	0	7
Nejapa (adults)	2	7	9
Guazapa (young women)	0	8	8
Guazapa (adult women)	0	8	8
Jiquilisco (young men)	6	0	6
Puerto El Triunfo (young women)	0	8	8
San Dionisio (mixed adults)	3	4	7
California (adult women)	0	5	5
Total	18	40	58

3.1.2. Approach

The focus groups facilitators were local qualitative experts and trained by the research team using a technical guide that included all of the focus group questions. During the semi-structured focus group meetings, the facilitator asked the participants open-ended questions that could lead to additional questions and responses.

Special care was taken to preserve participant anonymity and freedom to consent. Before starting each of the focus group sessions, a team member talked with the potential participant and read a consent statement that included the objective of the focus group, how participation was voluntary, and how the information shared would be used based on the IRB's approval of the experiment design. A facilitator conducted each of the discussions with the support of two team members who were responsible for listing the main work activities that the participants identified during the discussion on a whiteboard. The discussions lasted up to one and a half hours.

3.1.3. Focus Group Questions

The semi-structured technical guide helped the facilitators lead the focus group discussions. The guide included two main components. The first component included questions to elicit a list of general activities in which the community members engage. From there, the group identified which activities are considered work and which employment. The discussions started with the following question, *“Think about a typical day in your life. Which activities do you perform?”*,

followed by *“Now, I would like us to talk about things that members of the community generally do. Could you describe to the group what some of the most common activities are within your community?”* After identifying all the common activities, the facilitator asked: *“Which of these activities are considered work?”* Using different colored markers, the participants highlighted the activities that could be classified as work. After identifying the work activities, the facilitator asked about the heterogeneity of the activities through the following questions: *“Do both men and women perform these activities? Or do the men and women in your community do different things?”*, and *“Do young people also perform these activities?”*

The second component of the focus group aimed to identify the participants’ main barriers to finding stable employment within the community. In this section of the discussion, the facilitator started by asking whether people in the community have difficulty finding stable jobs. All of the focus groups participants agreed that it was difficult to find stable employment. Afterwards, the facilitator asked: *“What are the main obstacles to fostering stable employment within your community?”* The facilitators also explored heterogeneity in the job barriers by asking the following: *“Do women face different obstacles? What job characteristics would entice women to work or enable them to work?”* and *“Do young people face different barriers than women and men? What job characteristics would entice young people to work or enable them to work?”*

3.2. Training of Interviewers for the Survey

Personnel were selected from the pool of FUSADES interviewers that have experience with the Survey Solutions platform and have previously conducted phone surveys. Prior to the start of every round of the survey, a group of main interviewers, callback interviewers, and supervisors were trained to conduct the survey. The trainings lasted 2-3 days and focused on providing participants with detailed understanding of the survey and the Computer Assisted Personal Interview (CAPI) techniques, elucidation of important concepts and questions in the questionnaire.

The trainings were led by FUSADES with additional inputs provided by the LSMS team. Following the training, the interviewers carried out mock interviews in order to become more familiar with the questionnaire and interviewing techniques before interviewing respondents. After the mock interviews, a 1-day pilot was conducted with households in another community. Each interviewer conducted around 4 interviews during the pilot. The survey materials (presentations, manual, questionnaire and Survey Solutions assignments) were sent to trainees prior to the start of the trainings. This was to ensure that they become familiar with the materials and also have a smooth training.

3.3. Composition of field teams

Seven teams were deployed to the field for data collection. Each team consisted of eight professionals—field supervisor (team-leader), mappers/listers (2) and enumerators (5). These teams were supported in their day-to-day operations by several technical staff at FUSADES headquarters in San Salvador.

3.3.1. Supervisors

Each team had one supervisor who travelled with the team at all times. The team supervisor is the overall coordinator and supervisor for the fieldwork, as well as the final responsible for ensuring the team is completing all duties in a timely and high-quality manner. If the enumerators have any problems or need assistance, they report to the team supervisor. The team supervisor may also conduct household interviews when necessary in order to keep fieldwork on schedule.

3.3.2. Enumerators

The enumerators are responsible for accurately and completely administering household and community questionnaires by locating assigned households, using GPS technology to mark and record household locations, administering the household and community questionnaires, and finally communicating any problems to the supervisor.

3.4. Data Monitoring and Evaluation

As an additional aid to ensuring good quality data, extensive monitoring was performed throughout the fieldwork for each round of the survey. Two monitoring exercises were implemented during data collection. First, Survey Solutions' audio recording functionality was activated for 25 percent of the sample. These interview recordings were audited by 3 trained monitors, though not all recorded interviewers were able to be reviewed due to personnel constraints. On a daily basis, the monitors will listen to these recordings and fill in a structured questionnaire with their observations on interviewer performance. The feedbacks from these audio audits are then filtered to the respective interviewers.

The second quality check implemented were multiple visits to contacted households. Multiple visits were conducted by trained interviewers who are not part of the main data collection interviewers. Each day, up to 36 households that were contacted by the interviewing team are called by these call back interviewers. These interviewers conduct a short interview with the

household to confirm that the interviewer did indeed conduct the interview, that certain key elements were clearly stated to the respondent, that the interviewer conducted themselves in a professional manner, and other details on the interview process. Further, the team asked several time-invariant questions of the respondent to further confirm the interview was fully conducted and the interviewer captured the information correctly. Feedback from this quality checks were routed to the respective interviewers to improve on identified areas. Further, the interviewers also visited households that were not successfully contacted by the main interviewer. In some cases, the interviewer was able to reach the household. In such cases, the case was sent back to the interviewer to conduct the interview.

As a result of these quality checks, some of the interviewers were dropped from participating in the survey. There were also regular check-ins to address questions and issues the interviewers might have.

4. Data Management

4.1. Computer Assisted Personal Interview (CAPI)

This exercise was conducted using Computer Assisted Personal Interview (CAPI) techniques. The household questionnaire was implemented using the CAPI software, Survey Solutions. The Survey Solutions software was developed and maintained by the Data Analytics and Tools Unit within the Development Data Group (DECDG) at the World Bank. Each interviewer was given two tablets, which they used to conduct the interviews. Overall, implementation of survey using Survey Solutions CAPI was highly successful, as it allowed for timely availability of the data from completed interviews.

Basic information on every household was pre-loaded in the CAPI assignments for each interviewer. The information was pre-loaded to (1) assist interviewers in identifying the household and (2) ensure that each pre-loaded person is properly addressed and easily matched to the most recent interviews. Basic household information (location, household head name, phone number, etc.) was pre-loaded. The list of individuals from the previous interview and their basic characteristics were uploaded. This helped maintain the panel of individuals and ensured the status of each individual in the subsequent round of the survey.

4.2. Data Communication System

The data communication system used in the experiment was highly automated. Each interviewer was given a mobile modem allowing for internet connectivity and daily synchronization of their tablet. This ensured that head office in San Salvador has access to the

data in real-time. Once the interview is completed and uploaded to the server, the data is first reviewed by the Supervisors, and then routed for call back or audio audit if selected. A feedback questionnaire was also designed in Survey Solutions where interviewers receive respective feedback on their tablet from the various monitoring stages. This activity is done on a daily basis throughout the duration of the data collection.

4.3. Data Cleaning

The data cleaning process was done in three main stages. The first stage was to ensure proper quality control during the fieldwork. This was achieved in part by incorporating validation and consistency checks into the Survey Solutions application used for the data collection and designed to highlight many of the errors that occurred during the fieldwork.

The second stage cleaning involved the use of Supervisors in Survey Solutions. As indicated above, once the interview is completed and uploaded to the server, the Supervisors reviewed completed interviews for inconsistencies and extreme values. Depending on the outcome, they can either approve or reject the case. If rejected, the case goes back to the respective interviewer's tablet upon synchronization. The supervisor will provide general and question-specific comments when rejecting a particular completed interview. These errors were then corrected based on another visit to the household on the instruction of the supervisor. The data that had gone through this first stage of cleaning and has no issues is then approved by the Supervisor.

The third stage of cleaning involved a comprehensive review of the final raw data following the first and second stage cleaning. Every variable was examined individually for (1) consistency with other sections and variables, (2) out of range responses, and (3) formatting. Some minor errors remain in the data where the diagnosis and/or solution were unclear to the data cleaning team.

4.4. Confidential Information

For purposes of maintaining the confidentiality of the data, all names, phone numbers, and addresses have been removed from the datasets.

5. Using the Data

5.1. File Structure

The data should always be used in conjunction with the questionnaire. Where there are no issues of confidentiality, all the variables from the questionnaire have been included in the data sets. In some cases, there is an additional variable which contains the "other specify"

information that was written in the questionnaire. So, for example, the other specify variable will be indicated with an “_oth” attached to the variable name such as q5_oth containing the “other specify” information for a variable q5.

Every effort was made to keep question numbers (and thus variable names) as consistent as possible. If questions were added in the middle of a section, a letter was added to the question number at that space in the sequence (e.g., if added before question 2, the question number would be 2a). This was done to make utilization of the data sets across the rounds as consistent as possible.

5.2. Merging Datasets

The dataset contains the variable ***hhid***, which is a unique identifier for the household. This variable is used as the unique key variable in the merging of all household type datasets. In addition, the dataset includes the variable ***r_members__id*** that identifies the household member.